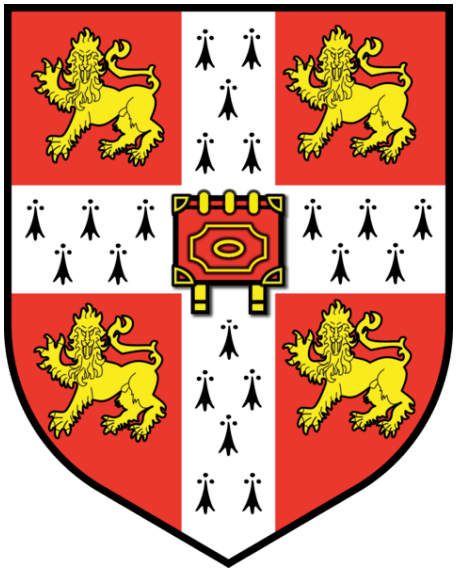
Genetic and molecular characterisation of metabolic subphenotypes and their contribution to type 2 diabetes aetiology

Alice Catherine Williamson

Department of Clinical Biochemistry
Girton College, University of Cambridge



Supervisors: Professor Sir Stephen O'Rahilly & Professor Claudia Langenberg

March 2023

This thesis is submitted for the degree of Doctor of Philosophy

Declarations

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.

The work presented in this thesis was completed under the supervision of Professor Sir Stephen O'Rahilly and Professor Claudia Langenberg and was conducted at the Institute of Metabolic Science Metabolic Research Laboratories and MRC Epidemiology Unit at the University of Cambridge since October 2019.

As prescribed by Degree Committee of the School of Clinical Medicine and Veterinary Medicine, this thesis does not exceed the word limit of 60,000 words, exclusive of figures, tables, appendices, and bibliography.

Alice Catherine Williamson

March 2023

Summary of Thesis: Alice Catherine Williamson

Genetic and molecular characterisation of metabolic subphenotypes and their contribution to type 2 diabetes aetiology

Systematic assessment of the genetic architecture of type 2 diabetes (T2D) has allowed the identification of hundreds of genetic loci contributing to T2D risk. However, understanding how these loci contribute to the aetiological complexity of T2D has been more challenging. Using an interdisciplinary approach this thesis aimed to investigate the genetic contribution to T2D, and related metabolic subphenotypes, to further the understanding of disease aetiology. This incorporated large-scale population genetic analyses of T2D and fine-scale intermediate traits, with targeted *in vitro* functional follow up of genes and variants of interest. Focused *in vitro* and genetic investigation of a single gene, *MC3R*, of which loss of function was previously thought to result in obesity and metabolic dysfunction, revealed that this was not the case. Combining functional assessment of coding variants in *MC3R* and association analyses at population-scale revealed a key role of *MC3R* in relaying nutritional cues to the timing of puberty, accrual of lean mass and height. At the other end of the spectrum, hypothesis-free genome-wide approaches examining traits intermediate to T2D, in combination with *in vitro* follow up, also have great utility in understanding T2D aetiology. A genetic discovery including > 55,000 individuals was conducted for insulin resistance after a glucose challenge (post-challenge insulin resistance) to approximate insulin resistance specifically in the post-prandial state, a key but understudied contributor to T2D onset. This revealed ten novel genetic loci for post-challenge insulin resistance ($P<5x10^{-8}$), including an association at *SLC2A4*, encoding the key insulin-stimulated glucose transporter GLUT4. Most of these post-challenge insulin resistance loci are shared with T2D, suggesting potential mechanisms behind these disease associations. Functional follow up using cell-based assays identified 9 novel candidate genes involved in insulin-stimulated glucose uptake and GLUT4 trafficking. Finally, integration of information of T2D genetic risk, T2D incidence, and untargeted plasma metabolomics identified distinct metabolomic signatures of specific T2D genetic endophenotypes associated with T2D development, which together with mediation analyses prioritised a subset of potential causal pathways. This work demonstrates the power of integrating population genetic analyses, refined traits intermediate to disease, and tailored *in vitro* follow up in prioritising key genes and pathways involved in T2D aetiology, and additionally highlights several candidates for further investigation.

Acknowledgements

I would like to thank the incredible group of people I have been fortunate to work with and learn from throughout my studentship. I have been amazed by the huge amount of support I have been given by my supervisors, lab mates, and collaborators throughout my PhD, and I hugely appreciative of everyone who has helped on this PhD journey.

First and foremost, I would like to thank my supervisors Professor Sir Stephen O'Rahilly and Professor Claudia Langenberg, for their guidance and support, and for always reminding me to think of the bigger picture throughout my research. Thank you to both of you for your openness to supervise an interdisciplinary research project, and for your trust in my abilities in doing this. You are both amazing scientific role models and I feel incredibly fortunate to have learnt from and worked with both of you.

Steve, thank you giving me the confidence and the opportunity to pursue ambitious research projects and always ensuring I was on a path to succeed during my PhD. I am grateful for the opportunity to have been involved in such a wide range of projects throughout this time, and the experience I have built in doing so. Scientific discussions with you have always been inspiring and often gave me completely new perspective of my research and the connections of this to the wider field.

Claudia, thank you for always challenging me to be ambitious in my research and encouraging me to push myself out of my comfort zone. Thank you for giving me so many opportunities and always opening many doors to build experience and self-belief as a researcher. I am particularly grateful for the opportunity to have led a collaborative research project during my PhD and your endless support throughout this. Thank you for always being so generous with your time and for always having my back.

I would also like to express an immense amount of gratitude to the post-doc researchers I have been lucky to work with throughout my PhD. I am particularly grateful for the support of Dr Eleanor Wheeler, Dr Maik Pietzner and, Dr Brian Lam. Ellie, thank you for advising me in all things GWAS, and help navigating the learning curve of working with a consortium. Thank you for time in answering my sometimes

never-ending list of questions and for always going the extra mile to help. Maik, your supervision and guidance throughout my rotation project and during my PhD provided me with the foundation of skills and confidence I needed to succeed in many of the analytical parts of my research, and I am so grateful for your time and patience. Brian, I can't thank you enough for all your help and advice in several aspects of my PhD both in the lab and computationally. Thank you for always being so generous with your time and for never making it feel like any question was too small. I am also grateful to Dr Isobel Stewart, Dr Felix Day, and Dr Nuno Rocha, whose support and advice was critical in several projects during my PhD.

I would like to thank Dr Daniel Fazakerley and Dr Dougall Norris, for the opportunity to collaborate and work with them during my PhD. This project allowed me to achieve a long-term ambition of taking findings from GWAS into the lab. I am grateful for their patience, guidance, and positivity. I would also like to thank the MAGIC "Refined Insulin" traits working group for their collaboration and invaluable feedback on the genetic analyses of post-challenge insulin resistance.

I am grateful everyone in the wider COY group at the IMS-MRL and Aetiology group at the MRC Epidemiology Unit. I feel very lucky to have had the opportunity to work with such a friendly and collaborative group of colleagues. Thank you particularly to Professor Nick Wareham and Professor Giles Yeo for their valuable feedback and advice. Thank you to Kara, Katie, Deb, and Nicola for their help and for ensuring that all projects run as smoothly and efficiently as possible. I would also like to thank fellow PhD students both past and present: Julia, Mine, Chunxiao, Sam, Katie, Xiaoming, Vicky, Nick, Lina, and Maria. Thank you to Carole, Meriel, and Laura for their administrative support that ensured this PhD journey was as smooth as possible.

Huge thank you to my wonderful friends in Cambridge, particularly Kat, Evelyna, and Julia. Thank you for making this time so much fun. We did it!

Finally, I would like to thank my parents and sister for their unconditional love, encouragement, and support. Mum and Dad - thank you for everything you have done which set me on a path to succeed and for always encouraging me to aim beyond what I think I am capable of. I am beyond grateful for everything.

Publications and Presentations:

Publications

Published and accepted manuscript related to the work presented in this thesis.

Lam B.Y.H.*, **Williamson A**.*, Finer S.* [34 authors] Langenberg C., Perry J.R.B., Yeo G.S., O'Rahilly S., 2021. MC3R links nutritional state to childhood growth and the timing of puberty. **Nature**, 599(7885), pp.436-441. * Joint first author

Williamson A, [100 authors], Mohlke K.L., Wheeler E, O'Rahilly S, Fazakerley D.J., Langenberg C. Genome-wide association study of postprandial glucose metabolism and functional characterisation identifies candidate genes for insulin-stimulated glucose uptake. **Nature Genetics** (2023, accepted)

Publications co-authored since starting PhD:

Broadaway K.A., Yin X, **Williamson A**, [64 authors], Langenberg C, Wheeler E, Mohlke K.L. Loci for insulin processing and secretion provide insight into type 2 diabetes risk. (2023) **American Journal of Human Genetics**; 110(2):284-299.

Wade K.H., Lam B.Y.H., Melvin A, [10 authors], **Williamson A**, Zhang C, Farooqi I.S., Yeo G.S.H., Timpson N.J., O'Rahilly S. Loss-of-function mutations in the melanocortin 4 receptor in a UK birth cohort. **Nature Medicine** 2021;27(6):1088-1096. doi: 10.1038/s41591-021-01349-y.

Papers related to the thesis in preparation for submission:

Williamson A *et al* Genetic Heterogeneity of Type 2 Diabetes is reflected in the circulating metabolome. In preparation.

Scientific Conferences and Meetings (relevant chapter indicated in brackets):

Genetic architecture of post-challenge insulin resistance (Chapter 2 and 3)

- Keystone Symposia – Adipose Tissue: Energising Good Fat (January 2023, US): Oral and poster presentation.
- American Society of Human Genetics Conference (October 2022, US): Poster presentation.
- Metabolism In Action Meeting (September 2022, Denmark): Poster presentation
- Danish Diabetes Academy PhD Summer School in Diabetes and Metabolism (August 2021, Denmark): Flash Talk and Poster Presentation.

Metabolomic Signatures of type 2 diabetes (Chapter 4)

- European Association for the Study of Diabetes, Study Group for Genetics of Diabetes (EASD-SGGD, May 2022, France): Oral presentation.

Role of MC3R in human physiology (Chapter 5)

- International Congress of Neuroendocrinology (August 2022, UK): Poster presentation
- American Society of Human Genetics Meeting (October 2021, Online): Oral Presentation in Platform Session.
- UK Biobank Scientific Meeting (November 2021, Online): Talk shortlisted as part of Doctoral Student of the Year Competition.

Commonly Used Abbreviations

$logEC_{50}$	log half maximal effective concentration
^{1}H-NMR	proton nuclear magnetic resonance
2DOG	2-deoxyglucose
AD-AMR	Admixed American Ancestry
adjBMI	adjusted for body mass index
AFR	African ancestry
ALSPAC	Avon Longitudinal Study of Parents and Children study
BCAAs	branched chain amino acids
BMI	body mass index
cAMP	cyclic adenosine 3',5'-monophosphate
CI	confidence interval
cLoF	complete loss of function
ddCT	double delta CT
DNA	deoxyribonucleic acid
EA	effect allele
EAF	effect allele frequency
EAS	East Asian ancestry
E_{max}	average relative maximal efficacy
EMSA	Electrophoretic mobility shift assays
EPIC	European Prospective Investigation of Cancer study
eQTL	expression quantitative trait locus
EUR	European ancestry
FDR	false discovery rate
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
G&H	Genes & Health Study
GRS	genetic risk score
GSVs	GLUT4 secretory vesicles
GWAS	Genome-wide association study
HbA1C	glycated haemoglobin
HEK293	Human Embryonic Kidney 293 cells
HIS-AMR	Hispanic American ancestry
HOMA	homeostatic model assessment
IFC	insulin fold change during an oral glucose tolerance test
indel	insertion or deletion
IR	insulin resistance
ISI	Modified Stumvoll Insulin Sensitivity Index
IVW	inverse variance weighted
LD	linkage disequilibrium
LDs	Lipid droplets
LoF	loss of function
MAC	minor allele count
MAF	minor allele frequency
MODY	maturity onset diabetes of the young
mQTL	metabolite quantitative trait locus
MR	Mendelian Randomisation
MS	mass spectrometry
NDP-MSH	[Nle^{4}, D-Phe^{7}]-a-melanocyte-stimulating hormone

NSS	normal swine serum
NT/ NT ctrl	non-targeting control
OA	Other/alternate allele
OGTT	oral glucose tolerance test
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PCs	principal components
PGS	polygenic score
pLoF	partial loss of function
PPA	posterior probability of association
pPGS	partial polygenic risk score
pQTL	protein quantitative trait locus
QC	quality control
QR	quasi-random
QTL	quantitative trait locus
RBC	Red blood cell
RNA	ribonucleic acid
RNAseq	RNA sequencing
rsid	SNP identifier, rs number
RT	room temperature
SAS	South Asian ancestry
scRNA-seq	single cell RNA sequencing
SD	standard deviation
SE/SEM	standard error
siRNA	short interfering RNA
SNP	single nucleotide polymorphism
SNVs	single nucleotide variants
sQTL	splice quantitative trait locus
T2D	type 2 diabetes
TfR	transferrin receptor
UKBB	UK biobank study
WES	whole exome sequencing
WGS	whole genome sequencing
WHR	waist to hip ratio
WT	wild type

List of Figures

Chapter 1

Figure 1.0: Interdisciplinary framework to understand disease aetiology using genetics **P1**
Figure 1.1: Tissue specific responses to insulin in healthy and insulin resistant states in key tissues of insulin action. **P9**
Figure 1.2: Downstream effects of proximal insulin signalling in metabolic control. **P10**
Figure 1.3: The pathophysiological sequence of insulin resistance and beta cell dysfunction in T2D onset. **P16**
Figure 1.4: Relative effect size of variants at different allele frequencies and their primary method of discovery. **P22**
Figure 1.5: integration of genomics with other 'omics measures to identify quantitative trait loci (QTLs). **P30**

Chapter 2

Figure 2.1: Study design to assess genetic architecture of post-challenge insulin resistance. **P44**
Figure 2.2: Meta-analysis workflow for genetic discovery analyses for ISI and IFC. **P52**
Figure 2.3: Summary of integrated GWAS approach strategies. **P57**
Figure 2.4: Observational Correlation (pairwise Spearman's Rank) of Insulin Fold Change and Modified Stumvoll ISI with metabolic traits in the Fenland Study. **P64**
Figure 2.5: Observational Correlation (pairwise Spearman's Rank) of insulin sensitivity and clearance related traits in the RISC cohort. **P66**
Figure 2.6: Loci associated with post-challenge insulin resistance at genome-wide significance overlap with associations with other glycaemic traits and T2D. **P68**
Figure 2.7: Two independent signals were identified at PPP1R3B for insulin fold change adjusted for BMI. **P70**
Figure 2.8: Manhattan plot of genetic associations with ISI in European only meta-analysis unadjusted for BMI. **P72**
Figure 2.9: Regional association plot at BICC1 for IFC adj.BMI in non-European meta-analysis. **P73**
Figure 2.10: Forest plot of beta estimates of associations of rs60453193 with insulin fold change in individual cohorts. **P74**
Figure 2.11: LD matrix for lead SNPs in IFC, ISI, and T2D GWAS at SLC2A4 locus in European ancestry **P80**
Figure 2.12: SLC2A4 locus regional association plot for Type 2 Diabetes across studies in different ancestry groups. **P81**
Figure 2.13: rs117643180 at SLC2A4 affects GLUT4 expression and transcription regulation **P85**
Figure 2.14: Statistical colocalisation of IFC with conditional summary statistics for eQTL in skeletal muscle at the SLC2A4 locus. **P86**
Figure 2.15: Statistical colocalisation of ISI with conditional summary statistics for eQTL in skeletal muscle at the SLC2A4 locus. **P87**
Figure 2.16: rs117643180 exhibits allelic differences in transcription factor binding. **P90**
Figure 2.17: Integrated GWAS approach highlights additional loci including those implicating glucose transport. **P95**
Figure 2.18: Joint colocalisation at loci associated with IFC or ISI with related traits **P96**

Chapter 3

Figure 3.1: Schematic overview of the insulin-stimulated glucose uptake pathway in muscle and adipose tissue. **P112**

Figure 3.2: Transport and recycling of GLUT4 from the plasma membrane. **P113**

Figure 3.3: Confirmation of knockdown of positive control genes by Western Blot. **P117**

Figure 3.4: GLUT4 screen results for 3T3-L1 WT and 3T3-L1 HA-GLUT4-mRuby cells for all targets. **P126**

Figure 3.5: Average percentage of mRNA detected in knockdown samples compared to pooled control RNA. **P127**

Figure 3.6: 14 genes at loci associated with insulin fold change were identified to significantly impact GLUT4 trafficking in adipocytes. **P131**

Figure 3.7: Glucose Transport relative to non-targeting control. **P135**

Chapter 4

Figure 4.1: PRISMA diagram of systematic literature review on the metabolomic signature of type 2 diabetes. **P153**

Figure 4.2: Metabolites associated with incident T2D identified by systematic literature review. **P161**

Figure 4.3: 382 metabolites are associated with incident Type 2 Diabetes in the Epic-Norfolk. **P162**

Figure 4.4: T2D PGS associations mirror those seen observationally. **P164**

Figure 4.5: number of metabolites significantly associated (FDR < 0.05) with each SNP in the T2D PGS. **P165**

Figure 4.6: SNP-wise metabolite associations for the SNPs included in the T2D PGS, plus GCKR. **P166**

Figure 4.7: Metabolite associations with partial genetic risk scores for T2D. **P168**

Figure 4.8: Causal metabolites mediating type 2 diabetes genetic risk to incidence of Type 2 Diabetes. **P170**

Figure 4.9: The inverse association of glycine with T2D and T2D genetic risk is a consequence of insulin resistance. **P175**

Figure 4.10: Proposed genetically driven framework to refine and assess the causal role of 'omics measures in common diseases such as type 2 diabetes. **P179**

Chapter 5

Figure 5.1: Workflow summary of functional reverse genetic approach to study role of MC3R in humans. **P185**

Figure 5.2: Posttranslational processing of POMC into POMC derive peptides. **P187**

Figure 5.3: Downstream signalling of MC3R/MC4R in neurons. **P190**

Figure 5.4: Schematic overview of the location of variants of interest in MC3R. **P213**

Figure 5.5: Functional characterisation of MC3R coding variants. **P215**

Figure 5.6: Association of MC3R variants with relevant traits in UK Biobank **P217**

Figure 5.7: PheWAS analysis of MC3R genetic risk score. **P223**

Figure 5.8: Characteristics of an individual who is homozygous for the MC3R p.G240W mutation. **P226**

Figure 5.9: A bifurcating model of central melanocortin signalling. **P238**

Chapter 6

Figure 6.1: Integration of population genetics and in vitro follow up to identify novel candidates involved in post-prandial insulin action and T2D. **P242**

Appendix

Supplementary Figure S2.1: Manhattan and QQ plots for European-only meta-analysis of Modified Stumvoll ISI adjusted for BMI. **P300**
Supplementary Figure S2.2: Manhattan and QQ plots for European-only meta-analysis of Modified Stumvoll ISI unadjusted for BMI **P300**
Supplementary Figure S2.3: Manhattan and QQ plots for European-only meta-analysis of insulin fold change adjusted for BMI **P301**
Supplementary Figure S2.4: Manhattan and QQ plots for European-only meta-analysis of insulin fold change unadjusted for BMI **P301**
Supplementary Figure S2.5: rs117643180 exhibits allelic differences in transcriptional activity **P302**
Supplementary Figure S4.1: Comparison of significant PGS-metabolite associations for BMI-unadjusted and adjusted T2D PGS. **P330**
Supplementary Figure S4.2: Comparison of significant PGS-metabolite associations for the 236 and 93 SNP T2D PGS. **P331**
Supplementary Figure S4.3: Metabolomic signature for incident diabetes upon adjustment for established T2D risk factors. **P332**
Supplementary Figure S4.4: number of metabolites nominally associated ($P < 0.05$) with each SNP in the T2D PGS. **P333**
Supplementary Figure S4.5: Dosage plots of effect sizes of glycine variants on standard deviations of glycine levels versus the log odds for T2D unadjusted for BMI. **P334**
Supplementary Figure S4.6: Dosage plots of effect sizes of glycine variants on standard deviations of glycine levels versus the effect sizes for Fasting insulin unadjusted for BMI. **P335**
Supplementary Figure S4.7: Dosage plots of effect sizes of glycine variants on standard deviations of glycine levels versus the effect sizes for Waist to hip ratio adjusted for BMI (WHR adjBMI). **P336**
Supplementary Figure S4.8: Dosage plots of the effect sizes for fasting insulin unadjusted for BMI verses effect sizes of glycine variants on standard deviations of glycine levels. **P337**
Supplementary Figure S4.9: Dosage plots of the effect sizes for Waist to hip ratio adjusted for BMI (WHR adjBMI) verses effect sizes of glycine variants on standard deviations of glycine levels. **P338**
Supplementary Figure S4.10: additional adjustment with WHR attenuates almost all candidate causal mediator metabolites in the mediation framework. **P339**

List of Tables

Chapter 1

Table 1.1: Five phenotypic subtypes of adult-onset diabetes defined by Ahlqvist et al (2018). **P27**
Table 1.2: Genetic subtypes of Type 2 Diabetes defined by Udler et al (2018) **P28**

Chapter 2

Table 2.1: Contributing study sample size and ancestry for insulin fold change (IFC). **P47**
Table 2.2: Contributing study sample size and ancestry for Modified Stumvoll insulin sensitivity index (ISI). **P47**
Table 2.3: Variant weights used in Genetic Risk Score (GRS) calculation for IFC and ISI, adjusted and unadjusted for BMI. **P55**
Table 2.4: Primers to test rs117643180 for allelic effects on transcriptional regulation. **P60**
Table 2.5: Genome-wide significant loci for insulin fold change and/or Modified Stumvoll ISI in meta-analysis of cohorts of European ancestry. **P69**
Table 2.6: 99% credible sets defined by fine mapping for Insulin Fold Change adjBMI in European only meta-analysis. **P71**
Table 2.7: 99% credible sets defined by fine mapping Modified Stumvoll Insulin Sensitivity Index adjBMI in European only meta-analysis. **P71**
Table 2.8: colocalisation of T2D with IFC/ISI at post-challenge insulin resistance genetic loci. **P76**
Table 2.9: Reported associations of post-challenge insulin resistance loci with type 2 diabetes and other glycaemic traits. **P77**
Table 2.10: Differences in allele frequency and LD across populations may explain the lack of colocalisation of T2D and ISI/IFC in Europeans. **P79**
Table 2.11: Statistical colocalisation of IFC and ISI with conditional statistics for eQTL for SLC2A4 in skeletal muscle (GTEx v8). **P88**
Table 2.12: Integration of additional phenotypic layers identifies additional post-challenge insulin resistance loci of interest. **P92**
Table 2.13: Annotation of loci identified in the integrated GWAS approach with their association with other traits. **P93**
Table 2.14: eQTL colocalisation at loci identified in the integrated GWAS approach. **P98-99**

Chapter 3

Table 3.1: TaqMan probes (Thermo Fisher) used to quantify mRNA expression. **P123**
Table 3.2: Quantitative PCR (qPCR) assessment of knockdown efficiency of targets in 3T3-L1 adipocytes. **P128**
Table 3.3: GLUT4 screen results in 3T3-L1 WT cells: Prioritised and Control Genes. **P133**
Table 3.4: Mapping of genes where knockdown significantly impacted GLUT4 trafficking/regulation to insulin fold change associated loci. **P141**

Chapter 4

Table 4.1: Causal mediation results for metabolites mediating association of T2Dpgs to T2D. **P171**
Table 4.2: Genetic colocalisation of T2D and glycine at glycine associated loci. **P173**
Table 4.3: Mendelian Randomisation Results for glycine, T2D and insulin resistance related traits - Fasting insulin (FI) and WHR adjusted for BMI (WHRadjBMI). **P174**

Chapter 5

Table 5.1: Summary of key phenotypic features of reported genetic deficiency in the central melanocortin system **P195**
Table 5.2: Genotype concordance amongst Whole-exome Sequencing (WES), genotyping arrays and imputation in UK Biobank. **P198**
Table 5.3: MC3R Genetic Risk Score Weighting. **P201**
Table 5.4: Number of variants prioritised for in vitro functional follow up across 4 population studies. **P208**
Table 5.5: Frequencies of variants of interest in population cohorts. **P210**
Table 5.6: variants identified in ALSPAC, Genes & Health (G&H) and UK Biobank (UKBB) studies. **P211**
Table 5.7: MC3R cAMP assay results for variants of interest. **P214**
Table 5.8: Phenotype association of array genotyped MC3R variants identified in UKBB. **P218-219**
Table 5.9: Independent association of array genotyped MC3R variants with age of menarche in UK Biobank. **P220**
Table 5.10: MC3R GRS associations with traits of interest in UK Biobank. **P221**
Table 5.11: Clinical characteristics of homozygous MC3R G240W proband. **P225**
Table 5.12: Enrichment of MC3R predicted damaging variants in UKBB participants reporting delayed age of menarche (N = 246,262). **P228**
Table 5.13: Enrichment of MC3R predicted damaging variants in UKBB participants reporting delayed age of voice breaking. **P229**

Appendix

Supplementary Table S2.1: Association of Genetic Risk scores for IFC and MS ISI, and their comprising SNPs with relevant traits in UK Biobank (Unrelated individuals of European ancestry). **P303**
Supplementary Table S2.2: Association of Genetic Risk scores for IFC and MS ISI, and their comprising SNPs with relevant traits in Fenland (Unrelated Europeans).**P304**
Supplementary Table S2.3: LD score regression results for IFC adjBMI with cardiometabolic traits in cohorts of European ancestry. **P305-306**
Supplementary Table S2.4: LD score regression results for ISI adjBMI with cardiometabolic traits in cohorts of European ancestry. **P307-308**
Supplementary Table S3.1: Protein coding genes taken forward for screening at loci associated with insulin fold change. **P309**
Supplementary table S3.2: Horizon Discovery Dharmacon siGenome siRNA pools used for knockdown. **P310-314**
Supplementary Table S3.3: Effect of knockdown of targets of interest on Glucose Transport in 3T3-L1 adipocytes upon stimulation with 0.5nM insulin. **P315**
Supplementary Table S3.4: GLUT4 screen results in 3T3-L1 WT cells (N = 3). **P316-319**
Supplementary Table S3.5: GLUT4 screen results in 3T3-L1 HA-GLUT4 cells (N = 5). **P320-323**
Supplementary Table S3.6: GLUT4 screen results in 3T3-L1 wildtype cells (N = 3) for lipid droplets and nuclei. **P324-325**
Supplementary Table S3.7: Effect of knockdown of targets of interest on Glucose Transport in 3T3-L1 adipocytes upon stimulation with 0.5nM insulin. **P326**
Supplementary Table S4.1: Systematic Review Search Terms used in PubMed. **P328**

Supplementary Table S4.2: Identified studies examining the association of plasma metabolite levels with incident T2D in systematic literature review (up to 12th October 2022). **P340-249**
Supplementary Table S4.3: SNPs and their weighting in the 236 SNP T2D PGS. **P350-353**
Supplementary Table S4.4: SNPs and corresponding weights used in the partial PGS from Udler et al (2018). **P354-356**
Supplementary Table S4.5: SNPs and the corresponding weights used in the construction of the 93 SNP T2D PGS. **P3357-358**
Supplementary Table S4.6: Metabolite associations with the T2D BMI unadjusted Genetic Risk Score in EPIC-Norfolk Quasi-random subcohort. **P359-363**
Supplementary Table S4.7: Metabolite associations with the partial T2D genetic risk scores defined by Udler et al (2018) in EPIC-Norfolk quasi-random subcohort. **P364-367**
Supplementary Table S5.1: Proteomics associations with MC3R variants of interest in the Fenland Study. **P368**
Supplementary Table S5.2: Metabolomics associations with MC3R variants of interest in the EPIC-Norfolk Study. **P369-372**

TABLE OF CONTENTS

Declarations I

Summary of thesis II

Publications and Presentations: V

Commonly Used Abbreviations VII

List of Figures IX

List of Tables XII

Introduction 1

AIMS: 1

RATIONALE 3

OUTLINE OF THESIS 4

Chapter 1: Overview of the literature 6

1.1 Background 8

1.2 Normal physiology: Maintenance of whole-body glucose homeostasis 9

1.3 Molecular mechanisms in insulin resistance 13

1.4 Genetic approaches to understand the aetiology of metabolic disease 17

1.4.1 The genetic architecture of type 2 diabetes 17

1.4.2 Genetic architecture of intermediate traits: glycaemic traits and T2D 18

1.4.3 Rare and low frequency variants and T2D risk 21

1.5 The genetic and phenotypic heterogeneity of T2D 26

1.6 From GWAS to function: going beyond genetic associations with disease 30

Integration of 'omics layers 30

Mendelian Randomisation and genetic colocalisation 32

What is the causal gene at a genetic locus? 32

1.7 Conclusions 35

Chapter 2: Genetic architecture of insulin resistance post-glucose challenge 36

2.1 AIMS AND OBJECTIVES 39

2.2 INTRODUCTION 40

2.3 METHODS 44

2.3.1 Study Design 44

2.3.2 Correlation of ISI and IFC with relevant traits 48

2.3.3 Genetic discovery: Study-level GWAS and meta-analysis 49

2.3.4 loci characterisation 53

Fine mapping 53

Statistical colocalisation of IFC and ISI with T2D 53
Phenotypic annotation of associated loci 53
Genetic risk score analysis 54
LD Score Regression 56
2.3.5 Integrated GWAS approach 56
2.3.6 Tissue of action 58
2.3.7 Characterisation of the SLC2A4 locus 59
2.4 RESULTS 62
2.4.1 Study overview 62
2.4.2 Observational correlation of post-challenge insulin resistance with related traits: 62
2.4.3 Genetic discovery of post-challenge insulin resistance 66
2.4.3 Assessing overlap of post-challenge insulin resistance with type 2 diabetes 75
2.4.4 Genetic risk scores and LD score regression analysis 83
2.4.5 In silico and in vitro characterisation of the SLC2A4 locus (rs117643180) 84
2.4.6 Integration of additional phenotypic layers to identify additional loci for post-challenge insulin resistance 91
2.4.7 Tissue of action 97
2.5 DISCUSSION 100
Strengths and limitations 102
Conclusion 103

Chapter 3: Investigating candidate regulators of insulin-stimulated glucose transport and GLUT4 trafficking 104

3.1 AIMS AND OBJECTIVES 108
3.2 INTRODUCTION 109
3.3 METHODS 114
3.3.1 Gene selection: 114
3.3.2 Assay Overview 115
3.3.3 Cell culture 115
3.3.4 siRNA-mediated gene knockdown in differentiated adipocytes 116
3.3.5 GLUT4 translocation assay 118
3.3.6 Radiolabelled glucose transport assay 120
3.3.7 Confirmation of gene knockdown 121
3.4 RESULTS 125
3.4.1 Rationale for studying genes at loci associated with insulin fold change (IFC) 125
3.4.2 Evaluation of the system used 125
3.4.3 Known regulators of GLUT4 traffic were genetically prioritised and shown effect GLUT4 trafficking 129
3.4.4 Novel genes involved in GLUT4 trafficking 131
3.4.5 Glucose transport 134
3.4.6 Other impacts of knockdown of the 9 novel GLUT4 regulators of interest 135
3.4.7 Possible mechanisms of action for the novel regulators of GLUT4 traffic 136

3.4.8 Evaluation of gene selection strategy 139
3.5 Discussion 142
Strengths and limitations 144
Conclusions and future directions 145

Chapter 4: Dissecting the genetic heterogeneity of type 2 diabetes using untargeted plasma metabolomics. 147

4.1 Aims and Objectives: 149
4.2 Introduction 150
4.3 Methods 152
4.3.1 Systematic literature review 152
4.3.2 The EPIC-Norfolk Study 153
4.2.3 Polygenic Score Calculation 155
4.2.4 Statistical Analyses 156
4.2.5 Causal mediation analyses 157
4.2.6 Assessing the causal role of circulating glycine in T2D 157
4.3 Results 160
4.3.1 A systematic literature review reveals a plethora of metabolites associated with T2D risk 160
4.3.2 More than 300 metabolites are significantly associated with incident T2D in the EPIC-Norfolk T2D case cohort 161
4.3.3 Convergence of metabolite associations for genetic susceptibility to T2D and metabolic profiles associated with incident T2D 163
4.3.4 Genetic T2D endotypes have distinct metabolomic signatures 167
4.3.5 Mediation analysis highlights metabolites on a potentially causal pathway to T2D 169
4.3.6 Changes in circulating glycine are likely a consequence rather than a cause of insulin resistance in T2D 172
4.4 Discussion 176
Strengths and limitations 180
Conclusion 181

Chapter 5: Phenotypic consequences of loss of function mutations in the human Melanocortin 3 Receptor (MC3R) 182

5.1 Aims and Objectives 185
5.2 Introduction 186
5.2.1. Melanocortin receptor agonists and antagonists 186
5.2.2 Peripheral role of melanocortin receptors (MC1R, MC2R, MC5R). 188
5.2.3 The central melanocortin system (MC3R and MC4R) 188
5.2.4 Genetic evidence of the role of proteins in the central melanocortin system. 191
5.2.5 Melanocortin 3 Receptor (MC3R) 196
5.3 Methods 197

5.3.1 Variant identification across the different cohorts ... *197*
5.3.2 In-vitro cyclic-AMP (cAMP) accumulation assay ... *199*
5.2.3 Phenotypic Association of MC3R in UK Biobank ... *200*
5.3.4 Phenome-wide association study ... *202*
5.3.5 Genes & Health ... *202*
5.3.6 The Fenland Study ... *204*
5.2.7 The European Prospective Investigation of Cancer (EPIC)-Norfolk study ... *205*
5.2.8 Assessment of frequency of MC3R variants in individuals with delayed puberty ... *205*
5.3 RESULTS ... 208
5.3.1 Prioritisation of MC3R variants in several population cohorts: ... *208*
5.3.2 Functional characterisation of MC3R coding variants identifies variants that result in loss of function. ... *209*
5.3.3 Heterozygous carriage of dysfunctional MC3R variants ... *216*
5.3.4 Phenotypic follow up ... *220*
The Fenland Study: association of MC3R LoF with additional metabolic traits ... 222
Phenome-wide association study (pheWAS) of MC3R ... 222
'Omic associations with MC3R LoF ... 223
5.3.5 A human knockout: homozygous carriage of MC3R LoF variant MC3R p. G240W ... *223*
5.3.6 Investigation of the enrichment of MC3R variants in individuals who reported delayed puberty in UK Biobank ... *227*
5.4 SUPPORTING EVIDENCE: RESULTS FROM COLLABORATION. ... 230
5.5 DISCUSSION ... 232
Strengths and Limitations ... *236*
Conclusions ... *237*

Chapter 6: Concluding Discussion ... 239

6.1 SUMMARY OF FINDINGS ... 240
Refined glycaemic traits intermediate to T2D have great utility in furthering the understanding of T2D aetiology ... *240*
Studying the genetic architecture of post-prandial insulin resistance reveals loci contributing to T2D risk via this pathway ... *240*
Genetically informed functional follow up reveals novel regulators of GLUT4 trafficking ... *241*
Genetic subtypes of T2D have distinct metabolomic signatures, implicating specific aetiological pathways. ... *242*
A functional reverse genetic approach allowed identification of the role of MC3R in humans. ... *243*
6.2 STRENGTHS AND LIMITATIONS ... 245
Strengths ... *245*
Limitations and considerations ... *247*
6.3 IMPLICATIONS ... 253
6.4 FUTURE DIRECTIONS ... 255
Conclusion ... *257*

References: **258**

Appendix **299**

Supplementary Material: Chapter 2 300

Supplementary Material Chapter 3 309

Supplementary Material Chapter 4 328

Supplementary Material: Chapter 5 368

Analysis Plans 373

Introduction

Aims:

The overarching aim of the work described in this thesis is to identify novel aetiological pathways in type 2 diabetes (T2D) and related subphenotypes, through the integration of large-scale genetic data, refined glycaemic traits, 'omics measures and *in vitro* follow up (**Figure 1.0**). A primary focus of this thesis is to leverage genetics to explore and further investigate the role of post-prandial insulin resistance in T2D, but other sections of this thesis explore other aspects of T2D more broadly, as well as T2D risk factors.

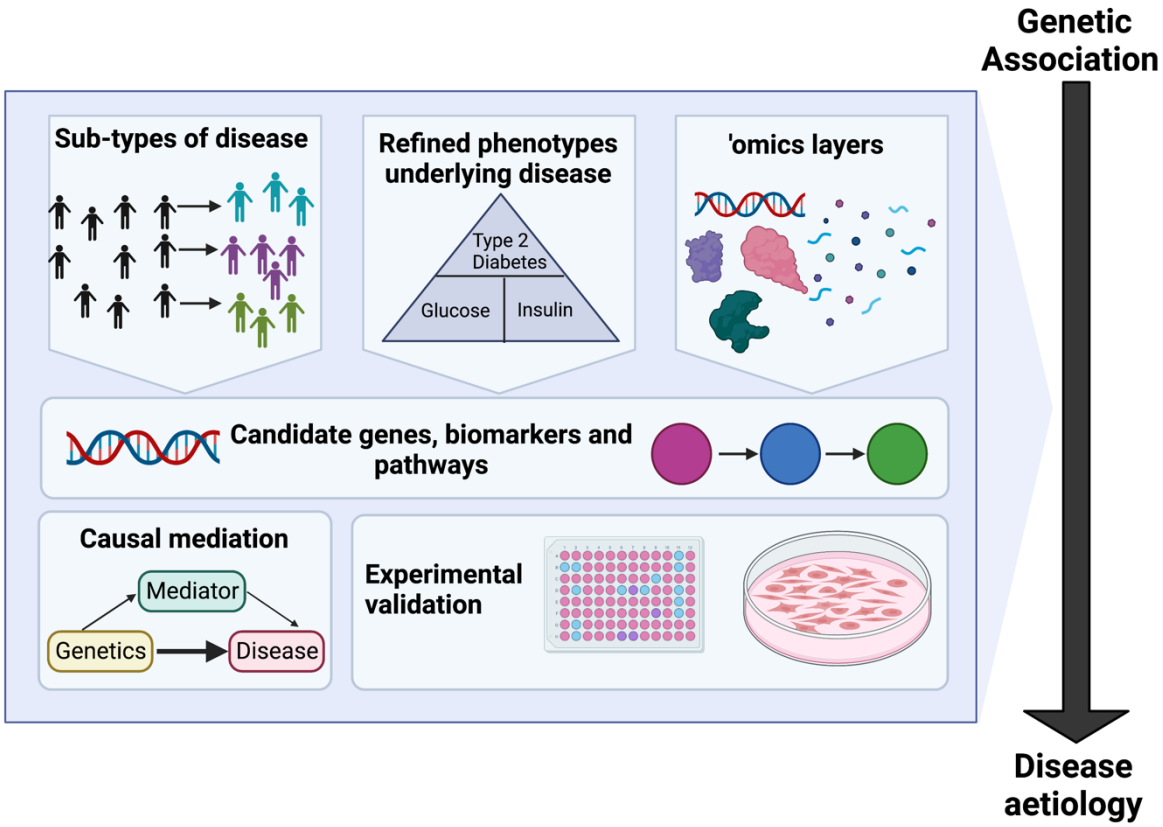


Figure 1.0 Interdisciplinary framework to understand disease aetiology using genetics

This overall aim was addressed through the following sub-aims and objectives:

Sub-aim 1: Assess the genetic architecture of post-prandial insulin resistance.

1.1 Identify genetic loci associated with dynamic glycaemic traits measured during an oral glucose tolerance test (OGTT) which approximate post-prandial insulin resistance through genome-wide meta-analysis.

1.2 Use an integrative GWAS approach to identify additional loci of interest below the traditional genome-wide significance threshold through joint discovery with related traits.

Sub-aim 2: *In vitro* characterisation of prioritised candidate genes hypothesised to play a role in post-prandial insulin resistance.

2.1 Identify a target list of candidate genes located at genetic loci associated with post-challenge insulin resistance.

2.2 Using cellular models assess the role of these genes in GLUT4 trafficking and insulin-stimulated glucose uptake.

Sub-aim 3: Identify a refined metabolomic signature for T2D and T2D endotypes, and investigate the potential causal role of metabolites in T2D incidence.

3.1 Systematic literature review to assess reported associations of metabolites with T2D incidence.

3.2 Identify metabolomic signatures associated with T2D incidence and T2D genetic risk.

3.3 Assess the metabolomic associations with genetic risk scores representing different T2D endotypes.

3.4 Assess the role of metabolites as candidate causal mediators from T2D genetic risk to T2D incidence.

Sub-aim 4: Investigate the role of *MC3R* in human physiology using a functionally informed reverse genetics approach.

4.1 Identify coding variants in *MC3R* found in population-based studies that are predicted to be damaging *in silico*.

4.2 Characterise the true damaging impact of these *MC3R* variants using *in vitro* assays.

4.3 Assess the association of truly damaging *MC3R* variants with phenotypic traits in population-based studies.

Rationale

T2D is a global problem, and the health and socioeconomic burden of T2D is rising due to ever increasing numbers of diagnoses. Whilst genome-wide association studies (GWAS) have identified hundreds of robustly associated genetic loci associated with increased risk of T2D, the molecular mechanisms behind such associations are still poorly understood. Increasing understanding of such mechanisms will allow novel insight in T2D aetiology, and potential avenues for treatment and intervention. Human genetics is a tractable way to explore such questions. Traits intermediate to disease and large-scale 'omics measures have become increasingly available in population studies and provide one avenue to identify biological mechanisms behind genetic associations with disease at scale.

Such insights can be coupled with *in vitro* functional characterisation which allows direct assessment of the role of candidate genes in disease-associated pathways and disease processes. *In vitro* characterisation of genetic variants of interest can further inform genetic analyses through prioritisation of variants of interest.

A cross-discipline approach allows more refined analyses to identify the layers of complexity in the aetiology of T2D, and thus novel candidate genes and pathways. Similarly, focused examination of a single gene using such approaches can provide valuable insight into gene function where this is still poorly understood.

Outline of thesis

Chapter 1 gives an overview of the literature, highlighting what is currently known about the genetic architecture of T2D and related traits, and how genetically driven approaches have furthered the understanding of the aetiology of T2D.

Chapter 2 focuses on the genetic characterisation of postprandial insulin resistance in > 55,000 individuals. Postprandial insulin resistance was estimated using two indices calculated from dynamic measures of insulin and glucose during an OGTT – insulin fold change and Modified Stumvoll insulin sensitivity index (collectively referred to as post-challenge insulin resistance). I describe the results of this genetic discovery across 3 ancestry groups – European, Hispanic American, and East Asian. I conducted follow up analyses to understand the relationship of these loci with T2D and related traits, as well as expression quantitative trait loci (eQTLs), aiming to identify candidate mechanisms of action at these loci. An integrated GWAS approach allowed the identification of an additional set of prioritised loci at suggestive significance threshold.

Chapter 3 outlines results of *in vitro* follow up of genes located at post-challenge insulin resistance loci (Chapter 2) using cellular models to investigate their role in insulin stimulated glucose uptake and GLUT4 trafficking. I used an siRNA knockdown assay in 3T3-L1 adipocytes to identify changes in GLUT4 localisation upon stimulation compared to control cells, as well as changes in other cellular phenotypes such as adipocyte differentiation through information on lipid droplets. Literature review of identified genes allowed insight into possible mechanisms of action implicating these genes in GLUT4 trafficking.

Chapter 4 focuses on the identification of metabolomic signatures of T2D genetic risk using genetic risk scores for T2D and T2D endotypes in the EPIC-Norfolk study. I characterised the metabolomic signature of these risk scores and identify metabolites that overlap with those associated with T2D incidence. I present a framework to assess the causality of these metabolites in T2D using causal mediation analyses, Mendelian randomisation and genetic colocalisation.

Chapter 5 describes the assessment of the role of melanocortin 3 receptor (MC3R) in human physiology through the combination of *in vitro* variant characterisation with

population genetics analyses to identify phenotypic associations with truly damaging missense variants in MC3R in UK biobank.

The concluding discussion (Chapter 6) gives an overview of findings in this thesis and evaluates the strengths and limitations of the approaches used. The implications of this research are highlighted along with discussion of potential future research.

Chapter 1: Overview of the literature

Metabolic disease, and genetically driven approaches to understand aetiology

Overview of chapter 1

Metabolic diseases including type 2 diabetes (T2D) and obesity are a global health concern, affecting an ever-increasing number of individuals worldwide. These common metabolic diseases are multifaceted in aetiology, and disease risk is contributed by a wide range of endogenous and exogenous factors, including genetics, diet, and lifestyle.

In this chapter, I first provide an overview of common metabolic diseases – T2D and obesity, and key pathways and mechanisms contributing to their pathophysiology. I outline the pathways involved in the maintenance of glucose homeostasis, at the tissue and molecular level, and describe how this is dysregulated in insulin resistance.

I next describe the use of genetics to further the understanding of the aetiology of metabolic disease, with a focus on T2D. Here, I outline the current understanding of the genetic architecture of T2D, in both the common and rare variant space, and describe how intermediate traits (e.g., glycaemic traits), 'omics layers', and functional information from *in vitro* studies can be used to understand the mechanisms behind T2D associated loci, and reveal pathways and genes contributing to disease risk.

1.1 Background

Metabolic diseases, including obesity and T2D, are some of the biggest health and economic burdens worldwide. Focusing on diabetes, in 2021 the International Diabetes Federation estimated that approximately 537 million individuals, 1 in 10 adults, are diabetic globally, 90% of which (483 million) have T2D.[1] By 2045 the number of individuals diagnosed with T2D is estimated to rise to approximately 705 million.[1] There is great overlap and a complex interplay between common metabolic disorders. For instance, overweight and obesity are major risk factors for T2D onset. Globally, the World Obesity Federation estimates that there will be more than 1 billion individuals living with obesity by 2030, with the largest increase in prevalence expected in low and middle income countries[2]. In Europe, it is predicted that approximately 1 in 3 individuals will be living with obesity by 2030.[2] Therefore, research aimed at furthering the understanding of the aetiology of these metabolic diseases is warranted and necessary to increase the understanding of disease pathology and identify novel candidates for pharmacological intervention.

Type 2 diabetes

T2D is defined by chronically elevated blood glucose levels, and while often diagnosed by a single diagnostic criterion (random plasma glucose > 11.1 mmol/L or fasting plasma glucose > 7.0 mmol/L, or HbA1C > 6.5%).[3] T2D is multifactorial and is influenced by genetic and environmental factors.[4] T2D is contributed by dysfunction of several pathways, which further influences susceptibility to a host of downstream disease complications.[4]

Obesity

Obesity is defined as a condition where long-term energy intake exceeds energy demands, resulting in an accumulation of excess fat. Individuals are classified as overweight with a body mass index (BMI; Kg/M^2) > 25 or as obese with a BMI > 30.[5] Like T2D, obesity is contributed to both genetic and environmental factors.

Obesity is associated with several complications, including an increased risk of diseases such as T2D, illustrated by the fact that 90% of patients with T2D being obese or overweight.[2] Obesity is causal for insulin resistance; however, it should be noted that while most individuals with T2D are obese, not all obese individuals with insulin

resistance progress to being diagnosed with T2D. Therefore, research which seeks to understand the maintenance of glucose homeostasis is highly relevant to both T2D and obesity.

1.2 Normal physiology: Maintenance of whole-body glucose homeostasis

Insulin is the key player when considering the orchestration of glucose homeostasis. The liver, adipose, and muscle tissues are key sites of insulin action in the fasting and post-prandial states[6], and are effected in insulin resistance (**Figure 1.1**). [7]

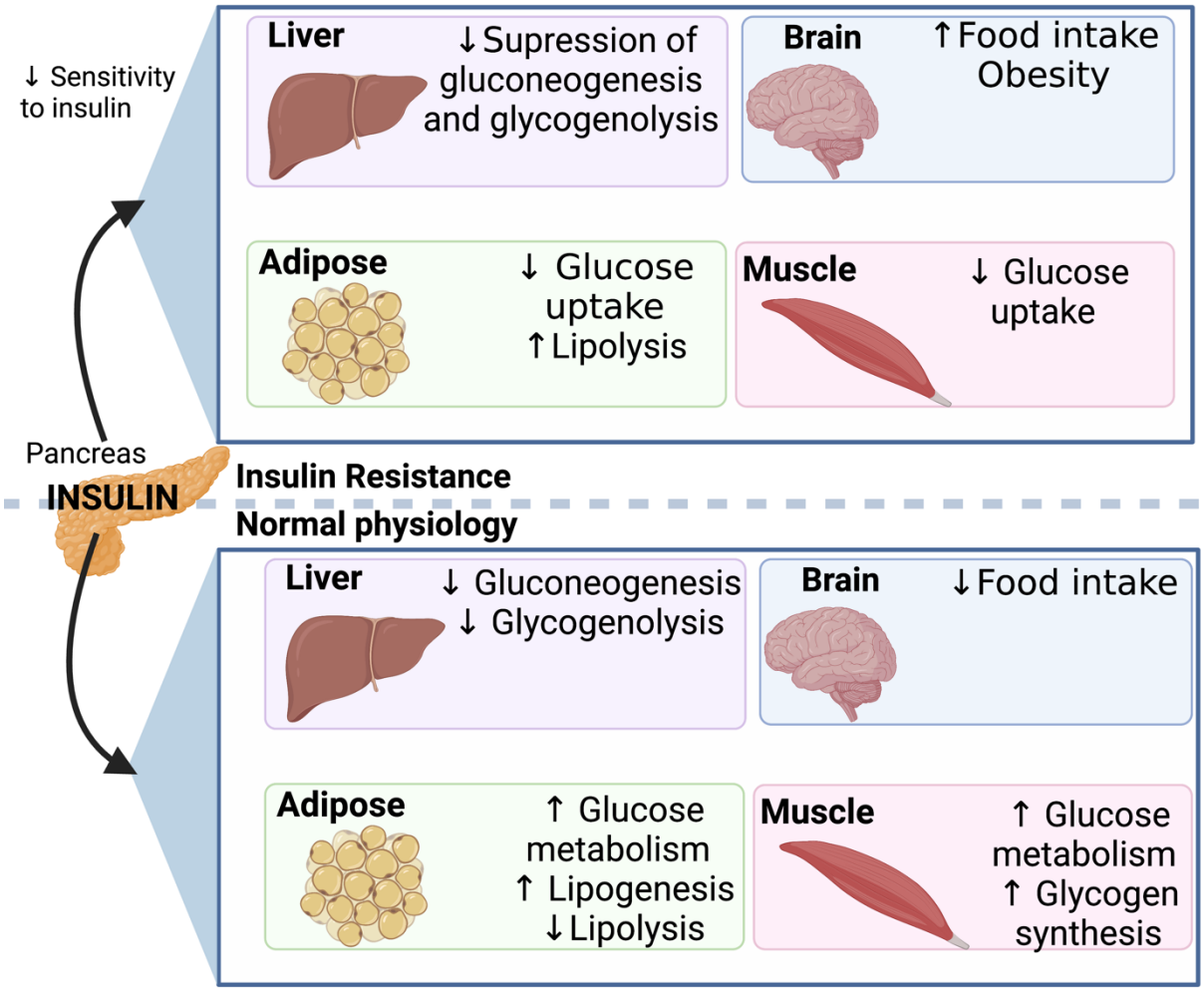


Figure 1.1 Tissue specific responses to insulin in healthy and insulin resistant states in key tissues of insulin action.

Canonical insulin signalling pathway

Insulin binding to the insulin receptor (IR) is the first stage in downstream signal transduction, which results in phosphorylation and activation of insulin-responsive substrate 1 or 2 (IRS1, IRS2). There are two downstream pathways, the first being the promotion of cellular proliferation via the GRB2/MAPK pathway, which will not be

discussed further in this introduction.[5] The second is the promotion of metabolic-related pathways through the activation of the downstream PI3K/AKT pathway.[8–10] The PI3K/AKT pathway activates several downstream processes through the amplification of signal transduction into various intracellular processes to promote glucose homeostasis through increased glucose uptake (via TBC1D4[11]), reduction of gluconeogenesis (via FOXO[11]), and increased glycogen synthesis (through GSK3[12,13]) (**Figure 1.1**). This pathway also promotes protein biosynthesis (via mTORC1[14]) and increased lipid synthesis (via mTORC1[14] and FOXO[15,16]) **(Figure 1.2**). These downstream effects depend on the tissue of action. Activation of the PI3K/AKT arm of the insulin signalling pathway further promotes the non-metabolic pathways of cell survival and prevents apoptosis (reviewed in [17]).

The role of insulin in promoting downstream control of glucose uptake, gluconeogenesis and glycogen synthesis is key in the maintenance of glucose homeostasis.

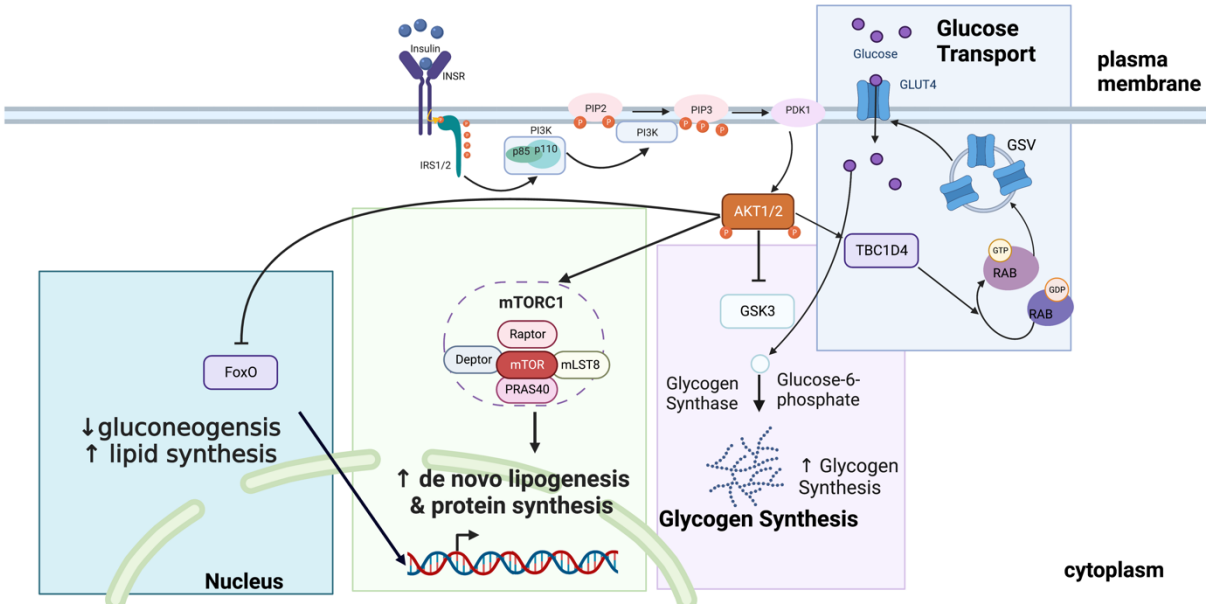


Figure 1.2 Downstream effects of proximal insulin signalling in metabolic control.

Different signalling arms downstream insulin signalling and their effects on metabolism. These effects depend on the respective tissues of action. Figure created using biorender.com.

The liver

In hepatocytes, glycogenolysis is the main source of production and release of glucose in short-term fasting periods.[18,19] In extended periods of fasting, gluconeogenesis becomes the main source of glucose production.[18]

In the presence of insulin, hepatic gluconeogenesis is suppressed, and glycogenesis is promoted. This glycogenesis is activated via the insulin signalling pathway, where AKT is involved in the inactivation of glycogen synthase kinase-3 (GSK-3), a serine/threonine protein kinase.[12] This results in the activation of the rate-limiting enzyme glycogen synthase, which catalyses glycogen chain elongation through the incorporation of UDP-glucose derived from intracellular glucose.[13]

Insulin-mediated suppression of gluconeogenesis results from the phosphorylation of transcription factors by active AKT, which is downstream of insulin receptor activation by insulin. This includes FOXO1, whose phosphorylation by AKT disrupts the interaction of FOXO1 with PGC-1alpha.[20–22] This prevents FOXO1 from entering the nucleus and therefore prevents the activation of transcriptional programs that promote gluconeogenesis.[21,22] There is further indirect action of insulin to results in suppression of gluconeogenesis via the suppression of release of gluconeogenic precursors from adipose (free fatty acids) and muscle (e.g. amino acids).[23,24]

The liver is further the primary organ responsible for insulin clearance, clearing approximately 50% of endogenous insulin.[23,24] After insulin is secreted from the beta cells of the pancreatic islets, it is transported to the portal vein, thus hepatocytes are exposed to much higher levels of insulin than in the general circulation.[25,26]

Peripheral tissues: Adipose and Muscle

In the fasting state, adipocytes in adipose tissue increase lipolysis and release free fatty acids and glycerol. Contrastingly, in the presence of insulin, lipolysis is suppressed, and glucose uptake is promoted. Similarly, in muscle, insulin acts to promote glucose uptake in the postprandial state when circulating insulin levels are high.

Insulin action on muscle

Muscles, including skeletal and cardiac muscle, are a key site of insulin-dependent glucose uptake.[27] Skeletal muscle is the key tissue involved in post-prandial glucose disposal, and hyperinsulinemic-euglycaemic clamp studies in humans have demonstrated that 70-80% of circulating glucose in the post-prandial state is taken up

by this tissue, which is paramount for whole body glycaemic control.[28] Glucose uptake in skeletal muscle is also stimulated by muscle contraction during exercise, independently of insulin signalling.[29]

Upon stimulation with insulin, or exercise, the glucose transporter GLUT4 is translocated to the cell surface in GLUT4 secretory vesicles (GSVs) and glucose is taken up into muscle cells.[27] Here, glucose is utilised as an energy source through glycolysis, or stored as glycogen where insulin activates the AKT/GSK-3/GS pathway, promoting glycogen synthesis. Skeletal muscle acts as the primary site of glycogen storage in the body and therefore this is a major path for glucose disposal.[30] In the liver, where glycogen stores are used to maintain blood glucose levels during fasting catalysed by glucose-6-phosphatase. However in skeletal muscle glycogen acts as an energy source during exercise and is therefore not released, as glucose, into the circulation.[31.] Instead, muscle releases lactate during exercise which can be used as a gluconeogenic precursor by the liver.[24]

Cardiac muscle is primarily fuelled by fatty acid oxidation under healthy conditions, however, when the heart is under stress, in anoxic conditions, glucose becomes the primary fuel for the heart to maintain energy production.[32,33] Glucose is predominantly taken up in cardiomyocytes via GLUT4.[34]

Insulin action on adipose tissue

As in skeletal muscle, insulin stimulates glucose uptake into adipocytes in the postprandial state via GLUT4, with GLUT4 being the predominant glucose transporter expressed in adipose tissue. However, unlike the liver and muscle, where most of the glucose taken up from the circulation is stored as glycogen[19,35], in adipocytes most of the glucose taken up is eventually stored as triglycerides through *de novo* lipogenesis[36], or converted to lactate through the glycolysis pathway, which produces ATP.[30] Less than 5% of glucose taken up into adipocytes is stored as glycogen.[37] The promotion of *de novo* lipogenesis occurs when glucose activates the transcription factors chREBP and SREBP1c.[19,35,38]

Insulin stimulation of adipose tissue also indirectly affects liver glucose homeostasis.[30] Upon insulin stimulation, lipolysis and therefore the release of free fatty acids (FFA) are suppressed in adipocytes. This decrease in circulating FFA availability enhances the suppression of glycogenolysis in the liver.[30]

The Brain: A Link Between T2D and Obesity?

In the brain, insulin action allows modulation of peripheral metabolism including regulation of appetite and therefore food intake.[39] When insulin binds to the insulin receptor in insulin-sensitive neurons, such as proopiomelanocortin (POMC) and Agouti-related peptide (AgRP) neurons in the arcuate nucleus of the hypothalamus, this activates pathways that allow the brain to respond to peripheral nutritional status, such as changes in food intake, and in peripheral metabolism.[40,41]

An example of such pathways downstream of insulin is the central melanocortin system.[42] This pathway is activated by peripheral nutritional cues including insulin and leptin. The downstream effects of this pathway and key players are described in Chapter 5, which focuses on the results of the genetic and functional characterisation of Melanocortin 3 Receptor (*MC3R*).

Other tissues and their role in glucose homeostasis

Several other organs play an important role in the maintenance of glucose homeostasis. As a key focus of this thesis is insulin resistance in peripheral tissues, the role of these other systems will not be widely discussed. These include the critical role of pancreatic beta cells in insulin secretion in response to increased blood glucose levels.[43] Other insulin-responsive tissues play a role in glucose homeostasis, including the vasculature, where insulin increases vasodilation and therefore improves perfusion[44], and the kidney, which is involved in insulin clearance, gluconeogenesis, and glucose uptake.[45,46]

1.3 Molecular mechanisms in insulin resistance

Insulin resistance is one of the first defects on the path towards T2D onset

Insulin resistance is one of the first defects that is detectable in individuals who are on the path towards T2D development.[4,47] Insulin resistance is characterised by a state in which cells display reduced insulin response, resulting in an impaired ability to reduce blood glucose levels. This is contributed to by several factors including obesity, low levels of physical activity, and genetic risk. Insulin resistance results in a number of metabolic effects that differ depending on the tissue type, and these effects are believed to potentially occur at different time points on the path towards T2D onset.[7,48–50]

Peripherally, insulin resistance affects muscle and adipose tissue which results in reduced glucose uptake, and thus hyperglycaemia.[27,51] This primarily is due to decreased insulin sensitivity in the post-prandial state, which is thought to be one of the primary defects on the path towards T2D onset.[47] Additionally, in adipose, insulin resistance results in effects on lipid metabolism through reduced suppression of lipolysis and therefore elevated plasma free fatty acids.[36,51,52] This results in effects on other tissues through ectopic lipid accumulation, which in turn further exacerbates the insulin resistance state.[36]

In obesity, increased adipose tissue mass results in increased release of free, non-esterified fatty acids (NEFA), glycerol, pro-inflammatory cytokines (e.g. IL-6 and TNF-alpha), and adipocyte-derived hormones, including leptin and adiponectin (reviewed in [53]). For example, increased NEFA release from fat is associated with obesity and T2D in humans, with increased circulating NEFAs resulting in both hepatic and peripheral (adipose and muscle) insulin resistance.[54]

The link between insulin resistance and obesity can also be explained independent of adiposity, through the influence of insulin on central control of energy expenditure and food intake via the central melanocortin system, and vice versa.[55–57] Further, ectopic lipid deposition such as in the liver and muscle in obesity can occur due to overnutrition and result in loss of insulin sensitivity in these tissues.[58–61]

In the liver, reduced sensitivity to insulin can results in impaired suppression of gluconeogenesis and glycogenolysis through decreased sensitivity of the insulin receptor to amplify signal transduction pathways or through changes in transcriptional regulation such as those induced by epigenetic alterations (reviewed in[62]).

The pathophysiological sequence of T2D onset

Whilst it is known that insulin resistance or beta cell dysfunction are key players in T2D pathogenesis, whether either are the primary initiators towards T2D onset has not been fully agreed.[63,64] Since T2D is highly heterogeneous it is possible the primary contributor could vary on an individual level, with baseline changes in both insulin secretion or sensitivity evident in longitudinal studies, prior to T2D onset.[65] There are two main pathophysiological sequence leading T2D onset that have been proposed involving both processes (**Figure 1.3**):

1. Insulin resistance results in pancreatic beta cell dysfunction and insulin hypersecretion, causing exhaustion of the pancreatic beta cells.[66–69]
2. Beta cell dysfunction occurs early on prior to impairment of glucose tolerance, and this results in T2D on a background of insulin resistance.[70,71]

Chronic excess of energy intake over energy expenditure, resulting in obesity, is strongly causal for insulin resistance. Impaired ability to store fat safely in adipose tissue greatly increases the risk of overnutrition and thus insulin resistance.[60] This insulin resistance will necessitate a chronic increase pancreatic beta cell effort. GWAS studies of T2D are enriched in loci implicating the pancreatic beta cells, with beta-cell dysfunction being essential in the progression to T2D.[72,73] It is possible that in individuals who are genetically protected from beta cell dysfunction, or have high functioning beta cells, will be able to maintain insulin hypersecretion over their lifetime, even in the background of stress from insulin resistance. However, those who have genetic predisposition to dysfunctional pancreatic beta cells will get diabetes. When insulin resistance occurs in these individuals they will likely get earlier onset of T2D if beta-cell stress is severe.[74]

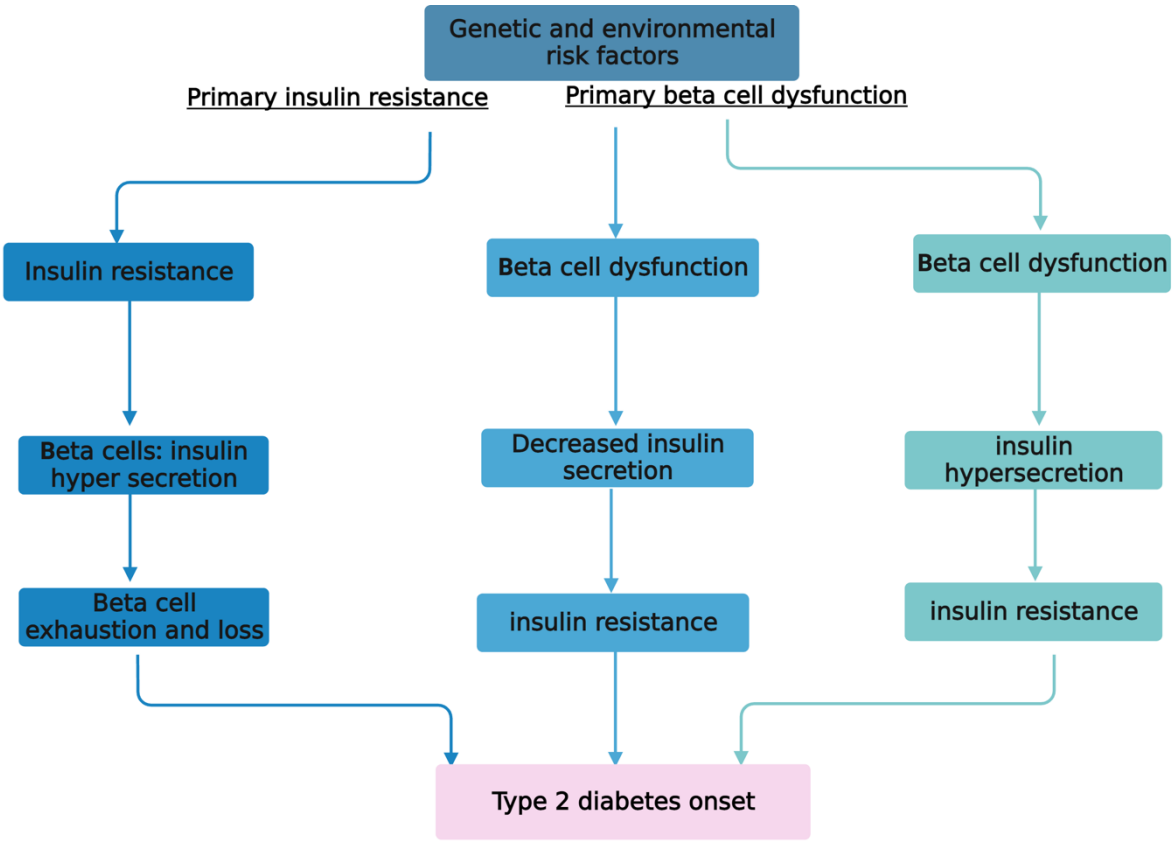


Figure 1.3: The pathophysiological sequence of insulin resistance and beta cell dysfunction in T2D onset.

1.4 Genetic approaches to understand the aetiology of metabolic disease

Human genetic studies have accelerated insights into metabolic disease aetiology. This has been aided by large-scale genome-wide association studies (GWAS), which have enabled the detection of an exponential number of common genetic variants influencing the disease risk, compared to earlier small-scale family-based linkage studies.

Leveraging genetic information allows investigation of disease aetiology in a conceptual framework that minimises the influence of confounding factors (e.g., diet) that affect observational studies of disease. Advantages include the fact that genetic variants are fixed at the point of conception and therefore avoid the influence of reverse causation, i.e., they always precede the exposure of interest. This is particularly important when the interest is on chronic conditions that develop over a longer time frame, where pathophysiological changes may precede the onset of clinically overt symptoms by years. The integration of disease genetic associations with those of other traits has allowed deep molecular characterisation of genetic loci, uncovered novel pathways and risk factors of interest which influence disease aetiology. This has aided the acceleration of drug development, with additional methods such as Mendelian randomisation (MR, see below) which allows approximating of a randomised control trial, leveraging genetic data at the population scale.[75,76] Such methods have been used to prioritise and refine potential targets for further investigation.

Genetic studies examining clinical measures of glycaemia including T2D, HbA1c and fasting and random glucose, all allow the identification of genetic variation influencing processes covering both insulin secretion and insulin action. Additional measures allow further disentanglement of insulin secretion (e.g. HOMA B) and insulin resistance (e.g. fasting insulin, HOMA IR) (reviewed in [77]). Therefore, the identification of variation influencing such processes in genetic studies is highly determined by the physiological measures examined and the processes these reflect.

1.4.1 The genetic architecture of type 2 diabetes

Common genetics: Genome-wide association studies

GWAS for T2D have grown in scale over almost two decades with meta-analysis efforts that allow the combined effort of tens of population studies, greatly increasing the power to detect associations at genome-wide significance. GWAS efforts superseded

linkage efforts that had limited success in determining genomic regions associated with T2D. The first GWAS for T2D, published in 2007, identified 4 loci in a total of 1,363 T2D cases and controls.[78] These loci included *TCF7L2,* which has remained the strongest associated genetic signal for T2D to date.[79,80] GWAS remain dominated by European (Caucasian) samples, and therefore recent GWAS efforts for T2D have aimed to expand analyses to include individuals of non-European ancestry.[80,81]

The most recent multi-ancestry GWAS effort from the DIAMANTE consortium includes 180,834 T2D cases and 1,159,055 controls across 5 ancestries.[80] Whilst European samples still dominate the overall sample size, making up 51.1% of individuals, this effort also included individuals of East Asian (28.4%), South Asian (8.3%), African American (6.6%) and Hispanic (5.6%) ancestry. This effort identified a total of 277 loci with T2D at the standard genome-wide significance threshold ($P < 5x10^{-8}$), 237 of which remained significant with a more stringent multi-ancestry genome-wide significance threshold ($P < 5x10^{-9}$). This more stringent significance threshold, accounts for the estimated number of independent SNVs across 5 ancestry groups using haplotypes from 1000 Genomes Project reference panel). A total of 338 distinct genetic signals were identified at these 237 loci using conditional analysis, which allows the identification of independent secondary association signals at a given locus.[82]

With the introduction of better imputation panels, GWAS have been able to identify associations with variants at low frequency as well as variants which are common in the population.[83] Further increased diversity of ancestries in meta-analyses effort have allowed improved resolution of fine mapping, with a lower number of candidate causal variants in the 99% credible sets from multi-ancestry analyses compared to those with European-only samples[84,85]. In the case of Mahajan *et al* (2022)[80], this allowed the number of variants in the 99% credible set to be cut by approximately 66% on average. However, despite this, the identification of causal genes and molecular mechanisms explaining why most of these loci are associated with a trait remains a challenge.

1.4.2 Genetic architecture of intermediate traits: glycaemic traits and T2D

The integration of associations with intermediate quantitative traits has allowed further insight into the mechanisms behind T2D disease associations. These efforts include

those studying the genetic architecture of glycaemic traits, which have been spearheaded by the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC). MAGIC is a consortium that aims to combine GWAS from multiple studies for glycaemic traits such as plasma insulin and glucose in individuals without T2D, thus allowing greater power to detect associations compared to looking at these studies in isolation. Identifying loci that overlap between T2D and glycaemic traits can greatly aid in furthering the understanding of the mechanism behind disease associations at a given locus and how these intermediate traits contribute to disease risk.[86]

Fasting traits that have been examined in such efforts include fasting insulin and glucose. Glycated haemoglobin (HbA1c), which indicates long-term glycaemic control has also been assessed. The most recent meta-analysis effort from MAGIC included genetic discovery for traits used in T2D diagnosis: fasting glucose, HbA1c and 2hr glucose, as well as fasting insulin, across up to 281,416 individuals without diabetes, representing 5 ancestry groups (70% European ancestry).[87] Here, they identified 235 loci in multi-ancestry analysis across all four traits and 7 in single ancestry analyses. Of these 242 loci, 114 were associated with T2D at a suggestive significance threshold of $P < 1x10^{-4}$.

Glycated haemoglobin (HbA1c)

HbA1C represents the average blood glucose level over a ~3 month lifespan of a red blood cell (RBC), it is unaffected by day-to-day variability in glucose levels[87], and hence is often the first line diagnostic test for T2D and prediabetes as this can be performed at any point of the day in non-fasting conditions.[88] The genetic architecture for HbA1c is reflective of both blood glucose control as well as genetic variants that influence the red blood cells directly for example through effect on RBC turnover. It has been shown that genetic variants are associated with HbA1c due to their effect on RBC traits do not predict T2D genetic risk, and therefore only a subset of the variants identified to be associated with HbA1c are likely of interest in understanding T2D pathogenesis.[89–91]

Fasting and random glucose

Fasting and random blood glucose are also used clinically in T2D diagnosis, with random glucose having the advantage of being able to be performed without prior notice of fasting. Random glucose is also more widely available in population cohorts

compared to fasting, with this being available in large population biobanks, including UK Biobank. Random glucose > 11.1 mmol/L or fasting glucose > 7.0 mmol/l is used for the diagnosis of diabetes.[92]

The most recent trans-ancestry GWAS for fasting glucose (N = 281,416) identified 102 loci, enriched in regions of pancreatic islet stretch enhancers and enriched for genes expressed in the pancreas.[87] A meta-analysis of GWAS for random glucose in 493,036 individuals allowed the identification of 162 distinct genetic signals at 128 loci, 77 of which had not previously been reported for glycaemic traits.[93] In this analysis time since last meal was accounted for, and authors demonstrated that random glucose is highly correlated genetically with fasting glucose (genome-wide genetic correlation (rg) = 0.88).[93] This study demonstrated that the genetic architecture of random glucose, whilst reflecting a wider range of physiological states than fasting alone, is reflective of fasting glucose.[93] Therefore, combined efforts across this more commonly measured trait and fasting glucose could be considered.

Fasting insulin

Fasting insulin is not measured routinely in the clinic but can provide important biological insights into the pathophysiology of T2D and insulin resistance. The same trans-ancestry effort described above for fasting glucose and HbA1c identified 48 loci (95 independent signals) significantly associated with fasting insulin.[87] Furthermore, these loci were significantly enriched in skeletal muscle and adipose stretch enhancers, with nominal enrichment in liver stretch enhancers.[87] Of the 95 signals, 41 were at least suggestively associated with T2D ($P < 1x10^{-4}$), suggesting that these T2D loci may contribute to T2D risk through an effect on fasting insulin resistance.[87]

Measures during an oral glucose tolerance test (OGTT)

An oral glucose tolerance test (OGTT) can be used to assess insulin sensitivity in individuals who show symptoms of T2D but have normal fasting glucose, and to diagnose gestational diabetes. The standard OGTT involves an overnight fast and obtaining a fasting blood sample, followed by the consumption of a 75g glucose drink (the challenge). Blood samples can then be obtained at various timepoints throughout the OGTT to assess insulin secretion and insulin sensitivity in response to the glucose

challenge. Much focus of genetic studies of OGTT traits have been on blood glucose levels after a glucose challenge, with fewer studies measuring insulin throughout the course of an OGTT. At 2 hours after the challenge, blood glucose levels are believed to reflect insulin resistance, while samples taken earlier in the OGTT, for example at 30 min, are thought to be more reflective of defects in insulin secretion and therefore impaired pancreatic beta cell function. Indices utilising OGTT measures of glucose and insulin have further been developed to approximate hyperinsulinemic-euglycaemic clamp measures of insulin sensitivity as these are more easily obtainable in large population studies (see chapter 2 for further discussion of OGTT indices and their genetic architecture).

However, the sample size for studies that measure insulin during an OGTT is typically lower for genetic discoveries (e.g. in Chen et al (2021) fasting glucose N = 281,416 vs 2h glucose N = 85,916).[87] The most recent study of 2 hour glucose identified a total of 28 signals at 21 loci in a meta-analysis of 41 cohorts, 13 of which overlapped with genome-wide T2D loci.[87] Studies of insulin secretion indices have also identified genetic loci that are thought to be involved in beta cell dysfunction.[94]

Other glycaemic traits

Other genetic discovery efforts have reported genetic associations with other glycaemic traits including proinsulin[95] and glucagon[96], which provide insights into the genetic contribution of insulin processing and hepatic glucose production, respectively.

1.4.3 Rare and low frequency variants and T2D risk

The heritability of T2D is estimated to be 30-70% from twin and family-based studies.[97–100] With the SNP-based heritability from GWAS for T2D now approximated to reach 19%, further genetic variation contributing to T2D remains to be uncovered.[101] This missing heritability will be identified with an increased sample size of common genetic discovery efforts, and through fine-scale whole exome (WES) and whole genome sequencing (WGS) efforts which allow better interrogation of the common and rare variants not covered by SNP arrays and imputation (**Figure 1.4**).[102,103]

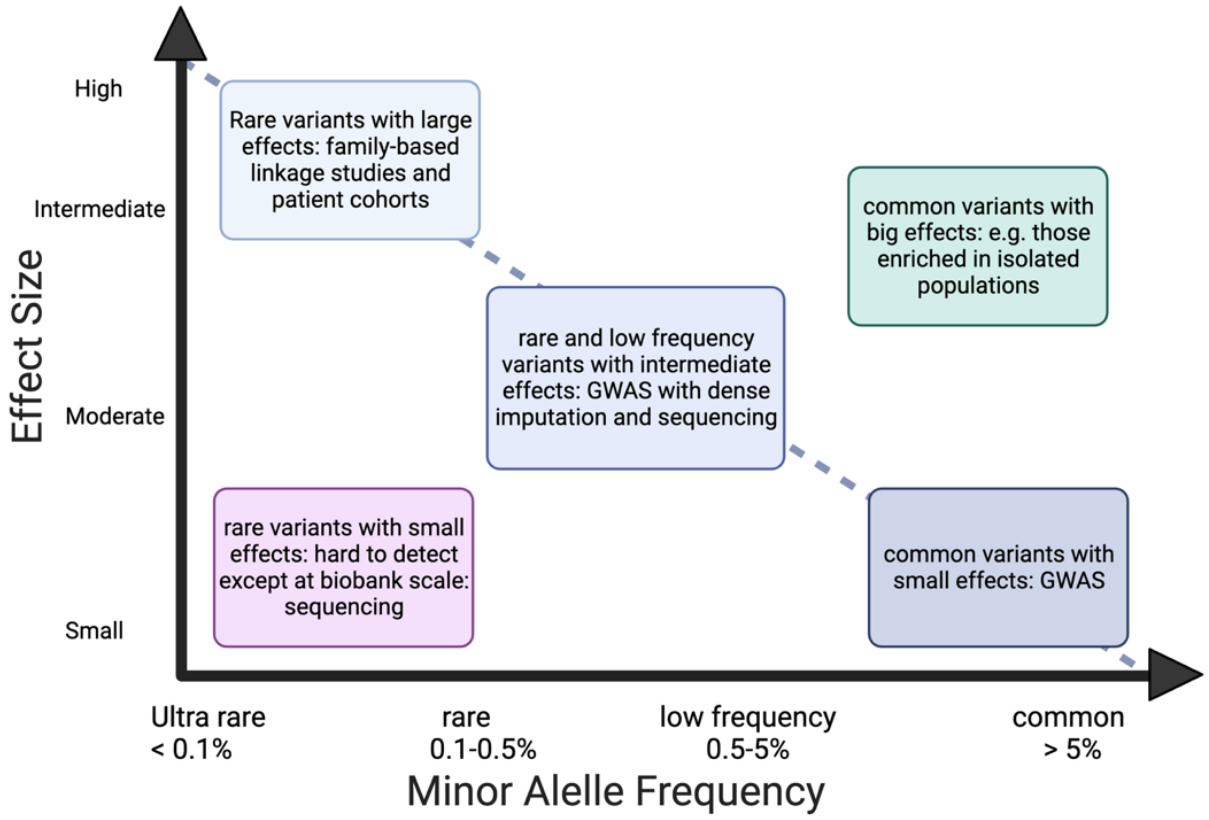


Figure 1.4 Relative effect size of variants at different allele frequencies and their primary method of discovery.

Like with GWAS T2D risk loci, rare variants contributing to T2D risk can broadly be split into those that affect genes involved in insulin secretion (*SLC30A8*[104], *MTNR1B*[105]) and insulin action (e.g., *AKT2*[106], *INSR*[107], *TBC1D4*[108]). There are several examples in which highly penetrant variants associated with T2D have been identified in genetically isolated populations (geographically or culturally). These population isolates are advantageous for studying the genetics of T2D, particularly where genetic drift results in variants at a high frequency that are rare or not found in the general population. Further, longer stretches of linkage disequilibrium, attributable to fewer variants mean that genetic heterogeneity is greatly reduced in these populations.

Such examples are coding variants in *TBC1D4*, a RAB-GTPase which is an important player in insulin-stimulated glucose uptake via GLUT4. These coding variants in *TBC1D4* have been identified in both isolated populations and patient cohorts. In the Greenlandic Inuit, a population that is isolated from the general outbred population, carries a common truncating variant (rs61736969; *TBC1D4* p.Arg684*ter*; MAF =

16.8%), which is associated with a 10.3 fold increase in T2D risk in homozygous carriers.[109] This variant is not detected in outbred populations globally.[110] The variant is estimated to be responsible for 10% of all T2D cases in Greenland and other isolated Arctic Inuit populations.[96,109,111]

It is thought that this variant, which is associated with post-prandial hyperglycaemia, may have been selected by positive selection for survival in the extreme artic environment where this population lives. The mutation could be seen as advantageous for survival with the traditional low carbohydrate diet of the Inuit population.[112] This variant is specifically associated with higher insulin and glucose levels 2hr after an OGTT in homozygous carriers, indicating post-prandial specific insulin resistance.[109,111] This suggests that there is value for screening for T2D by OGTT measures of glucose in these populations, with fasting measures likely to lead to an under diagnosis.[112] Further experiments have identified that *TBC1D4* p.Arg684*Ter* specifically affects the muscle-specific long isoform of *TBC1D4* and therefore this postprandial insulin resistance is specific to this tissue, the primary site of postprandial glucose uptake.[109] *TBC1D4* p.Arg684*Ter* results in reduced skeletal muscle GLUT4 content, the primary insulin-stimulated glucose transporter in the post-prandial state.[109] A study of gene-lifestyle interaction in Greenlandic Inuit demonstrated that increased physical activity is beneficial in reducing postprandial hyperglycaemia in carriers of *TBC1D4* p.Arg684*Ter*, and there was suggestive evidence that this physical activity reduces the risk of T2D.[112]

Another truncating variant in *TBC1D4* (p.Arg363*Ter*), characterised to be dominant negative, was first reported to specifically result in post-prandial hyperglycaemia and hyperinsulinemia in an individual heterozygous patient with severe insulin resistance.[108] Dominant negativity is defined as when the mutant protein disrupts the activity of the wildtype protein when expressed in heterozygosity. *TBC1D4* p.Arg363*Ter* was characterised in an *in vitro* overexpression system in 3T3-L1 adipocytes where it reduced GLUT4 at the plasma membrane in in the insulin-stimulated state.[108] Other missense variants in *TBC1D4* have also been identified in a patient cohort of severe insulin resistance and were associated with higher post-

prandial glucose insulin and glucose.[113] Unlike the truncating variant in Greenlandic Inuit, these do not appear to have muscle-specific effects.[113]

Like *TBC1D4*, mutations that implicate *AKT2* in postprandial insulin resistance were first identified in patients, before being identified in more isolated populations.[106,114] *AKT2* encodes a protein kinase that plays a key role in insulin signalling. In a single proband with early onset T2D, *AKT2* p.Arg274His was identified to result in severe insulin resistance with elevated fasting and post-prandial insulin as well as lipodystrophy.[106] *In vitro*, functional studies demonstrated that *AKT2* p.Arg274His, which lies within the kinase domain, has a dominant negative effect. This variant was also associated with lipodystrophy with reduced total body fat.[106] Other missense variants in *AKT2* have been identified in the Finnish population, which is more genetically isolated than other European populations.[115–118] *AKT2* p.Pro50Thr is found in 1.1% of individuals of Finnish ancestry and results in increased fasting and post-prandial insulin as well as a reduction in glucose uptake in insulin-responsive tissues, including muscle myotubes.[114,119,120]

More proximal insulin signalling mutations have also been identified to result in severe insulin resistance including insulin receptor (*INSR*), where mutations result in fasting hyperglycaemia, postprandial hyperglycaemia, and diabetes.[121,122]

There have also been several WES efforts for T2D. The T2D-GENES consortium identified 4 genes (*PAM, MC4R, UBE2NL and SLC30A80)* in which the burden of rare coding variants was associated at exome-wide significance in a multi-ancestry study of 20,791 individuals with T2D and 24,440 non-diabetic controls.[123] Predicted deleterious variants in these genes are associated with increased risk (*PAM, MC4R, UBE2NL)* and protection against T2D (*SLC30A80*). *PAM, MC4R*, and *SLC30A80* were identified in common variant GWAS before this study and contain low-frequency coding variants that have been previously reported for T2D.[123] In the UK Biobank, which is the largest widely accessible cohort with WES, burden of rare coding variants was identified to be associated with T2D. Examples of novel genes include *GIGYF1* and *TNRC6. GCK* and *HNF1A,* which are two genes where mutations are known to

cause a monogenic form of diabetes – maturity onset diabetes of the young (MODY), were also reported.[124–130]

Whilst some of the genes involved in large-scale WES studies such as those for T2D overlap with common genetic architecture it has been reported that there is limited overlap at the gene-level of rare and common genetic architecture in disease, with different genes, albeit implicated in similar biological processes being identified.[131] This suggests that there is value for continued genetic discovery in the rare and common variant space.

1.5 The genetic and phenotypic heterogeneity of T2D

Despite being diagnosed with a single diagnostic criterion – chronic hyperglycaemia – it is accepted that T2D is a heterogeneous disease which can result from multiple different pathways and mechanisms and have different characteristics in terms of progression, drug response and complications.[132,133] Therefore there have been efforts to use genetic and nongenetic information to identify subtypes of T2D, with the aim to identify personalised treatment strategies for patients.[134–136] Whilst it is generally accepted that the majority of individuals will have a mix of these different subtypes, rather than conform to an individual risk group[137], such profiling does identify individuals for which a particular T2D sub-type dominates and therefore may benefit from more targeted treatment strategies.

Phenotypic sub-types in T2D

Observationally, clustering efforts have focused on utilising multiple clinical measures to identify clusters of patients with differing disease pathology and progression. Ahlqvist and colleagues defined five subtypes of adult-onset diabetes, four of which reflect T2D, using six clinical measures[135] (**Table 1.1**). These clusters have been replicated in multiple independent cohorts and across different ethnic groups.[138–142] However other studies have also revealed alternative clustering of subtypes.[143,144] Similar phenotypic clustering methods have been used to stratify individuals without diabetes in different risk strata.[145] Further, these subtypes show different signatures of T2D genetic risk contributing to them[145].

Table 1.1: Five phenotypic subtypes of adult onset diabetes defined by Ahlqvist *et al* (2018)[135]

Reflects	Diabetes Cluster	Defining phenotypes	Proportion of ANDIS patients included in clustering (N total = 8980) [135]
T1D or latent autoimmune diabetes of the adult (LADA)	severe autoimmune diabetes (SAID)	Autoimmune component, GADA (glutamic acid decarboxylase antibody) positive, with insulin secretion defects. Early onset, relatively low BMI, insulin deficiency. Poor metabolic control.	6.4%
T2D	severe insulin deficient diabetes (SIDD)	insulin secretion defects, in the absence of GADA. Early age of onset, low BMI, low insulin secretion. Poor metabolic control.	17.5%
	Severe insulin resistant diabetes (SIRD)	high BMI with insulin resistance.	15.3%
	Mild obesity-related diabetes (MOD)	high BMI without insulin resistance. Modest metabolic dysfunction.	21.6%
	mild age-related diabetes (MARD)	typically, older than individuals in the other clusters with a mild T2D phenotype. Modest metabolic dysfunction.	39.1%

Genetic sub-types in T2D

In addition to phenotypic clustering of T2D, several clusters of GWAS loci associated with T2D have been developed to identify individuals at risk of T2D through changes in different pathways.

Udler *et al* (2018) aimed to categorise 94 T2D risk variants (from Morris *et al* , 2012[146]) into groups to identify different clusters of loci with different mechanistic pathways using a soft clustering approach, allowing genetic variants to belong to more than one cluster.[134] This differs from more historic efforts which used hard clustering meaning a

genetic variant can only belong to one group/cluster.[147–150] This revealed five clusters of variants reflecting different pathways and tissues of action, with different clinical associations when assessing the association of weighted genetic risk scores for each of these clusters[134] (**Table 1.2**).

These 5 clusters showed distinct associations with different traits related to T2D complications and related disease including stroke and coronary artery disease.[134] A recent expanded analysis including a larger number of T2D associated variants and 64 traits using the same approach replicated these 5 clusters and identified a further 5 novel clusters. Four of these clusters related to insulin deficiency and therefore defects in insulin secretion, the other 5 to insulin resistance and one with a yet to be identified mechanism.[151]

Table 1.2: Genetic subtypes of Type 2 Diabetes defined by Udler et al (2018)[134]

Reflective of	Genetic sub-score	Associated phenotypes	N SNPs
Reduced beta-cell function	Beta cell	Reduced insulin secretion, higher proinsulin levels.	30
	Proinsulin	Reduced insulin secretion, lower proinsulin levels.	7
Insulin resistance	Obesity	increased overall adiposity and fasting insulin. Obesity mediated insulin resistance.	5
	Lipodystrophy	associated with reduced insulin sensitivity, adiponectin, HDL, and increased triglycerides, as well as increased WHR in female participants. Lipodystrophy-like or fat distribution mediated form of insulin resistance	20
	Liver/lipid	Reduced levels of triglycerides, as well as urate, palmitoleic acid, and linoleic acid. Higher gamma-linolenic acid. 3 of the 5 SNPs in this cluster associated with non-alcoholic fatty liver disease (NAFLD).	5

Other studies have sought to identify genetic loci associated with factors that are known to contribute to T2D. For example, Lotta *et al* (2017) used an integrated genomic approach using several commonly measure hallmarks of insulin resistance– higher fasting insulin adjusted for BMI, lower HDL and higher triglyceride levels.[152] Combining and prioritising loci using this phenotypic signature allowed identification of

53 genetic loci, which when combined in a genetic score associated with higher risk of T2D and lower peripheral fat, suggesting impaired ability to store fat in these compartments.[152]

1.6 From GWAS to function: going beyond genetic associations with disease

To go beyond disease centred GWAS associations, efforts to study intermediate traits have been conducted to allow further mechanistic insight into loci associated with metabolic diseases. This has most extensively been done for T2D, with efforts characterising the genetic architecture of glycaemic traits (see above), as well as additional quantitative traits such as tissue-specific gene expression and 'omics profiling. 'omics measures are becoming increasingly available at the population scale, and scaling of *in vitro* functional assessment allows further understanding into the functional effects of genetic loci and therefore the ability to go from variants to mechanisms.

Integration of 'omics layers

'omics, describes the measurement of a given molecular entity in its entirety. This is perhaps most seen in genomics, which aims to characterise all genetic variation throughout the genome, but also includes transcriptomics, proteomics, and metabolomics.

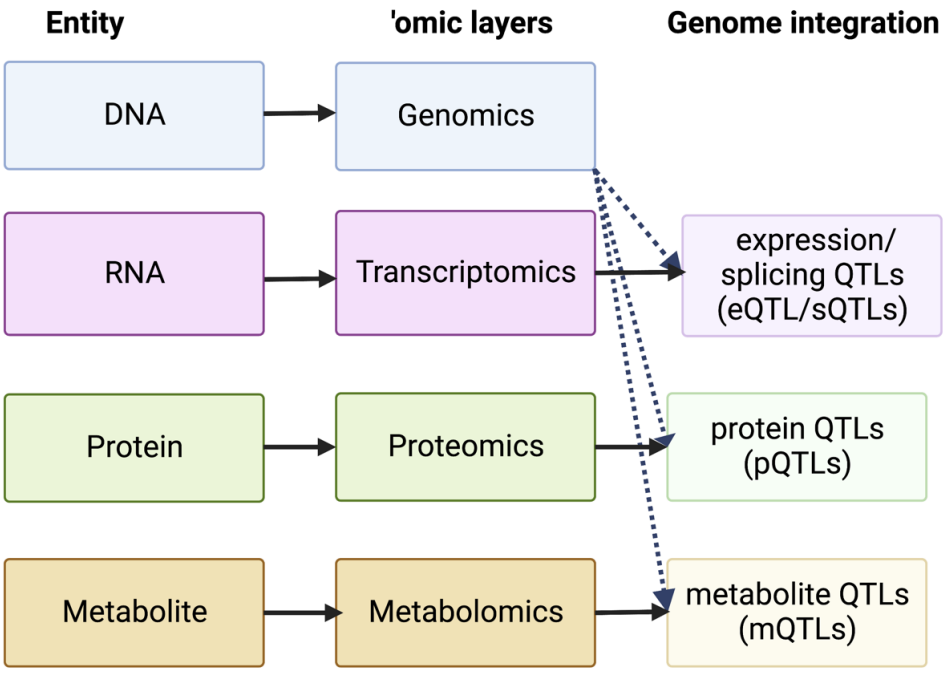


Figure 1.5: integration of genomics with other 'omics measures to identify quantitative trait loci (QTLs).

While in isolation such studies of 'omic layers give insight into biological pathways, these can be combined with genomics and disease-relevant traits and allow insight

into genetic aetiology of diseases at fine scale resolution (**Figure 1.5**). As the vast majority of GWAS hits are non-coding[153], such integration has the potential to allow increased insight into the regulatory impact of genetic variation on expression, protein levels, and small molecules; and further aid identification of candidate causal genes.

The GTEx project describes a population cohort in which genotyping has been combined with transcriptional profiling by RNA sequencing (RNA-seq) of 53 types of post-mortem tissues.[154–156] This has enabled the identification of tissue-specific expression quantitative trait loci (eQTLs) and splicing QTLs – genetic variants that are associated with changes in transcript expression or alternative splicing events leading to different levels of splice isoforms (Splicing QTLs, sQTLs), which allow insight into the regulatory architecture of the human genome, and further insight into genetic variants which influence transcription. The GTEx consortium have further expanded this resource by characterising proteomes of these same individuals across 32 tissue types.[157] More recently, efforts have aimed to identify cell-type specific eQTLs through the integration of genomic data with single-cell RNA-sequencing (scRNA-seq) in a given tissue of interest. Although typically scRNA-seq is not yet done at sufficient scale to identify single cell eQTLs, several methods have been developed to integrate small scale scRNA-seq data with bulk tissue QTL data to allow deconvolution of single-cell expression signatures from bulk RNA-seq.[158–162]

Metabolomics describes the characterisation of small molecules including amino acids, lipids and xenobiotics; and is most commonly is measured in blood samples.[161,162] The metabolome is affected by many endogenous (e.g. genetics) and exogenous factors (e.g. diet).[161,162] In combination with genetics, metabolomics allows the identification of metabolite QTLs (mQTLs) that indicate genetic variants associated with changes in the level of metabolites in a given tissue. These indicate genetically-driven changes in metabolite levels that result from the disruption of different pathways, therefore mQTLs may indicate candidate genes involved in catabolism of these.[163–167] Similarly, the integration of proteomics, the characterisation of proteins in a sample, with genomics has allowed the identification of genetic variants that influence the levels of proteins in a given sample.[168] Cis-pQTLs, protein QTLs that are near the gene encoding the protein measured, enable the prioritisation of candidate causal genes at genetic loci. Further the integration of disease GWAS identify proteins that are commonly implicated across diseases.[169–175]

Mendelian Randomisation and genetic colocalisation

The integration of QTL and disease information to gain mechanistic insights is most frequently done using the statistical approaches such as Mendelian Randomisation and genetic colocalisation. Genetic colocalisation methods aim to identify shared genetic signals for a given pair, or set, of traits, for example QTLs and disease-traits of interest. Different methods allow identification of shared genetic signals where there is a single causal variant[176,177] ,or multiple causal variants at a locus[178].

Mendelian Randomisation (MR) describes the use of genetic associations across different traits to approximate a randomised control trial, to identify if a particular exposure trait or measure is causally associated with an outcome trait.[76,179–181] Unlike genetic colocalisation, MR methods are commonly used to assess the causal relationship between trait pairs across a set of prioritised loci of interest – genetic instruments – but can also be used to assess the causal relationship of traits using a single genetic instrument.[178] Two-sample MR methods utilise summary level GWAS statistics for allowing the study of trait pairs using GWAS statistics from different sources[178], thus allowing the use of the maximum available sample size for each trait.[179,182] Combining both genetic colocalisation and MR allows the identification of trait pairs with a shared genetic architecture and causal relationship, providing greater impetus to prioritise for follow up studies.[178]

What is the causal gene at a genetic locus?

Although the molecular mechanism by which most common genetic loci are associated with T2D remains unknown, several of the identified genetic loci have been attributed to defects in insulin secretion through beta cell dysfunction and to a lesser extent insulin resistance. These mechanistic insights have been gained from a combination of focused genetic analyses in population and clinical cohorts, as well as through *in vitro* and *in vivo* functional characterisation of the predicted causal variant and/or affected gene.[183–185] Most of these studies have focused on a single locus, rather than screening across loci identified. Exome arrays and WES efforts have allowed the identification of coding variants associated with disease that are more readily interpretable in terms of the gene of action.[123,125,130] However, *in vitro* functional characterisation is often needed to understand the true impact of these coding variants

on the protein product, with such analyses relying on in silico prediction of variant impact.[186,187]

To go from non-coding genetic variant to gene is more challenging and integration of traits reflecting genome annotations such as eQTLs (described above), chromatin confirmation and epigenetic signatures can aid identification of the causal gene of interest when looking at such annotations in the context of tissues relevant to disease.[154,156,185,188,189]

A second strategy is to integrate information of intermediate traits, such as glycaemic traits, which may allow the identification of overlapping loci with specific T2D-relevent traits of interest. This could be useful in designing screening strategies to assess the role of genes at these T2D loci on a particular pathway of interest and prioritising the most relevant loci to that pathway in T2D. Such *in vitro* screens can be done in a targeted fashion focusing on a particular locus or candidate gene, this has been done for several loci including *TCF7L2*, *MTNR1B*, and *PPP1R3B*. Pinpointing the causal gene at a locus is challenging and this can be investigated using several lines of evidence.

TCF7L2 (rs7903146) was the first GWAS locus identified for T2D and remains the strongest associated locus to date.[80,81,190] rs7903146 has been shown to effect enhancer activity being located in pancreatic islet-specific open chromatin region with carriage of the risk allele resulting in allele-specific changes in chromatin accessibility and regulation in islets, surprisingly, the risk variant being associated with higher expression of *TCF7L2* in islets.[191,192] It has also been shown that this locus impacts insulin secretion and additionally results in a reduced response of beta cells to incretins such as GLP-1, suggesting that it disrupted incretin-stimulated insulin secretions[192,193], however, such an incretin effect has been debated.[194]

The T2D risk allele at the *MTN1RB* locus (rs10830963) has been associated with increased *MTNR1B* mRNA in human islets[195–199], and associated with increased fasting glucose and reduced beta-cell function.[200] In contrast, the characterisation of rare coding variants in *MTNR1B* has identified decreased melatonin signalling due to the dysfunction of *MTNR1B* resulting in increased T2D risk.[105] Similarly, in human clinical studies there are contradictory results with regards to benefit of melatonin

administration with acute treatment of melatonin resulting in impaired glucose tolerance in studies of non-diabetic[201] and diabetic[199] individuals. However, larger studies identified over time that baseline lower levels of melatonin secretion were associated with a higher risk of developing T2D.[201] This unclear picture stemming from extensive characterisation of this locus by multiple groups, highlights the challenges to resolve the functional impact at GWAS loci.

The *PPP1R3B* locus is an example of a T2D-associated locus that is thought to attribute its effect through a contribution to insulin resistance, rather than a defect of insulin secretion. *PPP1R3B* locus is associated with increased levels of fasting glucose and insulin, elevated levels of triglycerides, and an increased waist-to-hip ratio. Further, *in vitro* experiments from Manning et al (2020)[194] suggest that *LOC157273*, a long noncoding RNA (lncRNA) which is the nearest gene to lead SNP at the locus, is the effector transcript, which in turn affects *PPP1R3B* expression, and therefore the causal SNPs does not impact PPP1R3B expression/function directly.[194] *LOC157273* regulates *PPP1R3B* mRNA levels and the study showed risk allele carriers had reduced levels of *LOC157273*, and increased *PPP1R3B* expression and glycogen deposition.[194] This locus has been mis-labelled as *PPP1R3B* in GWAS studies as this is the closest protein coding gene, highlighting the importance of considering the non-coding RNAs in aetiology.

High-throughput approaches have further been employed to further aid the understanding of the molecular mechanisms behind T2D GWAS associations. As well as integration of 'omics data, high throughput *in vitro* screens have been employed to gain better understanding of GWAS loci[202,203]. Such approaches have to date predominantly focused on characterising loci and genes in beta-cell dysfunction rather than insulin resistance[204–208]. However similar studies, not necessarily at genome-wide scale, looking at T2D-relevent outcomes in liver and adipose have been conducted[209–211].

To date, most genetic studies related to T2D and their follow up have focused on genetic variants and their impact on pancreatic beta-cell dysfunction and fasting insulin resistance (characterised by fasting measures of glucose or insulin), which primarily implicates the liver. However other peripheral tissues, muscle and adipose, are known to play a key role in T2D onset, being the primary tissues that are affected in post-

prandial insulin resistance, the first defect on the path towards T2D. Therefore, greater understanding of the impact of the T2D genetic architecture on adipose and muscle tissue is needed. This is one of the keys aims of this thesis and further discussion into the role of adipose and muscle tissues in T2D is discussed in chapters 2 and 3.

1.7 Conclusions

T2D and obesity are complex metabolic diseases that are contributed to by many factors, including genetic risk. Insulin resistance is a common thread that links T2D and obesity and therefore its role in both conditions warrants further investigation. Current research on insulin resistance in human studies at population scale predominately focused on insulin resistance in the fasting state, with such measures being easier to obtain. However as illustrated above, insulin resistance effects many tissues and such studies likely miss important aetiological insights by not assessing the post-prandial role of insulin, with such measures reflecting insulin action on distinct tissues. Genomic studies for T2D have been successful in identifying hundreds of genetic loci associated with increased risk of disease, however there is still a significant knowledge gap in understanding the molecular mechanism by which these loci contribute to disease risk. Studies have aims to address this through the integration of additional genetic associations with intermediate traits (in the case of T2D – glycaemic traits) and 'omics measures to try and further pinpoint causal mechanisms and genes at these loci. Functional follow up of T2D loci has predominately focused on single candidate loci of interest to date, and large-scale *in vitro* screens have focused mainly on the role of T2D loci in beta-cell dysfunction and have not considered insulin resistance. Therefore, further research is needed to fully understand the genetic architecture of metabolic disease and the mechanisms dysregulated with increased genetic risk.

Chapter 2: Genetic architecture of insulin resistance post-glucose challenge

Overview of chapter 2

Insulin resistance and pancreatic beta cell dysfunction are the two key pathophysiological mechanisms predisposing to type 2 diabetes (T2D). The molecular mechanisms through which insulin controls glucose metabolism in fasting and post-prandial states are distinct, with the former primarily involving the control of expression and phosphorylation of enzymes for gluconeogenesis and glycogenolysis in the liver, while the latter involves the rapid translocation of an intracellular pool of stored glucose transporters to the plasma membrane of muscle cells and adipocytes. Despite this, large-scale genetic studies of insulin resistance have predominantly focused on the fasting state and therefore primarily capture insulin action in the liver.

I conducted a genetic discovery of post-prandial insulin sensitivity using two indices which integrate fasting and glucose-stimulated insulin levels during an oral glucose tolerance test (OGTT): insulin fold change and modified Stumvoll Insulin Sensitivity Index. I meta-analysed 28 genome-wide association studies (GWAS) for each of these traits in up to ~55,000 participants representing European, Hispanic American, and East Asian ancestries.

I identified 11 loci ($p < 5x10^{-8}$), of which 9 were novel, including a signal at *SLC2A4* encoding the canonical insulin sensitive glucose transporter GLUT4, associated with post-challenge insulin resistance and reduced expression of *SLC2A4* in skeletal muscle. All but one of the genetic signals were shared with risk of T2D (posterior probability of co-localisation > 80%). Using an integrative approach, incorporating information on glycaemic trait associations and skeletal muscle gene expression, I prioritised 6 additional loci of interest associated with post-challenge insulin resistance that did not meet the traditional genome-wide significance threshold.

In summary, through the integration of genetic data and *in silico* follow up, I identified loci that affect T2D risk through impaired insulin-stimulated glucose uptake in peripheral tissues.

Publications related to this chapter:

Williamson A, [100 authors], Mohlke K.L., Wheeler E, O'Rahilly S, Fazakerley D.J., Langenberg C. Genome-wide association study of postprandial glucose metabolism and functional characterisation identifies candidate genes for insulin-stimulated glucose uptake. **Nature Genetics** (2023, accepted)

Contributions and Collaborations

I conducted all analyses outlined below except the experimental work presented in the results under subheading: "Assessment of the transcriptional regulation of rs117643180" which was conducted by collaborators in Professor Karen Mohlke's group at University of North Carolina. Annotation of the *SLC2A4* locus in METSIM was done using GWAS data provided by Dr Anne U Jackson, as I do not have access to individual level data in this cohort.

I carried out study-level GWAS for Insulin Fold Change in the Fenland Study (for each of the 3 arrays) and Ely study. GWAS for Modified Stumvoll ISI in Fenland and Ely were completed by Dr Jian'an Luan in the MRC Epidemiology Unit, before I started by PhD. All remaining GWAS included in the meta-analysis were conducted by other study-specific analysts as I do not have access to individual genotype and trait data for these studies.

This work was done as part of the MAGIC (the Meta-Analyses of Glucose and Insulin-related traits Consortium) Refined Insulin Traits working group. Studies were recruited jointly as part of this working group and Quality Control (QC) of study-level GWAS was conducted by the working group using a harmonised pipeline. I conducted the meta-analysis and all downstream work presented in this chapter unless otherwise indicated.

2.1 Aims and Objectives

Aim:

This chapter aimed broadly to characterise the genetic architecture of post-challenge insulin resistance, and further annotate the identified loci of interest using several strategies.

Objectives:

The following objectives were addressed in this chapter:

1 Assessment of the observational correlation of glycaemic traits with those capturing post-challenge insulin resistance.

2 Genetic discovery of post challenge insulin resistance. This took place in 2 stages. First, genome-wide genetic discovery to identify loci at significance threshold ($P < 5x10^{-8}$). Second, integrative GWAS approach to identify loci of interest below genome-wide significance.

3 *In silico* follow up of identified loci, including annotation with additional traits.

4 Genetic risk score and genetic corelation analyses.

2.2 Introduction

Defects in insulin secretion and insulin action are the two key pathophysiological mechanisms predisposing to T2D[212], which develops when fasting and/or postprandial glucose homeostasis cannot be maintained. Insulin works through distinct molecular mechanisms to control glucose metabolism in fasting and postprandial states. Most genetic studies of common forms of insulin resistance have focused on the fasting state, where the relationship between blood glucose and insulin is largely reflective of liver insulin sensitivity and therefore high fasting insulin is indicative of liver insulin resistance. This focus on fasting measures is largely because these samples are easier to obtain compared to dynamic tests with repeated sampling to estimate post-meal insulin resistance.[87,213,214]

In the fasting state low circulating insulin concentrations increases gluconeogenesis and glycogenolysis in the liver (reviewed in[19]), promoting hepatic glucose output and maintenance of glycaemia. Conversely postprandially, increases in plasma insulin concentration result in supressed hepatic glucose output and promotion of glycogen synthesis, as well as glucose uptake into muscle and adipose tissues.[215,216]

This insulin-dependent glucose uptake into muscle and adipose tissue occurs through rapid translocation of an intracellular pool of the glucose transporter GLUT4 to the cell surface.[34,217,218] GLUT4, encoded by *SLC2A4*, is the primary insulin-dependent glucose transporter and is critical for the postprandial uptake of glucose into muscle and adipose tissues. Insulin resistance is associated with lower GLUT4 expression and translocation in adipose tissue[219], and impaired GLUT4 translocation in muscle tissue.[220] Insulin resistance in muscle tissue occurs long before the manifestation of hyperglycaemia and T2D.[28,221]

Measures of post-prandial insulin resistance

The gold standard method of measuring insulin resistance in these GLUT4-expressing tissues is the hyperinsulinemic-euglycaemic clamp combined with stable isotope-based measurement of glucose clearance, with [6,6-^{2}H]glucose or [1-^{13}C]glucose commonly used as tracers.[222,223] Utilising stable isotope-based measures allows the uncoupling of exogenous glucose uptake from endogenous glucose turnover[224]. Briefly, the hyperinsulinemic-euglycaemic clamp involves the raising of insulin to a constant concentration through continuous infusion, with glucose concentration held at

basal level through variable infusion.[224] Hyperinsulinemic-euglycaemic clamp measures of insulin sensitivity without the use of labelled glucose are more widely measured in small scale clinical and population-based studies, but still not at sufficient scale for large-scale genetic analyses. Without the use of stable isotopes, the glucose infusion rate is equal to whole body insulin sensitivity, reflecting both hepatic glucose output and insulin-stimulated glucose uptake.

Oral glucose tolerance test (OGTT) is more amenable to be conducted at a scale sufficient to be sufficiently powered for genetic studies, due to reduced cost, time and training involved to successfully conduct such measures. A standard OGTT involves the ingestion of a standard 75g glucose drink following a period of fasting (usually overnight, minimum 8 hours), where blood measures can be obtained at desired timepoints following ingestion. Due to the increased ease of conducting OGTT compared to hyperinsulinemic-euglycaemic clamps, many indices have been developed to assess insulin sensitivity using fasting and dynamic OGTT measures that approximate the "gold standard" clamp-based measures.[225–228]

These indices use fasting and/or post-challenge measures of insulin and glucose taken throughout the OGTT. Indies using exclusively fasting measures of insulin/glucose include HOMA[229] and QUICK1[230] indices. As these measures use only fasting measures, they predominantly represent hepatic insulin resistance.[231,232] For example, HOMA, homeostatic model assessment, includes two indices developed to estimate beta-cell function (HOMA-B) and insulin sensitivity (HOMA-IR), two key features of T2D, using fasting measures of insulin and glucose.[229,233] Additional indices have been developed using post-glucose challenge measures from these dynamic tests (reviewed in[77]). These include Modified Stumvoll insulin sensitivity index (ISI)[234,235] and Matsuda ISI[228] to assess whole body insulin sensitivity.

Genetic contribution to post-prandial insulin resistance

There are currently only a few examples of loss of function coding variants that contribute to tissue-specific changes in postprandial insulin action and increased T2D risk.[108,111,113,236] One of the most studied examples are coding mutations in *TBC1D4,* a key player in insulin-stimulated glucose uptake. *TBC1D4 (*also known as AS160*)*, encodes a Rab-GTPase activating protein (Rab-GAP) involved in insulin signalling and control of GLUT4 translocation.[237] TBC1D4 is regulated by insulin with the insulin-

stimulated activation of AKT resulting in the downstream phosphorylation of TBC1D4 by AKT. This phosphorylation of TBC1D4 results in the suppression of GAP activity, meaning that there is a reduction of conversion of active Rab-GTP bound proteins to inactive GDP bound forms (Rab-GDP). This releases GLUT4 secretory vesicles (GSVs) from their intracellular storage compartment to be translocated to the cell surface, allowing the uptake of glucose.[237]

As discussed in Chapter 1, loss of function mutations in *TBC1D4* have been associated with insulin resistance specifically following an oral glucose tolerance test (OGTT) in the Greenlandic Inuit population and in patients with severe insulin resistance.[108,111,236] In the Inuit population, common nonsense mutation *TBC1D4* Arg684ter is associated with higher 2hr glucose and insulin levels, and ~ 10x increased risk in T2D in homozygosity.[236] These mutations interfere with GLUT4 translocation in skeletal muscle, resulting in high blood glucose and insulin levels post glucose challenge, relative to fasting.[236] The carriers further have lower than normal fasting glucose and insulin levels. This is likely to be because the mutations in *TBC1D4* result in a continuous leak of GLUT4 to the plasma membrane and thus glucose uptake. This further results in a depletion of the intracellular pool of GLUT4 in GLUT4 secretory vesicles.

These data suggest that there is value in studying the genetic architecture of measures which differentially capture insulin action in the postprandial versus the fasting state.

Genetic architecture of post-challenge insulin resistance

Here, I studied two OGTT-derived measures of insulin resistance after a glucose challenge to identify genetic variants underlying these traits in a genome-wide meta-analysis (GWAS) of >55,000 participants without diabetes from 3 ancestry groups across 28 studies. These measures were the Modified Stumvoll Insulin Sensitivity Index (ISI) and the fold-change in insulin concentration (insulin fold-change; IFC) following a glucose challenge, both of which leverage measures of glucose and insulin during an OGTT.

IFC reflects the relationship between fasting and post challenge plasma insulin. Higher IFC values suggest more insulin is required to lower blood glucose after a standard glucose load and is therefore indicative of post-challenge insulin resistance. IFC is

dramatically increased in people who carry mutations that impair glucose transporter translocation and postprandial glucose uptake into skeletal muscle[108,111,236]Therefore, IFC has a strong face validity as an index of post-challenge, rather than fasting, insulin resistance.

Modified Stumvoll ISI (now referred to as ISI) incorporates measures of insulin and glucose at fasting and post oral glucose load, and has been demonstrated to approximate hyperinsulinemic-euglycaemic measures of whole-body insulin sensitivity.[235] Lower values of ISI reflect greater whole-body insulin resistance. Developed by Stumvoll et al (2000), ISI was demonstrated to be highly correlated with clamp-based measures of insulin sensitivity (r = 0.79) in a cohort of 104 non-diabetic individuals, and outperformed HOMA and other measures using fasting measures as well as the Matsuda index and 2 h insulin alone.[235] Formulae were subsequently developed to allow the estimation of ISI using different timepoints during the OGTT maximising its utility in studies with different timepoints, for example using only fasting and 120 mins (such as used in this project) had a correlation of r = 0.69 with clamp based measure insulin sensitivity.[234]

Previous meta-analysis efforts have examined the genetic architecture of ISI.[238,239] Whilst no genome-wide significant loci where identified by Dimas *et al* (2014)[239], Walford *et al* (2016) identified 3 genetic loci significantly associated with ISI ($P < 5x10^{-8}$; N = 16,753).[238] These include *BCL2* (rs12454712) and *NYAP2* (rs13422522; in Linkage Disequilibrium (LD) with previously known *IRS1* locus rs2943641). The *FAM19A2* locus (rs10506418) was identified to be significantly associated only in an interaction model testing the effect of genotype on ISI by BMI, which have not been explored in this present study.[238] The expanded meta-analysis in this chapter reflects over 3x increase in sample size from this previous meta-analysis of GWAS for ISI.[238]

IFC and ISI therefore can capture post-challenge insulin resistance and approximate postprandial insulin resistance in tissues such as skeletal muscle and adipose tissue. Using a genetically driven approach, 9 novel genetic loci specifically implicated in post-challenge insulin resistance were identified, in cohorts of European ancestry. This includes 4 novel loci for post-challenge insulin resistance significantly ($P < 5x10^{-8}$) associated with IFC and 6 novel loci for post-challenge insulin resistance associated with ISI, with one locus associated with both traits.

2.3 Methods

2.3.1 Study Design

A summary of the analyses conducted to address the aims and objectives of this chapter are summarised in **Figure 2.1**.

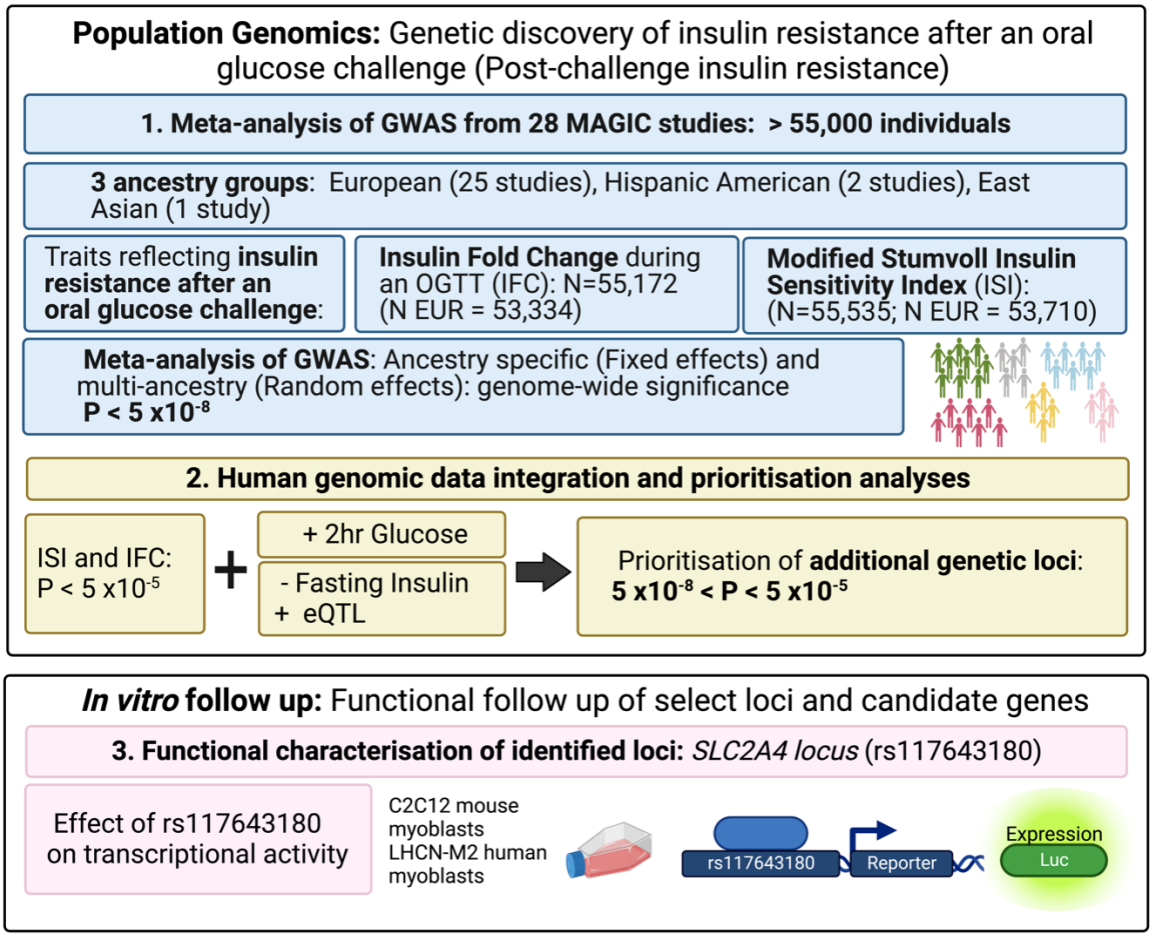


Figure 2.1: Study design to assess genetic architecture of post-challenge insulin resistance

Figure created using Biorender.com

Contributing studies

A total of 55,535 and 55,172 participants without diabetes across 28 studies contributing to the Meta-Analyses of Glucose and Insulin-related Traits Consortium (MAGIC) contributed to the analyses of ISI and IFC, respectively (**Tables 2.1 and 2.2**). Two of these studies included participants of Hispanic American ancestry (MACAD and HTN-IR; N ISI = 1095, N IFC = 1099) and 1 included participants of East Asian ancestry (TAICHI; N ISI = 730 N IFC = 739); the remaining 25 studies included individuals of European ancestry. Ancestry group was defined at the study level as the ancestry reported by each respective study at point of data collection. Within each study, we excluded individuals with missing traits or covariates from the analyses.

Table 2.1 Contributing study sample size and ancestry for insulin fold change (IFC)

Study	Max Sample Size	Population
Fenland: (Fenland-OMICS, Fenland-OMICS, Fenland-GWAS)	10123	EUR
METSIM	8440	EUR
INTER99	5051	EUR
Botnia-PPP	3875	EUR
Botnia_Family	3009	EUR
TUEF	2575	EUR
FHS	2527	EUR
LIFE-Adult	2426	EUR
MDC	2320	EUR
ELY	1248	EUR
ADDITIONPRO	1204	EUR
LURIC	1139	EUR
DRS EXTRA	1046	EUR
RISC	968	EUR
ULSAM	960	EUR
DIAGEN	891	EUR
EUGENE2	876	EUR
SORBS	838	EUR
TAICHI	739	EAS
HEALTH2008	717	EUR
KORA	715	EUR
FUSION_GWAS	589	EUR
HTN-IR	586	His-AMR
BC1936	579	EUR
RIGCOH	529	EUR
MACAD	513	His-AMR
FUSION2	364	EUR
dps	325	EUR
TOTAL	55172	
European	53334	
Non-European	1838	

Table 2.2 Contributing study sample size and ancestry for Modified Stumvoll insulin sensitivity index (ISI)

Study	Max Sample Size	Population
Fenland (Fenland-OMICS, Fenland-OMICS, Fenland-GWAS)	10785	EUR
METSIM	8440	EUR
INTER99	5032	EUR
Botnia-PPP	3875	EUR
Botnia_Family	3009	EUR
FHS	2527	EUR
TUEF	2479	EUR
LIFE-Adult	2426	EUR
MDC	2320	EUR
Ely	1244	EUR
ADDITIONPRO	1202	EUR
LURIC	1139	EUR
DR's EXTRA	1046	EUR
ULSAM	962	EUR
RISC	956	EUR
DIAGEN	891	EUR
EUGENE2	875	EUR
SORBS	838	EUR
TAICHI	730	EAS
KORA	715	EUR
HEALTH2008	713	EUR
FUSION_GWAS	588	EUR
HTN-IR	582	His-AMR
BC1936	576	EUR
RIGCOH	529	EUR
MACAD	513	His-AMR
dps	325	EUR
FUSION_2	218	EUR
TOTAL	55535	
TOTAL EUR	53710	
TOTAL NONEUR	1825	
Previously published Meta-analysis, discovery stage (Walford et al 2016) [238]	16,753	

Phenotype definition

Individuals were excluded from study-level analyses if they were diagnosed with T2D, reported use of diabetes medication, or were classed as having "biochemical diabetes": Fasting Glucose ≥ 7mmol/L, 2hr Glucose ≥ 11.1 mmol/L or HbA1c ≥ 6.5%.

Fasting measures of plasma insulin and glucose were collected before a standard 75g oral glucose tolerance test (OGTT) with 2hr glucose and insulin measures obtained 120mins after the ingestion of the glucose challenge.

Modified Stumvoll insulin sensitivity index (ISI) was calculated as previously described using the formula[234,235]:

$$Modified\ Stumvoll\ ISI = 0.156 - \left(0.0000456 \times insulin_{120\,\mathrm{min[pmol/L]}}\right) - \left(0.000321 \times insulin_{0\,\mathrm{min[pmol/L]}}\right) - \left(0.0054 \times glu\cos e_{120\,\mathrm{min[mmol/L]}}\right)$$

Insulin fold change (IFC) was calculated using fasting and 120 min insulin measures during OGTT using formula:

$$IFC = \ln\left(\frac{Insulin_{120mins[pmol/L]}}{Insulin_{0\ mins[pmol/L]}}\right)$$

Formulae terms: $insulin_{120min}$ – insulin levels at 120 mins post-ingestion of glucose challenge in pmol/L. $Insulin_{0min}$ – fasting insulin levels prior to glucose challenge in pmol/L. $Glucose_{120min}$ – glucose levels at 120 mins post-ingestion of glucose challenge in mmol/L.

2.3.2 Correlation of ISI and IFC with relevant traits

The correlation of relevant traits was assessed in the Fenland[240] and RISC[241] studies using Spearman's Rank correlation for pairwise complete observations. ISI and IFC were calculated as defined above and were restricted to non-diabetic individuals. Individuals who had diabetes, reported use of diabetes medication, or had fasting glucose ≥ 7 mmol/L, 2hr glucose ≥ 11.1 mmol/L or HbA1c ≥ 6.5% were excluded. Blood biochemistry measures were natural log transformed. IFC was calculated using OGTT insulin measures at 0 min and 120 min.

2.3.3 Genetic discovery: Study-level GWAS and meta-analysis

Genotyping, quality control, (QC), and imputation

Study-level QC, genotype imputation, and genome-wide association study (GWAS) analyses were conducted following a shared analysis plan (see **Appendix**) to ensure analyses were conducted consistently across studies. Most study samples were genotyped using commercially available genome-wide genotyping arrays, with a small subset of study samples genotyped on targeted arrays such as Illumina Cardio-Metabochip (Metabochip) array.

Genotype imputation was conducted for all autosomes and the X chromosome in all contributing studies on human genome build GRCh37 (Hg19) on the forward strand. Most studies of European ancestry were imputed to the Human Reference Consortium (HRC) reference panel.[242] Studies of non-European ancestry were imputed to population-specific 1000 genomes (1000G) reference panel.[243] Botnia-PPP was imputed to the SISu v3 reference panel in genome build GRCh38[244] and variants were mapped to build GRCh37 using UCSC LiftOver, after GWAS analyses.[245]

Study-level GWAS

Study-level GWAS were performed using an additive model, modelling on the inverse normal transformed residuals of each trait and adjusting for age, sex, population structure, and study-specific covariates with or without adjustment for BMI. Full analysis plan for IFC and ISI is included in the appendix. To control for type 1 error rate of low-frequency variants and remove trait-covariate correlations, covariate adjustment for each trait was conducted in two steps before association analysis. First, a given trait was adjusted for all covariates in a model and residuals were derived. These residuals were inverse normal transformed and then modelled adjusting for all covariates. The residuals from step 2 were used in the final GWAS analysis.

Study-level GWAS QC

Each contributing study shared GWAS summary statistics for both traits and both models. Study-level GWAS QC was performed by the central analysis group using easyQC, to assess trait transformation and exclude variants that failed QC from downstream analyses.[246] Variants were flagged if the differences of their minor allele frequency (MAF) between the study and the corresponding frequencies in the imputation reference panel were greater than 0.2. In addition, variants were excluded

if they met one or more of the following criteria: monomorphic, low genotype call rate (< 95%), poor imputation quality (INFO < 0.3), duplicate variants, minor allele count (MAC) < 3 per study, or not in Hardy Weinberg Equilibrium (HWE P < $1x10^{-5}$). Variants were additionally excluded if these had standard error of the effect size > 10 or were missing an effect estimate, standard error, or p-value in the summary statistic file. For the RISC study variants were excluded where imputation INFO score < 0.8.

Single-ancestry meta-analysis

An inverse variance fixed-effect meta-analysis of the study level GWAS was done using METAL for each ancestry, using beta-coefficients and standard errors from each study-level GWAS.[247] I applied genomic control correction to both study-specific GWAS results and to the meta-analysis results for each ancestry group (double genomic control). Study-specific Metabochip results were corrected by genomic control using test statistics for 4986 SNPs included on the Metabochip array associated with QT interval, a trait that is not correlated with glycaemic traits, consistent with previous analyses of other glycaemic traits.[87,248]

In each individual study variants from with a MAF < 0.005 were removed prior to meta-analysis. Genome-wide significance was considered at P < $5x10^{-8}$, with a genomic locus considered as a 1 Mb window (± 500kb) around the lead variant at each locus. Loci were annotated with the nearest protein-coding gene where applicable or based on previously reported gene annotation in the literature. Lead variant associations were manually inspected across all individual studies to ensure consistency of association.

Conditional analysis was conducted using the –cojo-slct option in GCTA (v1.26.0) using BMI adjusted summary statistics for IFC and ISI in European studies to identify independent signals at each locus.[82] A collinearity threshold of 0.8 and conditional p-value threshold of P < $5x10^{-8}$ were used to identify conditionally independent genetic signals at each locus. The Fenland-OMICS subset (N = 8,925) of the Fenland Study with dense genome-wide genotyping was used as an LD reference panel.

Post meta-analysis QC

Post-meta-analysis, variants with a total sample size less than 11,000 in the European ancestry only or multi-ancestry (EUR, HIS-AMR, and EAS) analyses were removed, ensuring that the associations observed were contributed to by at least 2 studies. In

the non-European ancestry (HIS-AMR and EAS) analysis, variants detected in < 1,000 individuals were removed, ensuring these associations were seen in at least 2 of the 3 studies.

Novel loci were defined as those loci which have not been reported to be associated with post-challenge glycaemic traits or indices reflecting post-challenge insulin resistance.

For both IFC and ISI, fixed effect meta-analyses of GWAS from BMI adjusted (adjBMI) analyses in European ancestry cohorts were considered the primary analyses. Unless specified otherwise, the results of this primary analysis are reported in the main text and supplementary tables.

Multi-ancestry meta-analysis

I conducted a multi-ancestry random-effects meta-analysis using random-metal (https://github.com/explodecomputer/random-metal) to meta-analyse the European, Hispanic American (His-Amr) and East Asian (EAS) ancestry specific analyses (see above), to delineate any ancestry specific effects on these traits.[247,249] The meta-analysis strategy is summarised in **Figure 2.2**.

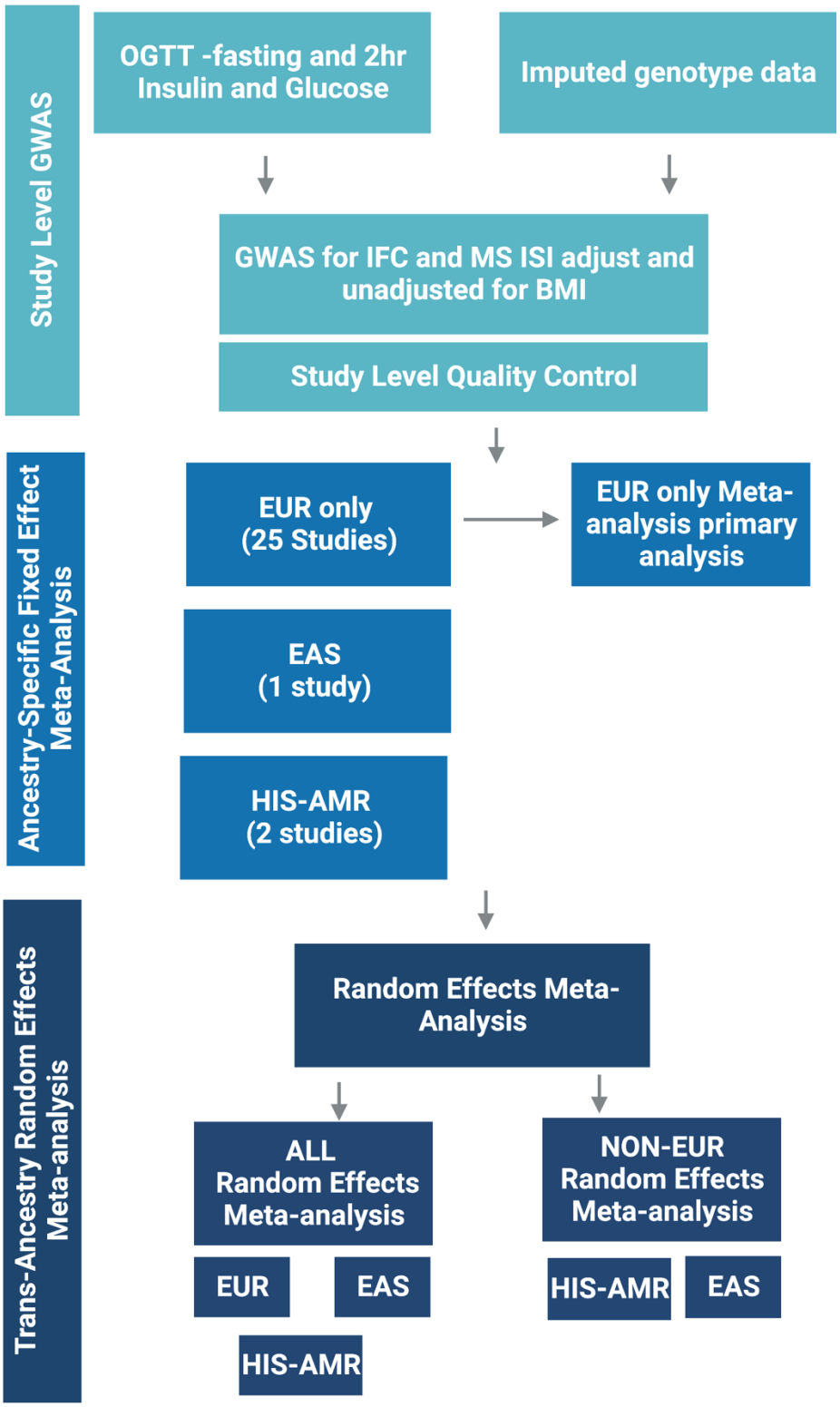


Figure 2.2: Meta-analysis workflow for genetic discovery analyses for ISI and IFC
Analysis workflow for the meta-analysis of study-level GWAS results for Insulin fold change and Modified Stumvoll ISI

2.3.4 loci characterisation

Fine mapping

I performed fine-mapping of the 11 loci significantly associated with IFC or ISI in European ancestry meta-analysis to define 99% credible sets for each locus. I generated approximate Baye's factors (ABFs) using the method described by Wakefield (2009)[250] and calculated the posterior probability that each variant was casual. Credible sets were defined including all variants that contained cumulative posterior probability greater than 99% in each region. Credible sets were calculated using marginal summary statistics for all loci where there was identified to only be one independent signal. At *PPP1R3B* where 2 independent signals were identified for insulin fold change, conditional summary statistics were used to calculate the 99% credible set for each signal.

Statistical colocalisation of IFC and ISI with T2D

Bayesian colocalisation analyses were conducted using coloc.abf from R package coloc[177] to test for genetic colocalisation of IFC and ISI with T2D[251]. BMI adjusted summary statistics in studies of European ancestry were used. Colocalisation was considered at a threshold of posterior probability > 0.7 for hypothesis "PP.H4.abf": a locus is associated with both traits and there is one shared causal SNP for the pair of traits being assessed.

Phenotypic annotation of associated loci

I used gene expression data from the GTEx consortium to assess whether lead variants are annotated as tissue specific expression quantitative trait loci (eQTLs) at a suggestive significance threshold of $P < 1x10^{-4}$ in tissues relevant to T2D – skeletal muscle, adipose (subcutaneous and visceral), liver, kidney[154]. I additionally used summary statistics for relevant traits that are publicly available to further annotate the loci of interest: T2D[251], WHR[252], fasting insulin[87], fasting glucose[87], 2hr glucose[87]. BMI-adjusted summary statistics in studies of European ancestry were used for all traits. These were additionally used to annotate lead variants at loci identified in the integrated GWAS approach (see below). Rare variant burden association statistics were obtained from the AstraZeneca PheWAS Portal (https://azphewas.com/) which includes pheWAS analysis across exomes from ~450,000 individuals of European ancestry in UK Biobank. I extracted summary statistics for rare damaging missense variants which were defined using the "raredmgmtr" variant mask – which includes high quality calls for missense variants that are rare (MAF < 0.005%), damaging (REVEL

score ≥ 0.25), and fall into a region of constrained variation according to the missense tolerance score metric (MTR < 0.78 or MTR_centil < 0.5). For more information please see the publication outlining this work.[253]

Genetic risk score analysis

I additionally constructed genetic risk scores (GRS) for IFC and ISI including lead variants at primary association signals in Europeans. GRS were constructed using the sum of allele scores for each individual where allele counts were weighted by the effect of this allele on IFC or ISI, adjusted or unadjusted for BMI as appropriate. Weights used in the construction of the GRS are outlined in **Table 2.3**. Allele counts were discrete for genotyped variants (0 = reference, 1 = heterozygous, 2 = homozygous) and weights were aligned to the dosage coding allele prior to weighting of each variant. Weighted allele counts were summed for each individual to generate the GRS. GRS were constructed for ISI and IFC adjusted and unadjusted for BMI in unrelated individuals of European ancestry in UK Biobank and the OMICS subset Fenland Study (N = 8925)..

Table 2.3 - Variant weights used in Genetic Risk Score (GRS) calculation for IFC and ISI, adjusted and unadjusted for BMI

GRS Trait	Locus	rsid	EA	NEA	EAF	P_adjBMI	P_noBMI	GRS_Weight_adjBMI	GRS_Weight_noBMI
IFC	MTN1RB	rs1387153	T	C	0.3023	2.25E-10	4.33E-10	0.0438	0.0431
IFC	PPP1R3B_primary	rs7012814	A	G	0.4796	1.35E-13	1.07E-13	0.0481	0.0483
IFC	PPP1R3B_secondary	rs2126259	T	C	0.1265	2.86E-12	6.12E-13	-0.0688	-0.0706
IFC	C2CD4A	rs7167878	A	C	0.4606	2.76E-11	3.41E-11	-0.0426	-0.0424
IFC	SLC2A4	rs117643180	A	C	0.0231	1.11E-10	5.26E-11	0.1428	0.1453
ISI	IRS1	rs2972144	A	G	0.366	1.10E-21	5.31E-12	0.0644	0.0458
ISI	COBLL1	rs12692738	T	C	0.7724	5.24E-14	3.44E-06	-0.0568	-0.0351
ISI	PPARG	rs11128603	A	G	0.8595	2.64E-12	2.33E-05	-0.0635	-0.0389
ISI	SLC2A4	rs117643180	A	C	0.0231	1.58E-10	1.67E-08	-0.1416	-0.1251
ISI	C5orf67	rs459193	A	G	0.2816	6.18E-09	7.95E-07	0.0409	0.0348
ISI	MTOR	rs2295080	T	G	0.6961	1.34E-08	0.000469	-0.0394	-0.0243
ISI	BCL2	rs12454712	T	C	0.5858	4.42E-08	0.000739	-0.0374	-0.0231
ISI	FAM101A	rs1906937	A	C	0.2894	4.86E-08	8.99E-06	-0.0384	-0.0313

Linear regression models were run for each genetic score and phenotype pair adjusting for age, sex and the 1st 10 genetic principal components to account for population structure. Phenotypes were scaled to a mean of 0 and SD of 1 to allow direct comparison across phenotypes. Blood biochemistry measures in Fenland and UK Biobank were natural log transformed prior to scaling to account for skewed distribution of these traits in both studies. Significance was considered at a Bonferroni significance threshold in UKBB ($P < 4.5x10^{-4}$; N GRS = 4, N phenotypes = 28) and nominal significance threshold in the Fenland Study ($P < 0.05$).

LD Score Regression

I applied LD score regression[254] using python-based software LDSC (accessed at https://github.com/bulik/ldsc) to assess genome-wide genetic correlations between IFC and ISI with other cardiometabolic traits. I calculated LD scores for ISI, IFC using BMI adjusted summary statistics from our analyses in cohorts of European ancestry. I further calculated LD scores for additional T2D related traits from publicly available summary statistics for 2 h glucose[87], fasting insulin[87], T2D[255] and WHRadjBMI[256] in cohorts of European ancestry. Further pre-computed LD scores were obtained online for additional cardiometabolic traits of interest (https://data.broadinstitute.org/alkesgroup/LDSCORE/). The genetic correlations were calculated restricting to HapMap3 SNPs with a MAF > 0.05, and used the 1000G European LD Score reference.[254]

2.3.5 Integrated GWAS approach

I employed an integrated GWAS approach, which has been successful in identifying variants associated with insulin resistance[257], to identify genetic variants which had a similar association pattern to that of the *SLC2A4* locus, using a more permissive significance threshold. Variants which were associated with IFC and ISI ($P < 5x10^{-5}$) were selected.

I subsequently integrated associations for IFC and ISI with the associations reported for fasting insulin[87], 2hr glucose[87], and skeletal muscle eQTLs[154]. BMI adjusted summary statistics from meta-analyses including studies of European ancestry were used for all traits. For IFC and Modified Stumvoll ISI, the European only BMI adjusted analyses were used. For 2hr glucose and fasting insulin, the largest available BMI

adjusted summary statistics in European studies were used.[87] Lead variants for IFC or ISI at the loci selected to not be associated with fasting insulin (strategy 2 outlined below), were subsequently assessed for their association and statistical colocalisation (see below) with mRNA expression levels in skeletal muscle (eQTLs) in GTEx v8.[154]

I used two variant selection strategies to refine loci which have the following phenotypic patterns of association:

1. Identify loci additionally associated 2hr Glucose ($P < 5x10^{-4}$)
2. Identify loci not associated with Fasting insulin ($P > 0.05$), and reported to be an eQTL in skeletal muscle in GTEx v8 ($P < 5x10^{-4}$)[154], with supportive evidence from the FUSION study ($P < 0.05$)[258].

Where conditions were met at a given locus for both ISI and IFC, the lead variant was selected to be that that met the threshold for both ISI and IFC. These prioritisation strategies are summarised in **Figure 2.3**.

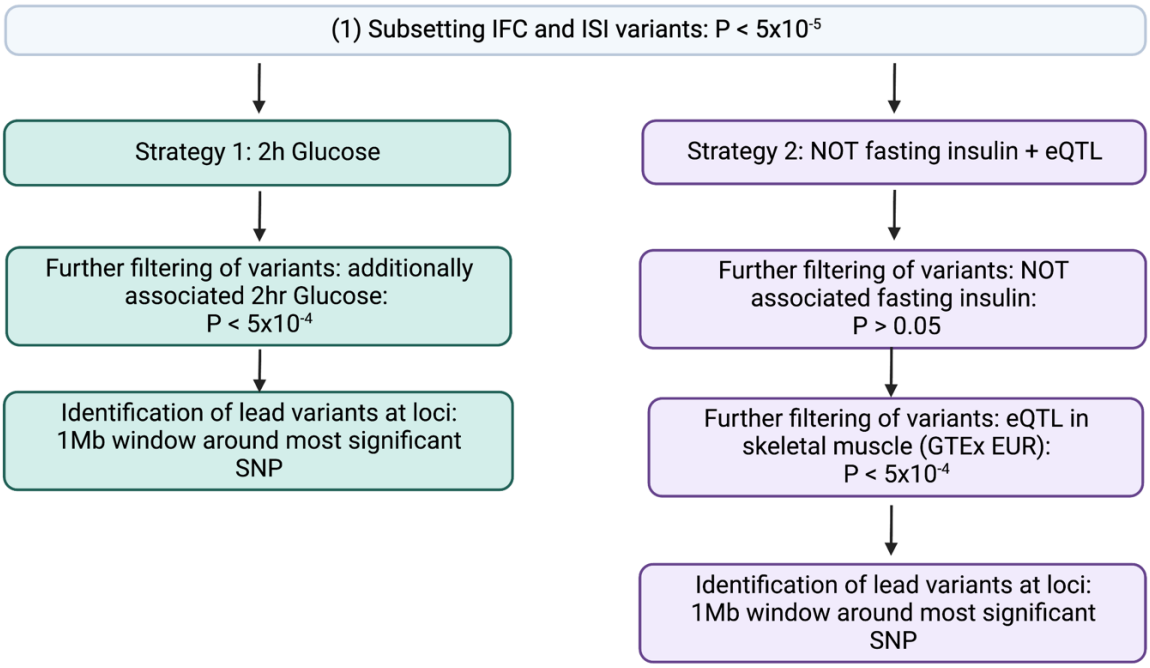


Figure 2.3 Summary of integrated GWAS approach strategies

Statistical colocalisation with relevant traits

I further assessed the identified loci using HyPrColoc[176] to identify colocalised signals across the following traits: T2D[251], WHR[252], fasting insulin[87], 2hr glucose[87], ISI, and IFC. BMI adjusted summary statistics in studies of European ancestry were used for all traits. Loci were defined as ±500kb around the lead variant for IFC and ISI, respectively with a significance threshold of $P < 5x10^{-5}$ used for ISI and IFC. Additional traits assessed were dropped from the analysis at each of the selected loci where $P > 5x10^{-4}$. HyPrColoc was run across default priors ($p1 = 1x10^{-4}$, p2 = 0.98 and threshold = 0.5) as well as looped over a grid of priors to allow for optimisation ($p1 = 1x10^{-4}$; p2 = 0.95, 0.98, 0.99, 0.999; threshold = 0.5,0.6, 0.7, 0.8, 0.9). Colocalisation of traits was considered at a threshold of posterior probability of association > 0.8. The largest stable cluster of traits that met this threshold was selected and the most stringent set of priors (higher values for p2) at each locus was used. Identified shared signals were further assessed by conducting pairwise colocalisation using HyPrColoc for all traits of interest at the selected loci to confirm the stability of these colocalised traits at each locus (pairwise colocalisation posterior probability > 0.7).

2.3.6 Tissue of action

Statistical colocalisation with eQTLs in relevant issues

I additionally tested for colocalisation of eQTLs in relevant tissues (skeletal muscle, adipose, liver, and pancreatic islet) with IFC/ISI at the loci of interest. Colocalisation was considered as significant when the colocalisation posterior probability was > 0.7. eQTL marginal summary statistics in individuals of European ancestry from the GTEx consortium (v8) were used in relevant tissues (Adipose – Subcutaneous, Adipose – Visceral Omentum, Muscle – Skeletal, Pancreas, Liver).[155,156,259] Bayesian colocalisation analyses were conducted using coloc.abf from R package coloc using default priors.[177]

Tissue specificity

To dissect the tissue of action of IFC and ISI loci, I used LDSC-SEG to assess the genome-wide enrichment of tissue-specific and cell type specific annotations.[260] In brief, LDSC-SEG employs stratified LD score regression to assess for genome-wide enrichment of sets of specifically expressed autosomal genes for a given trait of interest. LD scores available for a total of 205 specific tissue and cell types from GTEx (53 tissues; RNA-seq) and Franke Lab dataset which consists of expression data for 152 tissues of cell types in human, mouse, and rat[260] were used. In addition, LD scores

also generated by Finucane *et al* (2018)[260] from chromatin annotation from Roadmap[261] and ENTEx data from ENCODE[262] were examined. Suggestive enrichment was considered at a nominal significance threshold of $P < 0.05$, with no result surviving multiple testing correction.

2.3.7 Characterisation of the *SLC2A4* locus

In vitro assessment of rs117643180 at SLC2A4 (Collaboration: Prof Karen Mohlke)

Cell culture

C2C12 murine myoblasts (TCF-UNC/ATCC; CRL 1772) were grown in high-glucose Dulbecco's Modified Eagle's Medium (Sigma-Aldrich, St. Louis, MO) supplemented with 10% fetal bovine serum. LHCN-M2 immortalized human myoblasts[263] (Purchased from Evercyte, Vienna, AT) were grown on dishes coated with 0.1% pigskin gelatin (Sigma-Aldrich) in a medium consisting of four parts 4.5 g/L DMEM and one-part M199 (both Gibco, Waltham MA) and supplemented with 15% FBS, 20 mM HEPES buffer, 30 ng/mL zinc sulfate, 1.4 ug/mL vitamin B12, 55 ng/mL dexamethasone, 2.5 ng/mL HFG, and 10 ng/mL basic FGF. To differentiate LHCN-M2 myoblasts to myotubes, cells were cultured on gelatin-coated dishes for seven days in 4:1 DMEM: M199 supplemented with 20 mM HEPES buffer, 30 ng/mL zinc sulfate, 1.4 ug/mL vitamin B12, 10 ug/mL insulin, and 100 ug/mL apotransferrin. All cells were maintained in a humidified incubator at 37°C with 5% CO_2. Cells tested negative for *Mycoplasma* contamination in accordance with the MycoAlert Mycoplasma Detection Kit (Lonza, Morristown, NJ).

Transcriptional reporter assays

Dual-luciferase reporter assays were performed as previously described[264]. A 229-base pair fragment surrounding rs117643180 (chr17:7185651-7185879, GRCh37) was amplified via PCR and cloned into the firefly luciferase reporter vector pGL4.23 (Promega, Madison, WI) in both orientations with respect to the minimal promoter. This fragment is located in *SLC2A4* intron 1 and includes approximately 100 bp of an untranslated alternate first exon. Alleles for additional variants located within the insert were kept consistent across all plasmids. To create plasmids containing the rs117643180-A allele, we performed site-directed mutagenesis using QuikChange II (Agilent Technologies, Santa Clara, CA) and PAGE-purified primers (Integrated DNA Technologies, Newark, NJ). Four or five clones for each allele, in each orientation,

were purified and sequence confirmed (GeneWiz, RTP, NC). Primer sequences can be found in **Table 2.4**.

Table 2.4 Primers to test rs117643180 for allelic effects on transcriptional regulation.

Primers used for amplification of fragment spanning rs117643180		
Insert Orientation	**Primer Orientation**	**Sequence (5'to 3')**
Forward	Forward	GCGCGGTACCGCTCACGGGCAGTGTTTC
	Reverse	GCGCCTCGAGGGAGGAGGGCAGGAAGGT
Reverse	Forward	GCGCCTCGAGGCTCACGGGCAGTGTTTC
	Reverse	GCGCGGTACCGGAGGAGGGCAGGAAGGT
Primers used for site-directed mutagenesis (C>A_128)		
N/A	Forward	GGGACTGTCAGCAC**C**CCCTCCTCCA
	Reverse	TGGAGGAGGG**G**GTGCTGACAGTCCC
Primers used for electrophoretic mobility shift assays		
N/A		C Allele: CTGTCAGC**C**CCCCCTCC
		C Allele: GGAGGGGG**G**GCTGACAG
		A Allele: CTGTCAGC**A**CCCCCTCC
		A Allele: GGAGGGGG**T**GCTGACAG

Sequences are listed in the 5' --> 3' orientation with respect to the genome. rs117643180 is shown in underlined, boldfaced font where applicable.

Duplicate wells of C2C12 cells (35,000 per 1.9 cm^2) were co-transfected with 500 ng of luciferase plasmid and 80 ng of phRL-TK *Renilla* luciferase reporter plasmid (Promega) using transfection reagent Lipofectamine 3000 (Thermo Fisher, Waltham, MA). Duplicate wells of LHCN-M2 cells (40,000 per 1.9 cm^2) were co-transfected similarly either in an undifferentiated state or on day 6 of differentiation using transfection reagent LTX (Thermo Fisher). Reporter assays were performed 30 hours (LHCN-M2) or 48 hours (C2C12) after transfection. Luciferase measures were normalized to corresponding *Renilla* measures, and mean fold-change across duplicate wells was calculated relative to two to four independent preparations of empty vector pGL4.23. Alleles were compared using two-sided unpaired t-tests (α = 0.05). All transfections were repeated on a different day and showed consistent results.

Electrophoretic mobility shift assays

Assays were conducted as previously described using C2C12 or LHCN-M2 nuclear extracts.[265] Seventeen-bp complementary oligonucleotides centred around

rs117643180-C or rs117643180-A and biotinylated at the 5’ end (IDT) were annealed. Binding reactions were carried out with 6-10 µg of nuclear extract, 200-400 fmol of biotinylated probe, and 10- to 30-fold excess of unlabelled competing probe.

2.4 Results

2.4.1 Study overview

Gold standard measures of post-prandial insulin resistance, which can be obtained by hyperinsulinemic-euglycemic clamp coupled with stable radiolabelled isotopes, are not available at scale and therefore previous efforts have approximated post-prandial insulin resistance from measures of insulin and glucose during an OGTT.[225,235] In this study I utilised two measures of insulin resistance post glucose challenge (subsequently referred to as post-challenge insulin resistance) to capture, at least in part, post-prandial insulin resistance – Modified Stumvoll Insulin Sensitivity Index (ISI) and insulin fold change (IFC).

I assessed the genetic architecture of post-challenge insulin resistance through meta-analysis of GWAS from a total of 28 studies of European, Hispanic American and East Asian ancestry (**Figure 2.1 and 2.2; Tables 2.1 and 2.2**). I conducted ancestry specific fixed effect meta-analysis, with the primary analysis presented being the European only meta-analysis for IFC and ISI, adjusted for BMI. I additionally conducted random effects multi-ancestry meta-analysis across all ancestries as well as only the 3 non-European cohorts of East Asian and Hispanic American ancestry. Twenty five of the 28 studies included participants of European ancestry and represented 97% of total sample size.

2.4.2 Observational correlation of post-challenge insulin resistance with related traits:

First, I aimed to assess the observational correlation of the two post-challenge insulin resistance – IFC and ISI – to assess how specific these are to post-challenge insulin resistance, compared to whole body insulin resistance.

The Fenland Study

To determine the traits correlated with insulin action in the fasted and fed state, I conducted observational correlation analyses for fasting insulin and dynamic insulin traits Modified Stumvoll ISI and Insulin Fold Change during an OGTT, with metabolic traits measured in the Fenland Study (Max N for individual trait= 1454 (c-peptide)-11787).

IFC and ISI show strong correlation with measures of insulin and glucose taken at 2hrs during a standard OGTT, however are poorly correlated with fasting measures,

suggesting these are primarily capturing the post-challenge insulin response (**Figure 2.4)**. ISI is additionally inversely correlated with liver-related traits, as seen with fasting insulin (hepatic steatosis score and measures of circulating alkaline phosphatase, gamma-glutamyl transferase, alanine aminotransferase), suggesting that ISI may additionally capture insulin action on the liver in the fasted state (**Figure 2.4**). IFC was not correlated with any liver-related traits, suggesting that this potentially more specific to the post-prandial state. This is further evidenced by observations that ISI is more strongly correlated with HOMA-IR, an index of fasting insulin resistance (**Figure 2.4**). Further, IFC was weakly correlated with ISI (rho = -0.35), compared to other insulin related traits e.g., 2-hour insulin, suggesting that a GWAS of this measure may capture variants that are distinct from those that are driving 2-hour insulin or alternatively that these traits are able to capture different aspects and pathways that contribute to the post-challenge insulin response.

Fasting insulin was positively correlated with many metabolic traits including glycaemic traits fasting and 2hr glucose, proinsulin, HbA1C. Further fasting insulin is positively correlated with LDL triglycerides and liver biomarkers as well as BMI, WHR and blood pressure (**Figure 2.4**).

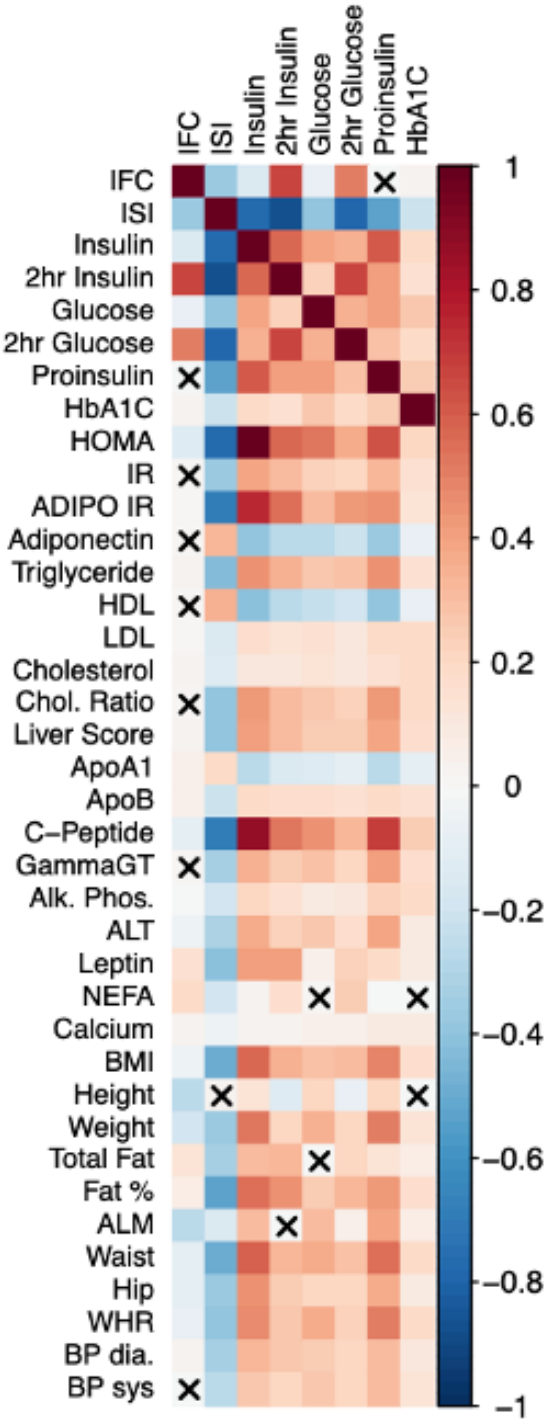


Figure 2.4: Observational Correlation (pairwise Spearman's Rank) of Insulin Fold Change and Modified Stumvoll ISI with metabolic traits in the Fenland Study.
Red shades denote positive correlation, blue shared denote negative correlation between trait pairs. X denotes no significant correlation (P > 0.05).

The RISC Study

I conducted similar analyses in nondiabetic participants in the RISC cohort (N=1,138; **methods**) where hyperinsulinemic-euglycemic clamp and OGTT-derived measures of insulin sensitivity are available. I assessed and compared the correlation of clamp and

OGTT-derived measures of insulin sensitivity with relevant traits, as was done in the Fenland Study above. This replicated the strong positive correlation of IFC with insulin at 120min during an OGTT (rho = 0.725, P = 4.21×10^{-177}) and identified a weak negative correlation of IFC with fasting insulin during an OGTT (rho = -0.165, P = 4.57×10^{-8}) as well as no correlation with insulin taken 20min before clamp (rho = 0.009, p =0.768) (**Figure 2.5**).

I further assessed the correlation of OGTT derived ISI and IFC with the clamp-based measure of insulin sensitivity (M/I) and identified a significant observational correlation of both OGTT-derived traits with this clamp-derived measure of insulin sensitivity in the RISC study (ISI vs M/I Spearman's rho = 0.522, P = 5.10×10^{-76}; IFC vs M/I rho = -0.180, P = 3.29×10^{-9}) (**Figure 2.5**), and identified we are able to approximate clamp-based measures with OGTT indices available in a greater number of studies. The weaker correlation with IFC and M/I is expected, with M/I being reflective of whole-body insulin sensitivity and not specifically post-challenge insulin sensitivity which IFC captures more explicitly.

For ISI there is an inverse correlation with insulin at fasting (rho = -0.68, P = 6.38×10^{-149}) and 120mins during an OGTT (rho = -0.869, P =0), as seen in the Fenland Study. Overall correlations were similar and consistent in direction between RISC and Fenland studies for primary overlapping traits of interest.

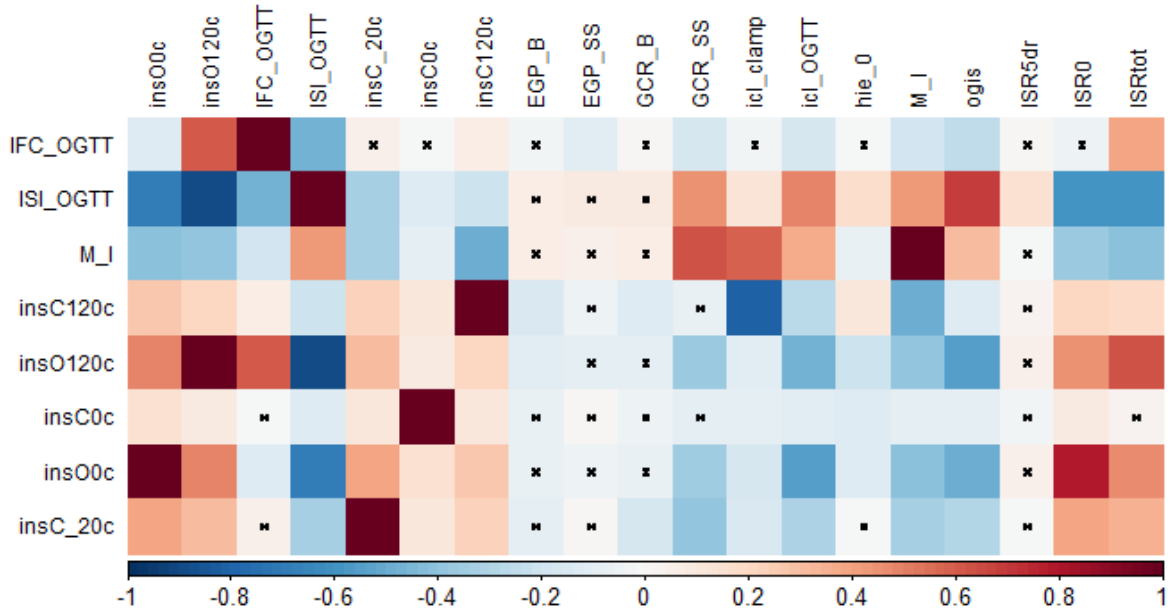


Figure 2.5: Observational Correlation (pairwise Spearman's Rank) of insulin sensitivity and clearance related traits in the RISC cohort.
Red shades denote positive correlation, blue shared denote negative correlation between trait pairs, legend along bottom of heatmap shows colour scale relative to rho value. X denotes no significant correlation (P > 0.05). Abbreviations denote: IFC_OGTT – IFC calculated using OGTT measures. ISI_OGTT – ISI. M_I – M/I index of insulin sensitivity. InsC0c – insulin measured at 0min during clamp. InsO0c - insulin measured at 0min during OGTT. InsC120c – insulin measured at 120min during clamp. InsO120c - insulin measured at 120min during OGTT. InsC_20c – insulin measured at 20min before clamp. EGP_B - basal glucose production, EGP_SS - glucose production during clamp, GCR_B - basal glucose clearance, ml/min/kg lean body mass, GCR_SS - steady state glucose clearance, ml/min/kg lean body mass, icl_clamp - peripheral insulin clearance (1/min/m²), icl_OGTT - endogenous "pre-hepatic" clearance during the OGTT, hie_0 - hepatic insulin extraction during clamp, OGIS - oral glucose insulin sensitivity index (ml min-1 m-2). ISR5dr - insulin secretion 5 mM glucose, beta cell dose response (pmol min-1m-2). ISR0 basal insulin secretion (pmol min-1m-2). ISRtot - total insulin secretion (nmol m-2).

Correlation between IFC and ISI

ISI and IFC are moderately correlated in Fenland (IFC vs ISI Spearman's rho = -0.375, $P < 1x10^{-300}$) and RISC studies (IFC vs ISI Spearman's rho = -0.458, $P = 3.14x10^{-57}$), suggesting that these capture in part similar phenotypes. However, this also suggests there is value in studying both of these with these additionally capturing differing aspects of post-challenge insulin resistance.

2.4.3 Genetic discovery of post-challenge insulin resistance

Primary results: European cohorts, BMI adjusted

Meta-analysis of GWAS identified 11 genome-wide significant loci across both traits ($P < 5x10^{-8}$), 9 of which were novel for post-challenge insulin resistance, as measured by ISI and IFC adjusted for BMI, in individuals without diabetes from 25 studies of European ancestry (**See methods, Figure 2.6**).

Four of these loci, all novel and not previously implicated in post-challenge insulin resistance, were significantly associated with IFC (N = 52,474; *SLC2A4, PPP1R3B, C2CD4A* and *MTNR1B;* **Figure 2.6 and Table 2.5**). The remaining loci were associated with ISI, including 6 novel loci not previously implicated in post-challenge insulin resistance: *MTOR, COBLL1, PPARG, C5orf67, FAM101A, SLC2A4*. This replicated 2 previously-identified ISI loci at *IRS1* and *BCL2* (**Figure 2.6 and Table 2.5**).[238] The *SLC2A4* (rs117643180) locus was the only locus significantly associated with both ISI and IFC. Manhattan plots of BMI adjusted and unadjusted results for both ISI and IFC in European cohorts, and associated QQ plots are outlined in **Supplementary Figures S2.1-S2.4**. The opposing direction of effect of these loci on IFC and ISI is expected, with higher ISI reflecting greater insulin sensitivity rather than greater insulin resistance.

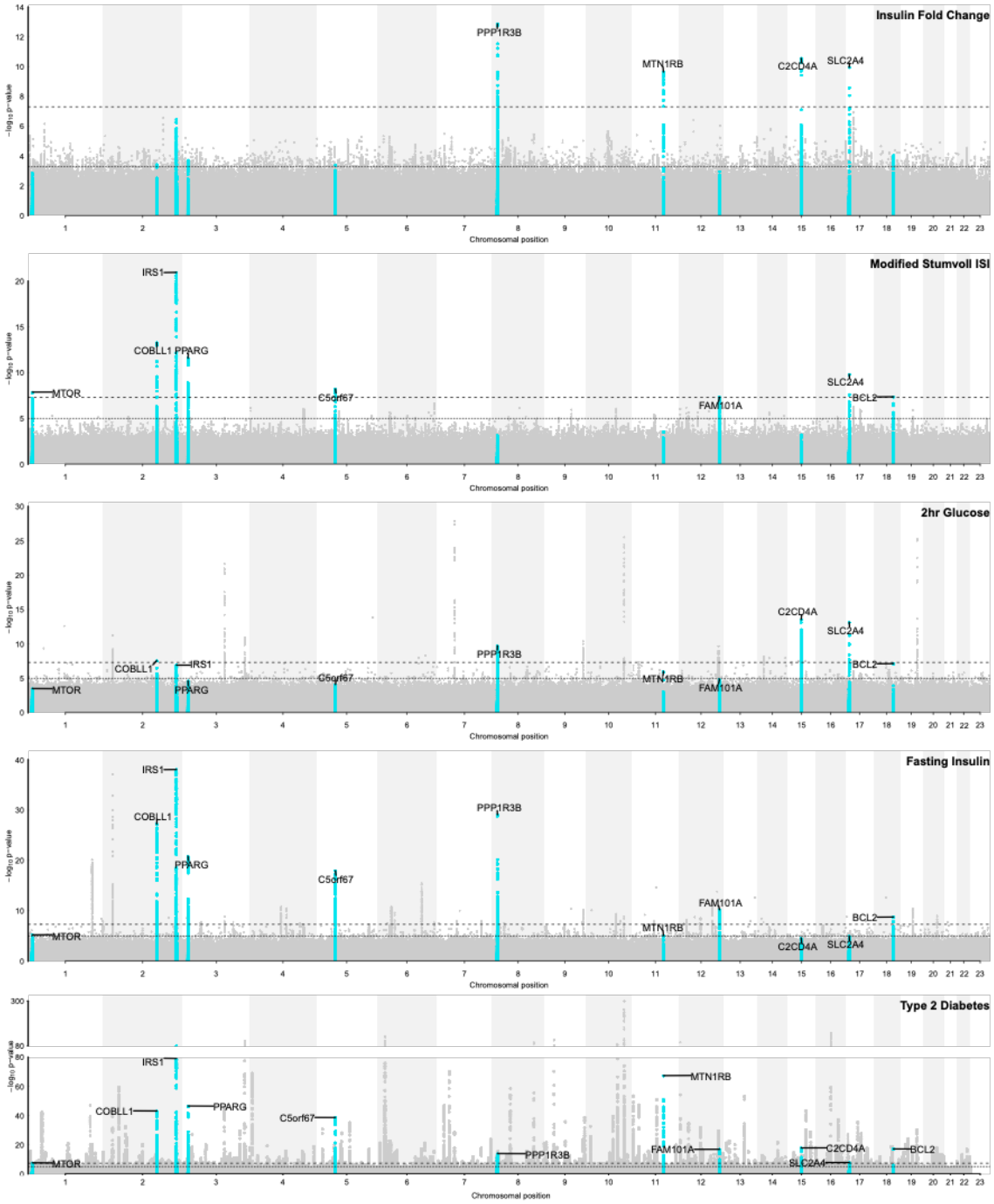


Figure 2.6: Loci associated with post-challenge insulin resistance at genome-wide significance overlap with associations with other glycaemic traits and T2D

Y-axis denotes the -$\log_{10}$(p-value) of association for each trait. The dashed line defines genome-wide significance threshold ($P < 5x10^{-8}$), and the dotted line denotes a suggestive significance threshold of $P = 1x10^{-5}$. Points highlighted in blue on each plot represent those that meet genome-wide significance ($P < 5x10^{-8}$) as well as additional loci prioritised using integrated approach for ISI and IFC. Labels for indicate the nearest gene at a locus defined at genome-wide significance or through the integrated approach. IFC and ISI statistics are from the meta-analysis of GWAS adjusted for BMI from cohort of European ancestry. Fasting insulin and 2 h glucose statistics are those reported by Chen et al (2021)[87] in European studies. T2D statistics are those reported by Vujkovic et al (2020)[251] in studies of European ancestry. For T2D y axis is cut at $1x10^{-100}$ for clarity, with the minimum p value truncated to P = $1x10^{-}$

Table 2.5 Genome-wide significant loci for insulin fold change and/or Modified Stumvoll ISI in meta-analysis of cohorts of European ancestry

							BMI adjusted				BMI unadjusted			
Trait	Locus	MarkerName	rsid	EA	OA	EAF	Beta	SE	P-value	N	Beta	SE	P-value	N
Insulin Fold Change	MTN1RB	11:92673828:C_T	rs1387153	T	C	0.3023	0.0438	0.0069	2.25E-10	52698	0.0431	0.0069	4.33E-10	52745
	PPP1R3B (primary)	8:9173358:A_G	rs7012814	A	G	0.4796	0.0481	0.0065	1.35E-13	52458	0.0483	0.0065	1.07E-13	52496
	PPP1R3B (secondary)	8:9185146:C_T	rs2126259	T	C	0.1265	-0.0688	0.0099	2.86E-12	52698	-0.0706	0.0099	6.12E-13	52745
	C2CD4A	15:62396189:A_C	rs7167878	A	C	0.4606	-0.0426	0.0064	2.76E-11	53287	-0.0424	0.0064	3.41E-11	53334
	SLC2A4	17:7185779:A_C	rs117643180	A	C	0.0231	0.1428	0.0221	1.11E-10	50671	0.1453	0.0221	5.26E-11	50708
Modified Stumvoll ISI	IRS1	2:227101411:A_G	rs2972144	A	G	0.366	0.0644	0.0067	1.10E-21	50511	0.0458	0.0066	5.31E-12	52872
	COBLL1	2:165558252:C_T	rs12692738	T	C	0.7724	-0.0568	0.0075	5.24E-14	53657	-0.0351	0.0076	3.44E-06	53710
	PPARG	3:12385828:A_G	rs11128603	A	G	0.8595	-0.0635	0.0091	2.64E-12	53657	-0.0389	0.0092	2.33E-05	53710
	SLC2A4	17:7185779:A_C	rs117643180	A	C	0.0231	-0.1416	0.0221	1.58E-10	51187	-0.1251	0.0222	1.67E-08	51230
	C5orf67	5:55806751:A_G	rs459193	A	G	0.2816	0.0409	0.007	6.18E-09	53657	0.0348	0.007	7.95E-07	53710
	MTOR	1:11322628:G_T	rs2295080	T	G	0.6961	-0.0394	0.0069	1.34E-08	53657	-0.0243	0.0069	0.000469	53710
	BCL2	18:60845884:C_T	rs12454712	T	C	0.5858	-0.0374	0.0068	4.42E-08	52701	-0.0231	0.0068	0.0007387	52754
	FAM101A	12:124539537:A_C	rs1906937	A	C	0.2894	-0.0384	0.007	4.86E-08	53657	-0.0313	0.007	8.99E-06	53710

Identification of independent genetic signals

I conducted exact conditional analysis revealed 2 distinct signals at *PPP1R3B* for IFC, with no additional signals identified at other loci (**Figure 2.7; Table 2.5**).

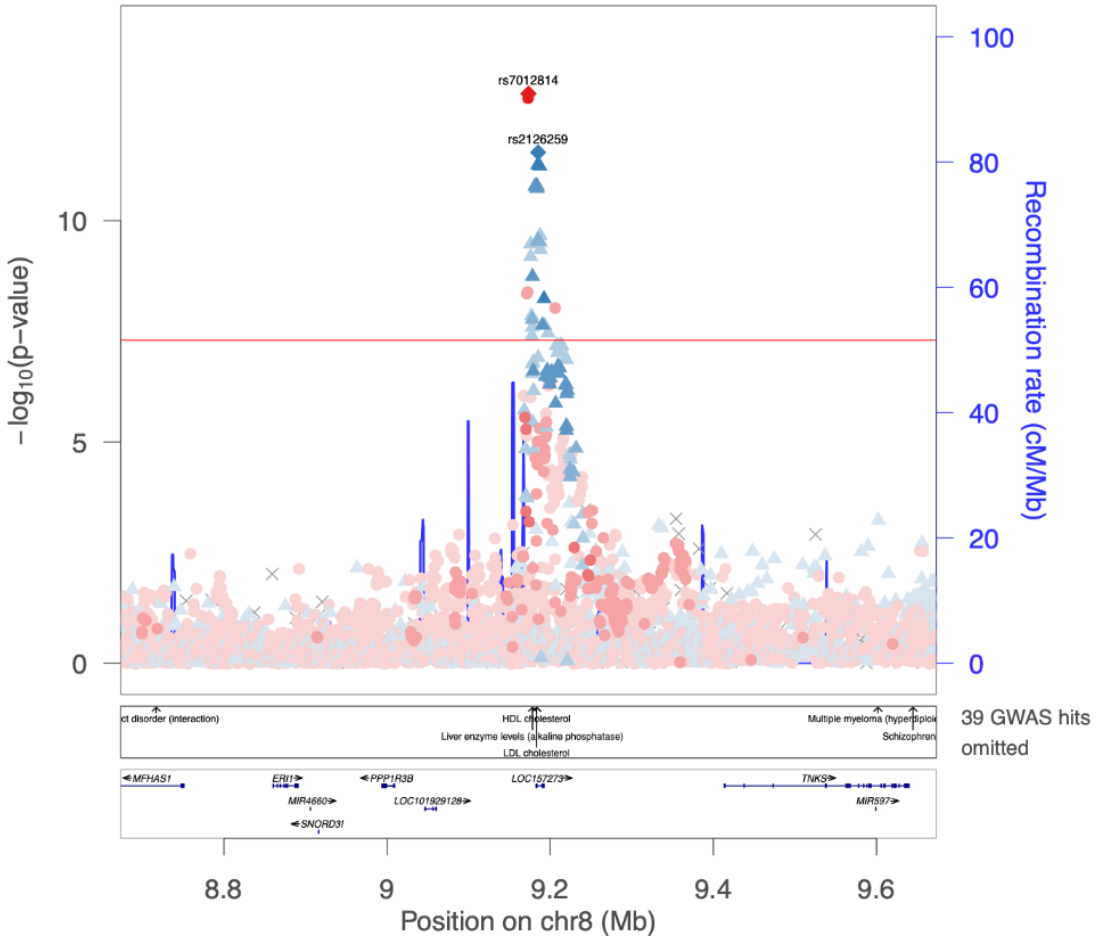


Figure 2.7: Two independent signals were identified at *PPP1R3B* for insulin fold change adjusted for BMI.
Conditional analyses (GCTA-COJO) using Fenland-OMICs as an LD reference panel, identifies a second independent signal at PPP1R3B for insulin fold change adjBMI. The regional association plot shows the primary signal in red and the secondary signal in blue for marginal summary statistics for insulin fold change adjBMI.

Fine-mapping: Defining 99% credible sets

I conducted fine-mapping to define 99% credible sets that accounted for ≥99% of the posterior probability of association (PPA) for each of the 12 independent association signals (**Tables 2.6 and 2.7**). At *SLC2A4* the 99% credible set included a single variant for ISI and 2 variants for IFC, with high confidence that lead variant rs117643180 is the candidate causal variant (IFC PPA for rs117643180 = 0.94; ISI PPA for rs117643180 = 0.98).

Table 2.6: 99% credible sets defined by fine mapping for Insulin Fold Change adjBMI in European only meta-analysis.

Locus	Lead variant at Locus	N variants selected in 99% credible set (including lead variants)	N cred. Set variants with Rsq≥0.8 with lead variant	N cred. Set variants with Rsq< 0.8 & ≥0.5 with lead variant	N cred. Set variants with Rsq<=0.5 with lead variant	Summary statistics used IFC adjBMI European Only	PPA Lead variant
C2CD4A	rs7167878	13	13	0	0	marginal	0.108
MTN1RB	rs1387153	32	8	16	1	marginal	0.127
SLC2A4	rs117643180	2	0	1	0	marginal	0.938
PPP1R3B (primary)	rs7012814	23	3	0	20	Conditioning on secondary signal	0.528
PPP1R3B (secondary)	rs2126259	45	24	18	4	Conditioning on Primary signal	0.36

Table 2.7: 99% credible sets defined by fine mapping Modified Stumvoll Insulin Sensitivity Index adjBMI in European only meta-analysis.

Locus	Lead Variant at this locus	N variants selected to be in 99% credible set (including lead)	N cred. Set variants with R^2≥0.8 with lead variant	N cred. Set variants with R^2< 0.8 & ≥0.5 with lead variant	N cred. Set variants with R^2<=0.5 with lead variant	PPA lead variant
BCL2	rs12454712	19	1	0	18	0.581
C5orf67	rs392794	19	19	0	0	0.0927
COBLL1	rs12692738	9	1	0	8	0.156
IRS1	rs2972144	44	44	0	0	0.296
MTOR	rs2295080	2	2	0	0	0.484
PPARG	rs11128603	24	24	0	0	0.104
SLC2A4	rs117643180	1	1	0	0	0.984
FAM101A	rs1906937	98	1	0	97	0.043

BMI unadjusted genetic discovery

To fully assess the effect of adiposity, I repeated analyses not adjusting for BMI. IFC associations were largely unaffected by BMI adjustment (effect size R^2 BMI adjusted vs BMI unadjusted = 0.98). In contrast, for ISI, adjusted and unadjusted beta estimates were moderately correlated (Pearson correlation between betas = 0.78), but significance levels were weaker in BMI unadjusted models, as previously reported for other glycaemic traits[87], with only *IRS1* and *SLC2A4* remaining significant at $p<5x10^{-8}$ (**Figure 2.8**).

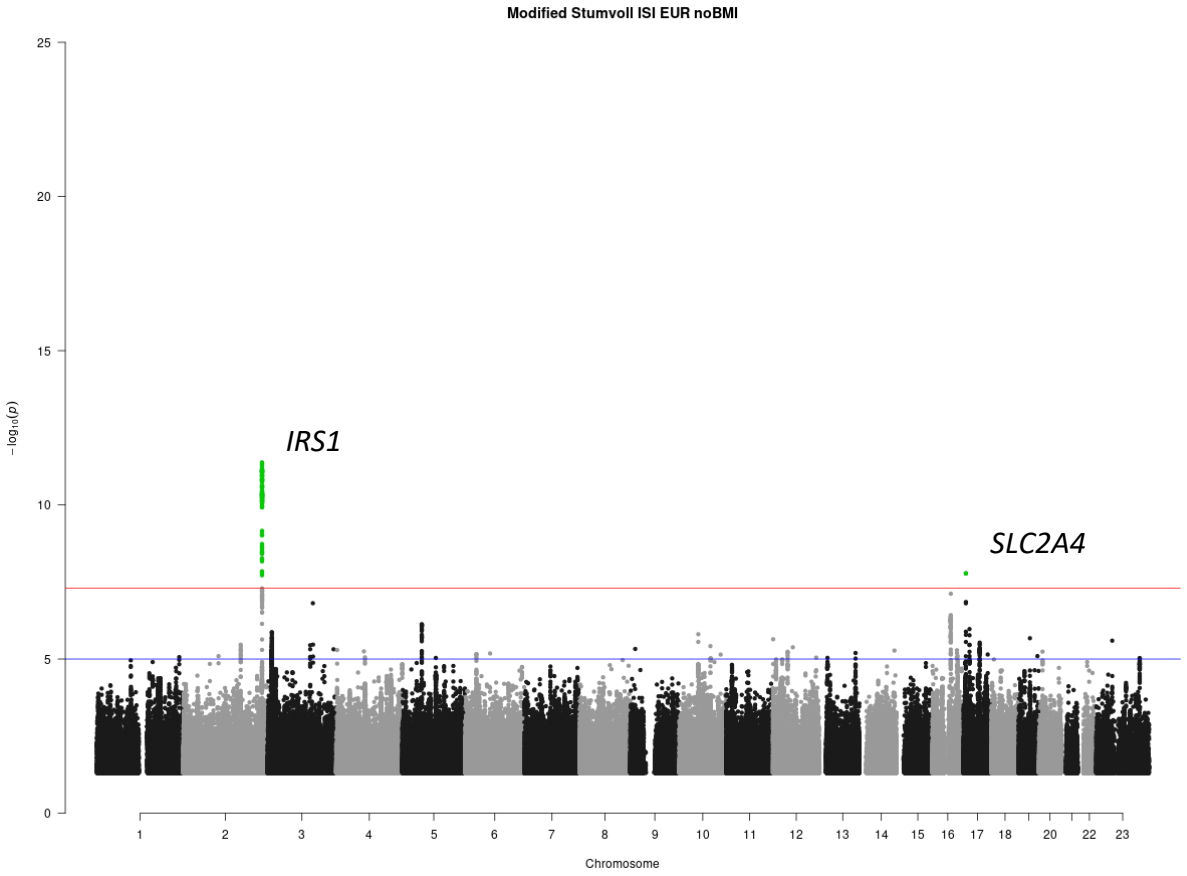


Figure 2.8 Manhattan plot of genetic associations with ISI in European only meta-analysis unadjusted for BMI.
Green represents variants that meet genome-wide significance. Labels represent genome-wide significant loci. Red line represents genome-wide significance threshold of P = $5x10^{-8}$, blue line represents suggestive significance threshold of P = $5x10^{-5}$.

Multi-ancestry genetic discovery

Meta-analyses of GWAS from all ancestry groups (European, East Asian, and Hispanic American) did not reveal any additional associations, with all 11 loci identified in the European only analyses remaining genome-wide significant ($P < 5x10^{-8}$). Further, no

genome-wide significant associations were identified in ancestry-specific analyses of cohorts of Hispanic American or East Asian ancestry for IFC or ISI.

Therefore, to increase power in studying cohorts of non-European ancestry, I conducted a random effects meta-analysis of studies of Hispanic American and East Asian ancestry. While no genome-wide significant ($P < 5x10^{-8}$) loci for ISI were identified in this meta-analysis of non-European cohorts, a single genome-wide significant locus for IFC was identified.

BICC1 locus

BICC1 (rs60453193, beta = 0.43, SE = 0.08, $P = 4.06x10^{-8}$, N = 1,837) was identified as a signal associated with IFC in a meta-analysis of GWAS in participants of Hispanic American and East Asian ancestry **(see Methods; Figure 2.9).** This locus has not been implicated in post-challenge insulin resistance previously. The effects were consistent in meta-analyses of East Asian (beta = 0.406, SE = 0.204, P = 0.046, N = 739) and Hispanic American ancestry studies (beta = 0.435, SE = 0.085, $P = 3.15x10^{-7}$, N = 1,098). No one study seems to be driving this association with effect sizes being comparable across all 3 studies of non-European ancestry (**Figure 2.10**).

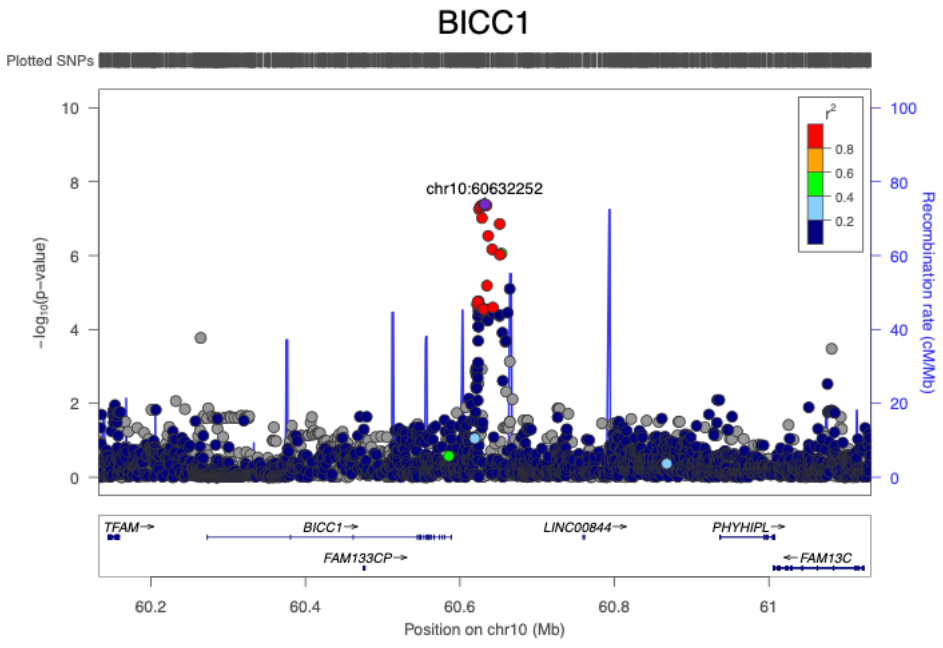


Figure 2.9: Regional association plot at BICC1 for IFC adj.BMI in non-European meta-analysis.
Colour of points represents related linkage disequilibrium of a variant with the lead variant at this locus: rs60453193 at chr10:60632252, as indicated by r^2. Relative thresholds of r^2 are summarised on the right hand side of the plot, using the 1000 Genomes ASN LD reference panel.

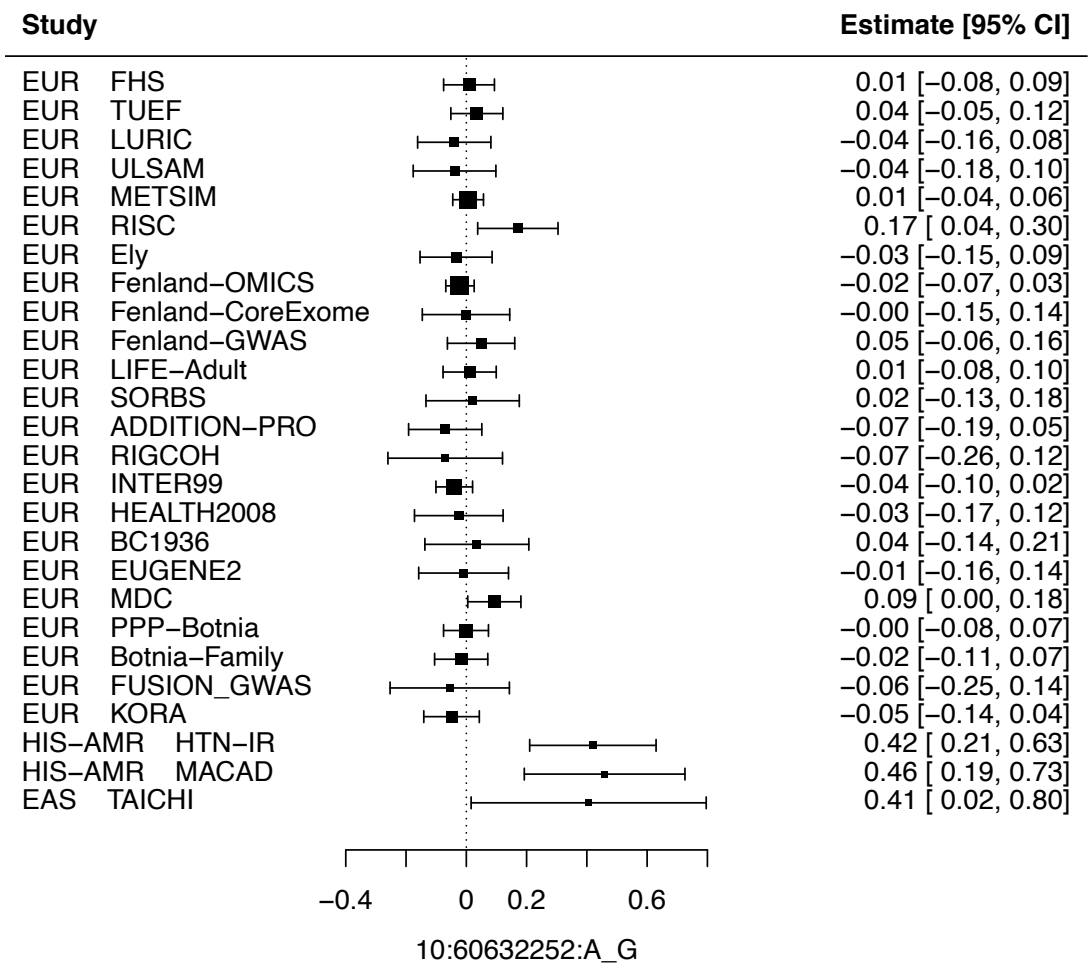


Figure 2.10: Forest plot of beta estimates of associations of rs60453193 with insulin fold change in individual cohorts.
Labels on the right-hand side indicate the ancestry of the study and study name. EUR- European ancestry, HISAMR – Hispanic American ancestry, EAS – East Asian ancestry. Left-hand side values are beta estimate and 95% confidence interval. Error bars denote a 95% confidence interval. X-axis denotes the beta estimate of associations with insulin fold change adjBMI.

BICC1 was not associated with IFC meta-analysis of studies only of European ancestry (beta = -0.0026, SE = 0.010, P = 0.791, N = 50,671); despite being common in all ancestry groups (MAF European = 12%, Hispanic American =7.8%, East Asian = 4.2%). Variants within this locus (±500kb of rs60453193) have been reported to be associated with T2D at suggestive significance in multi-ancestry analyses[251] (Minimum P = $1.0 \times 10^{-4.7}$), however rs60453193 itself is not associated (MVP multi-ancestry T2D: Beta = 0.0079, SE = 0.0062, P = 0.020).

Ancestry-specific loci with such effects have been identified previously for glycaemic traits, and it is possible that heterogeneous effect sizes despite similar EAF may be due to genotype-by-environment interactions or other epistatic effects.[87] Further, the

variants detected here may be in LD with ancestry specific causal variants that were not investigated in this study and differ in frequency across ancestries.[87]

Rare, damaging variants in *BICC1* nominally associated with T2D in individuals of European ancestry.[253] *BICC1* was not statistically significant in the multi-ancestry analyses of IFC, including all cohorts (beta = 0.26, SE = 0.18, P = 0.16, N = 52,508). In phenome-wide whole exome analyses in UK Biobank[253], burden of deleterious variants in *BICC1* have been reported to be associated suggestively with several kidney-related disorders and increased creatinine and urate, which are known to be present in diabetic kidney disease.

2.4.3 Assessing overlap of post-challenge insulin resistance with type 2 diabetes

There is strong evidence of a shared genetic signal between post-challenge insulin resistance and T2D[251] for 10 of the 11 loci IFC and ISI associated loci identified in studies of European ancestry, using statistical colocalisation method coloc (posterior probability of colocalisation > 0.7; **Table 2.8**), providing strong evidence of a shared genetic signal underlying both post-challenge insulin resistance and T2D at each of these loci. Of the 11 loci, 8 loci were associated with T2D at genome-wide significance and 2 loci at suggestive significance threshold of $P < 1x10^{-4}$ (**Table 2.9**).

Table 2.8 - colocalisation of T2D with IFC/ISI at post-challenge insulin resistance genetic loci.

Trait	Locus		rsid	PP abf (H4)	Summary statistics used for IFC/ISI and T2D colocalisation
IFC	chr8_9173358_A_G	PPP1R3B	rs7012814	80.40%	marginal
IFC	chr15_62396278_C_T	C2CD4A	rs7167932	98.70%	marginal
IFC	chr17_7185779_A_C	SLC2A4	rs117643180	4.36%	marginal
IFC	chr11_92673828_C_T	MTNR1B	rs1387153	95.20%	marginal
ISI	chr2_227101411_A_G	IRS1	rs2972144	96.20%	marginal
ISI	chr2_165555539_A_G	COBLL1	rs79953491	98.20%	ISI: marginal. T2D condition on secondary signal rs13389219
ISI	chr3_12385828_A_G	PPARG	rs11128603	98.10%	marginal
ISI	chr17_7185779_A_C	SLC2A4	rs117643180	4.55%	marginal
ISI	chr5_55806751_A_G	C5orf67	rs459193	98.70%	marginal
ISI	chr1_11322628_G_T	MTOR	rs2295080	85.10%	marginal
ISI	chr18_60845884_C_T	BCL2	rs12454712	99.90%	marginal
ISI	chr12_124539537_A_C	FAM101A	rs1906937	92.40%	marginal

Table 2.9 – Reported associations of post-challenge insulin resistance loci with type 2 diabetes and other glycaemic traits in cohorts of European ancestry.

						Type 2 Diabetes EUR (Vujkovic et al; 2020)				2hr Glucose adjBMI EUR (Chen et al, 2021)				Fasting Insulin adjBMI EUR (Chen et al, 2021)			
Trait	**Locus**	**rsid**	**EA**	**OA**	**EAF**	**Beta**	**SE**	**P-value**	**N**	**Beta**	**SE**	**P-value**	**N**	**Beta**	**SE**	**P-value**	**N**
	MTN1RB	rs1387153	T	C	0.3023	0.075	0.005	1.05E-50	1042540	0.0066	0.0063	0.2948	63163.8	-0.011	0.0045	0.01401	141953
	PPP1R3B (primary)	rs7012814	A	G	0.4796	-0.003	0.0063	0.6374	835859	0.0354	0.0058	1.04E-09	63395.8	-0.045	0.004	8.34E-30	148942
	PPP1R3B (secondary)	rs2126259	T	C	0.1265	-0.0521	0.0078	1.92E-11	1042540	-0.0567	0.009	2.98E-10	63392.9	0.0538	0.0064	4.76E-17	144629
	C2CD4A	rs7167878	A	C	0.4606	-0.0398	0.0046	4.63E-18	1042540	0.0324	0.0058	2.32E-08	63395.9	0.0018	0.0039	0.6416	151013
IFC	SLC2A4	rs117643180	A	C	0.0231	0.0462	0.0148	0.001837	1041220	0.1893	0.0253	7.31E-14	38302	0.018	0.0147	0.2193	112805
	IRS1	rs2972144	A	G	0.366	0.0899	0.0048	1.42E-79	1042540	-0.0312	0.0059	1.24E-07	63391	-0.0514	0.004	2.28E-38	150996
	COBLL1	rs12692738	T	C	0.7724	-0.0514	0.0074	3.42E-12	835859	0.0378	0.0068	2.72E-08	63386	0.0466	0.0046	2.51E-24	151010
	PPARG	rs11128603	A	G	0.8595	-0.0942	0.007	2.23E-41	1042540	0.0275	0.0082	0.00080	63396	0.052	0.0057	7.12E-20	151013
	SLC2A4	rs117643180	A	C	0.0231	0.0462	0.0148	0.001837	1041220	0.1893	0.0253	7.31E-14	38302	0.018	0.0147	0.2193	112805
	C5orf67	rs459193	A	G	0.2816	0.0467	0.0052	4.42E-19	1042540	-0.0145	0.0066	0.02802	63383	-0.0353	0.0045	3.16E-15	145254
	MTOR	rs2295080	T	G	0.6961	0.0265	0.0068	9.37E-05	835859	0.0157	0.0063	0.0127	63381.8	0.0164	0.0043	0.000124	149740
	BCL2	rs12454712	T	C	0.5858	-0.0412	0.0048	8.30E-18	1039830	0.0355	0.0066	7.50E-08	55370.4	0.03	0.005	1.78E-09	113902
ISI	FAM101A	rs1906937	A	C	0.2894	0.0328	0.0069	2.01E-06	835859	0.0189	0.0072	0.00867	51420.8	0.0182	0.0047	0.000101	136287

The only locus without evidence of colocalisation of the IFC/ISI and T2D signals was at *SLC2A4,* which is nominally associated with T2D in European cohorts (Beta = 0.046, SE = 0.015, P = 0.0018, N = 1,041,220).[251] However, lead variant rs117643180 is in perfect linkage disequilibrium (D' = 1) with lead SNPs at T2D loci identified in East Asian and Hispanic American populations (**Figure 2.11 and 2.12, Table 2.10**).[81,266]

Even though the *SLC2A4* lead variant (rs117643180) was associated with an increased risk of T2D at nominal significance (beta = 0.046, SE = 0.015, P = 0.0018, N = 1,041,200, OR = 1.047)[251], this was the only post-challenge insulin resistance locus identified in European ancestry cohorts without evidence of a shared signal with T2D using formal colocalisation testing. rs117643180 is in perfect linkage disequilibrium (LD; D' = 1) with lead SNPs at T2D loci identified in East Asian and Hispanic American ancestries.[81,266]

Table 2.10 Differences in allele frequency and LD across populations may explain the lack of colocalisation of T2D and ISI/IFC in Europeans.

				LD (POPULATION MATCHED)			MAF GNOMAD 2.1.1[110]				
RSID	**chr**	**pos_b37**	**TopSNP for:**	**IFC lead SNP ancestry Matched**	**D' ancestry matched IFC lead SNP**	**D' rs117643180 ancestry Matched**	**EUR (non-Finnish)**	**EAS**	**AFR/AFR AMR**	**AD-AMR**	**SAS**
rs13342232	17	6945940	T2D MVP ASN[255]	rs1641512	0	1	0.01053	0.09966	0.3354	0.2986	0.00588
rs193399	17	6962272	T2D MVP AFR[255]	rs9890111	0.154	1	0.009007	0.2237	0.1427	0.2612	0
rs858519	17	7531965	T2D MVP EUR[255]	rs117643180	0.269	0.269	0.5732	0.4922	0.2008	0.6026	0
rs13342232	17	6945940	T2D SIGMA[266]	rs9890111	0.1	1	0.01053	0.09966	0.3354	0.2986	0.00588
rs186568031	17	6953781	T2D AGEN EAS[81]	rs1641512	0	1	0.007273	0.1107	0.08837	0.2506	0
rs117643180	17	7185779	MS ISI, IFC, primary eQTL signal for SLC2A4 muscle[154]	rs117643180	1	1	0.02393	0.00423	0.004138	0.00906	0.01201
r222849	17	7185861	Secondary eQTL signal SLC2A4 muscle[154]	rs117643180	1	1	0.644	0.6398	0.8593	0.5979	0

GNOMAD genetic ancestry– EUR – European, EAS – East Asian, AFR – African, AFR AMR – African American, AD-AMR – Admixed American. SAS – South Asian.

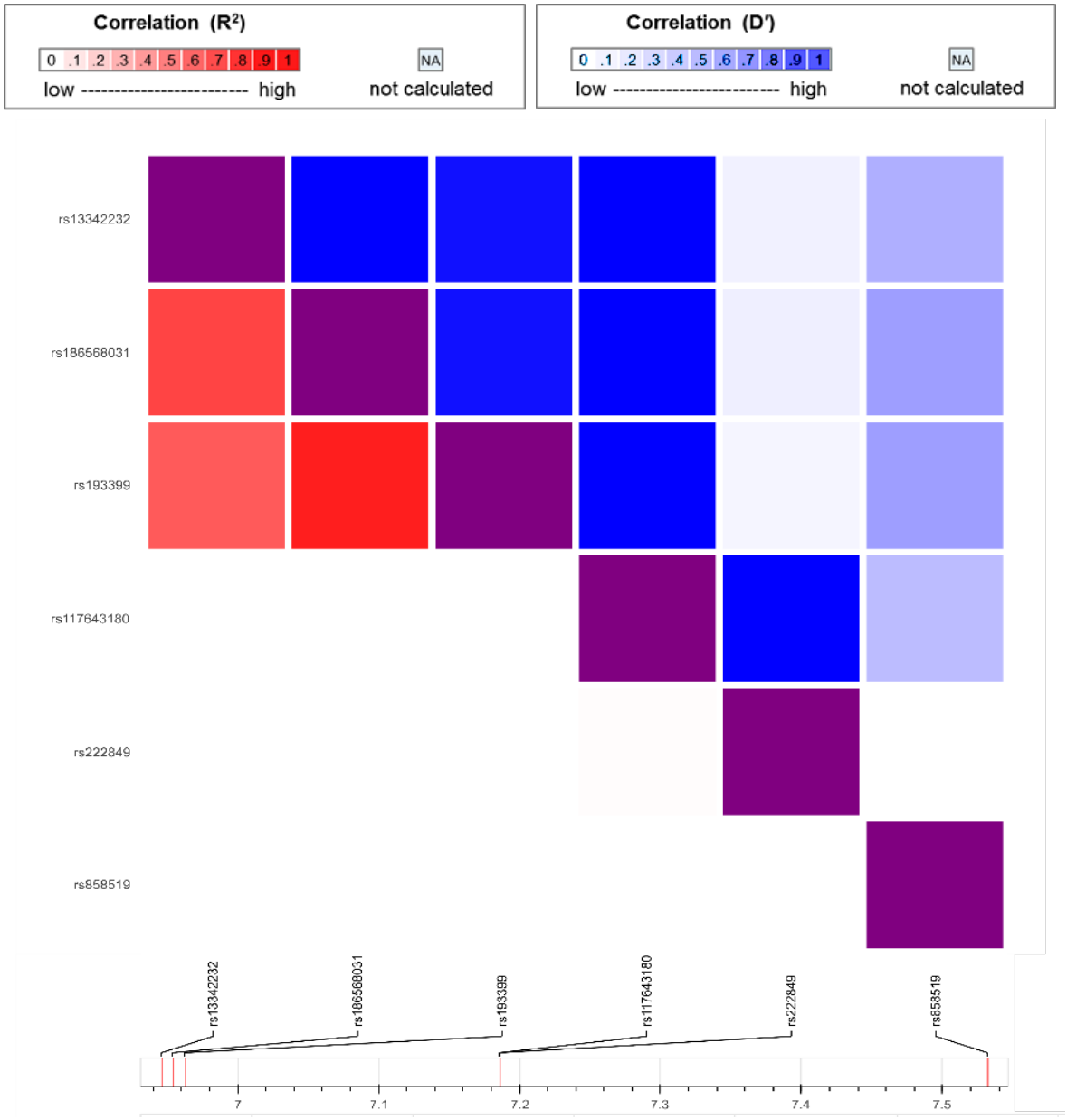


Figure 2.11: LD matrix for lead SNPs in IFC, ISI, and T2D GWAS at SLC2A4 locus in European ancestry

LD matrix generated for individuals of European ancestry using LDLink[267]. Scale for relative linkage disequilibrium of SNP pairs is indicated by scale at the top of the figure with D' in blue and R^2 in red.

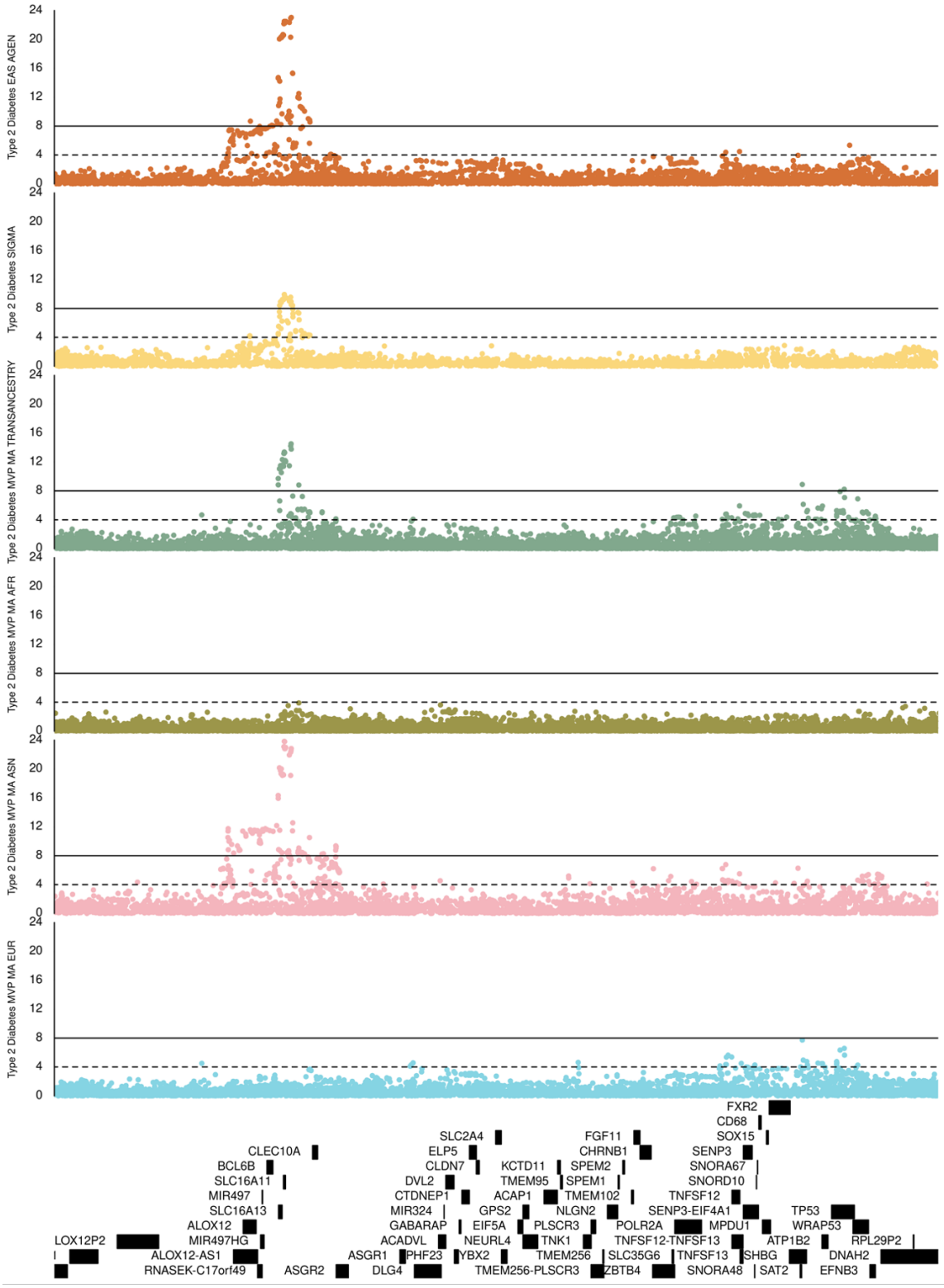


Figure 2.12: SLC2A4 locus regional association plot for Type 2 Diabetes across studies in different ancestry groups.
Locus is defined as +/- 500kb from lead ISI and IFC SNP rs117643180. From top to bottom the tracks denote the following publicly available Type 2 Diabetes meta-analysis results: Track 1: East Asian Ancestry (AGEN; Spracklen et al; 2021)[81]. Track 2: Hispanic American ancestry, SIGMA consortium (Wiliams et al 2014)[266]. Track 3: MVP meta-analysis (Vujkovic et al; 2020)[255] trans-ancestry. Track 4: MVP meta-analysis (Vujkovic et al; 2020)[255] African ancestry. Track 5: MVP meta-analysis (Vujkovic et al; 2020)[255] Asian ancestry. Track 7: MVP meta-analysis (Vujkovic et al; 2020)[255] European ancestry. Track 7: genes located at this locus. Y axis represents the –log10(p-value) of association.

The solid line refers to genome-wide significance threshold of P = $5x10^{-8}$, and the dashed line denotes to suggestive significance threshold of P = $1x10^{-5}$.

The lack of evidence of a shared signal between post-challenge insulin resistance and T2D at *SLC2A4* in cohorts of European ancestry is surprising, since lead variant rs117643180 is located in the first intron of *SLC2A4*, encoding GLUT4, a key player in post-prandial glucose uptake. rs117643180 is also associated with higher 2 h glucose (**Table 2.9**), one of the diagnostic criteria for T2D.[87,92] Several reasons may contribute to lack of colocalisation for example differences in the allele frequency and ancestry composition considered in the respective GWAS efforts.

First, is possible that rs117643180 is the causal variant underlying differences in post-challenge insulin resistance but is not sufficient to cause T2D unless accompanied by beta cell dysfunction.

The respective lead variants at this locus in each ancestry are rare in other ancestries, and therefore may not have been included in previous genetic discovery efforts for T2D or glycaemic traits. For example, in gnomAD (v2.1.1)[110], rs117643180 (*SLC2A4*) has a MAF of 2.3% in non-Finnish European ancestry populations but only 0.42% and 0.91% in populations of East Asian and Hispanic American ancestry, respectively. While formal statistical colocalisation of post-challenge insulin resistance and T2D is not possible in other ancestries due to insufficient sample size available for post-challenge insulin traits. The evidence of LD between T2D-associated variants and post-challenge insulin resistance across different ancestries, demonstrates the utility of examining associations across multiple ancestry groups.

It is further possible that epistatic effects may play a role and the A risk allele of rs117643180 for post-challenge insulin resistance co-segregates with the established nearby protective T2D C allele of rs858519 (T2D association at rs858519: Beta = -0.025, SE = 0.004, P = $1.28x10^{-9}$, EA = C EAF = 0.505)[255]. For instance in 1000 genomes[243], the rs117643180-A risk allele is present at a frequency of 3.1% in European ancestry, and when present, this co-segregates on the same haplotype with the T2D protective C allele at rs858519 approximately 42% of the time.[267]

Rare predicted damaging variants (**see methods**) in *SLC2A4* (MAF < 0.005%, N SNPs = 68) in ~450,000 European ancestry participants from the UK Biobank with whole

exome sequencing have been reported to be nominally associated with increased diabetes risk.[253] Burden of rare predicted damaging missense variants are reported to be nominally associated with increased diabetes risk (UK Biobank field 20002 #1220, OR = 1.79 [1.25-2.68], P = $5.5x10^{-3}$), increased random glucose (Beta = 0.11, SE = 0.05, P =0.03) and increased HbA1c (Beta = 0.13, SE = 0.06, P = 0.018).[253]

2.4.4 Genetic risk scores and LD score regression analysis

Genetic risk scores for post-challenge insulin resistance

Increased genetic risk of post-challenge insulin resistance, represented by genetic risk scores (GRS; **see methods**) of ISI and IFC, was associated (P < 0.0004, Bonferroni) with increased T2D risk, lower waist to hip ratio (WHR) adjusted for BMI, and increased HBA1c, weight, BMI, and body fat percentage in European participants of UK Biobank (UKBB; N max = 351,987; **Supplementary Table S2.1**). In a subset of the deeply phenotyped Fenland Study (N max = 8,925), IFC GRS was at least nominally associated (P < 0.05) with higher levels of fasting and post-challenge glucose, HbA1C, circulating levels of several lipid-related measures, and 2 h insulin, but lower levels of fasting insulin. The ISI GRS was at least nominally associated (P < 0.05) with lower levels of fasting insulin and post-challenge glucose and insulin levels (**Supplementary Table S2.2**).

LD score regression analysis

LD score regression analyses was used to assess genome-wide genetic correlations of ISI and IFC with a wider range of related biochemical and cardiometabolic traits. There are strong genetic correlations of both ISI and IFC with 2h glucose, consistent with insulin resistance and the observational correlations in the Fenland Study (IFC and 2 h glucose: genetic correlation (rg) = 0.713, SE = 0.089, P = $2.63x10^{-22}$; ISI and 2 h glucose: rg = -0.820, SE = 0.048, P = $2.12x10^{-22}$; **Supplementary Table S2.3 and S2.4**). In contrast to ISI, which was strongly correlated with fasting insulin levels, IFC showed only a weak correlation, highlighting greater specificity as an indicator of post-challenge insulin resistance. Both ISI and IFC showed only moderate genetic correlations with T2D, consistent with the fact that multiple pathways contribute to its genetic architecture, with postprandial insulin resistance being only one of several aetiologies (**Supplementary Table S2.3 and S2.4**).

2.4.5 In silico and *in vitro* characterisation of the *SLC2A4* locus (rs117643180)

Association of SLC2A4 with other glycaemic traits

The only locus identified at genome-wide significance for both ISI and IFC was *SLC2A4* (rs117643180). The lead variant at this locus, rs117643180, is located in the first intron of *SLC2A4*. rs117643180 was associated with post-challenge insulin resistance and was previously reported to be associated with higher 2 h glucose (Beta = 0.19, SE = 0.025, P = 7.31×10^{-14}, N = 38,302)[87] but not fasting insulin (P = 0.22)[87] or fasting glucose (P = 0.14)[87] suggesting the effect of this locus is specific to the postprandial state (**Figure 2.13**). This is, to our knowledge, the first time an insulin resistance trait association has been reported at this locus.

In humans, *SLC2A4* expression is highest in skeletal muscle[156] and rs117643180 is the lead SNP at a primary expression quantitative trait locus (eQTL) for *SLC2A4* expression in skeletal muscle, with the effect allele (A) associated with lower expression (GTEx v8, European ancestry samples (EUR); NES (normalised effect size) = -0.50, SE = 0.073, P = 2.51×10^{-11}).[154] These data suggested altered *SLC2A4* expression may underlie the association of rs117643180 with post-challenge insulin resistance and its post-challenge specific glycaemic effect is due to impact on expression of *SLC2A4* [155,156]. Collaborators (Prof Karen Mohlke, University of North Carolina) demonstrated that, rs117643180 is additionally in a region of accessible muscle chromatin since it overlaps with enhancer or promoter annotations in skeletal muscle, and adipose tissues (**Figure 2.13c**) and lies within a region of high conservation within intron 1 of *SLC2A4*. Together, these features suggest that rs117643180 is in a region of importance for *SLC2A4* gene regulation.

There is strong colocalisation of *SLC2A4* eQTL signal at rs117643180 in skeletal muscle with both IFC and ISI when conditioning on the secondary eQTL signal (rs222849) at this locus (posterior probability of colocalisation IFC/eQTL = 1.00, posterior probability of colocalisation ISI/eQTL = 1.00; **Table 2.11, Figure 2.14 and 2.15**). This colocalisation is specific to the primary eQTL signal for *SLC2A4* at rs117643180, the same genetic signal associated with IFC and ISI (**Table 2.11**).

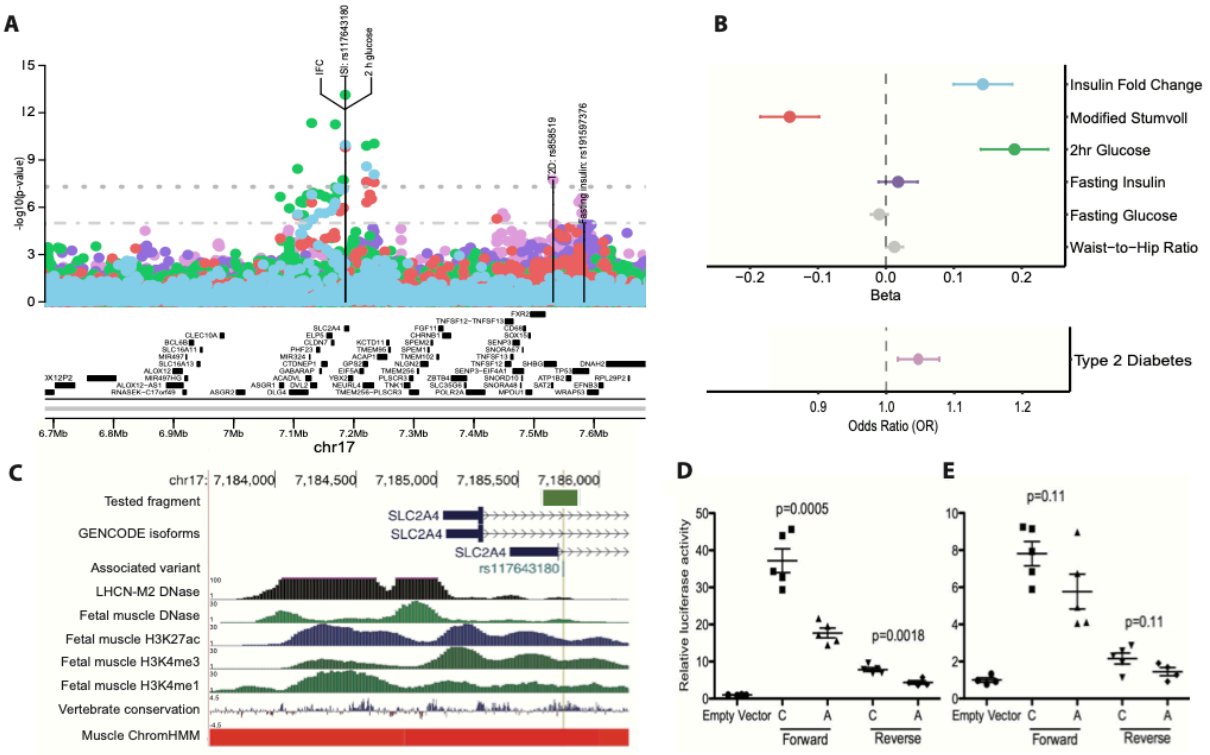


Figure 2.13: rs117643180 at *SLC2A4* affects GLUT4 expression and transcription regulation
A) Regional association plot shows associations at the *SLC2A4* locus with key traits of interest: insulin fold change (IFC) - blue, Modified Stumvoll (ISI) – red, and 2 h glucose[87] – green, fasting insulin[87] – purple, type 2 diabetes[251] – pink .Regional association plot showing ± 500kb around the lead IFC-associated variant, with points indicating the association and location of individual variants. Y-axis denotes the $-\log_{10}$(p-value) of association. The dashed lines and dotted line indicate genome-wide ($P = 5x10^{-8}$) and suggestive ($P = 1x10^{-5}$) significance threshold, respectively. X-axis denotes the genomic position. All association statistics are from BMI adjusted analyses in studies of European ancestry**.** Labels indicate the lead variant for each trait, with location indicating position. Integrated lead is the lead variant defined for this locus using the integrated GWAS approach (see below). B) **Association of rs117643180 with key metabolic traits of interest.** Plot shows beta estimate or odds ratio of association for rs117643180 with insulin fold change (IFC) - blue, Modified Stumvoll (ISI) – red, and 2 h glucose[87] – green, fasting insulin[87] – purple, type 2 diabetes[251] – pink, all included in a) as well as fasting glucose[87] and waist-to-hip ratio[256] plotted in grey. All association statistics are from BMI adjusted analyses in studies of European ancestry. Error bars indicate 95% CI. **C) Regulatory annotation shows the *SLC2A4* locus overlaps with a transcriptionally active region in muscle tissues**. rs117643180 is located ~730 bp downstream of the main *SLC2A4* transcription start site (TSS) and approximately 330 bp downstream of an alternate *SLC2A4* TSS. The 229-bp region flanking the variant, as shown by the orange bar on the UCSC Genome Browser, overlaps with accessible chromatin in LHCN-M2 from the ENCODE consortium, chromatin marks of regulatory activity in fetal muscle leg tissue from the Roadmap Epigenomics Consortium, and a region conserved across species. **D-E) rs117643180 exhibits allelic differences in transcriptional activity.** 229-bp fragments including the rs117643180 C or A alleles were cloned upstream of a minimal promoter driving luciferase expression in the forward and reverse orientations with respect to the promoter. Values represent fold-change of firefly luciferase/*Renilla* activity normalized to empty pGL4.23 vector in C2C12 cells (D) and LHCN-M2 myotubes (E). Error bars represent the SEM of four or five independent clones tested in duplicate wells. P-values are calculated from two-sided t-tests.

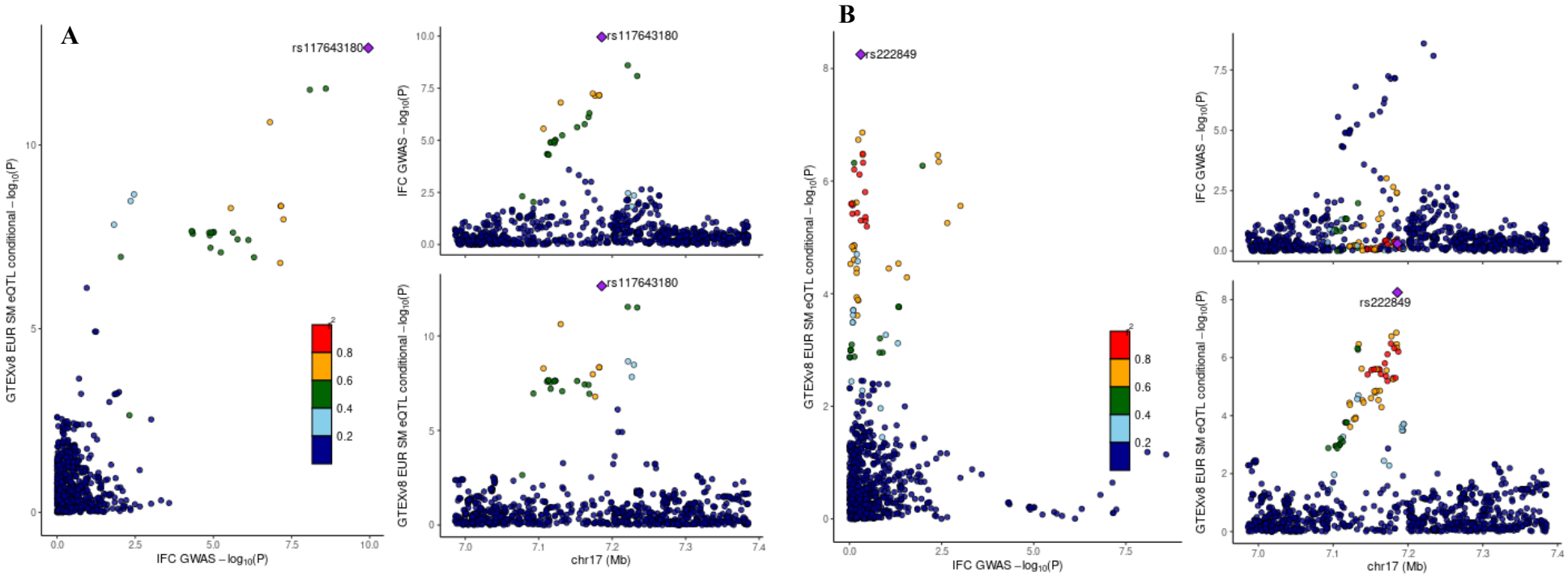


Figure 2.14: statistical colocalisation of IFC with conditional summary statistics for eQTL in skeletal muscle at the SLC2A4 locus.
A) ISI and SLC2A4 eQTL statistics conditioned on the secondary signal rs222849. B) ISI and SLC2A4 eQTL statistics conditioned on the primary signal rs117643180. Locus was defined as +/- 200kb of rs117643180 for colocalisation analyses. The Colour of points, as indicated in the legend, represent R^2 values compared to the lead SNP for each signal annotated with a purple diamond.

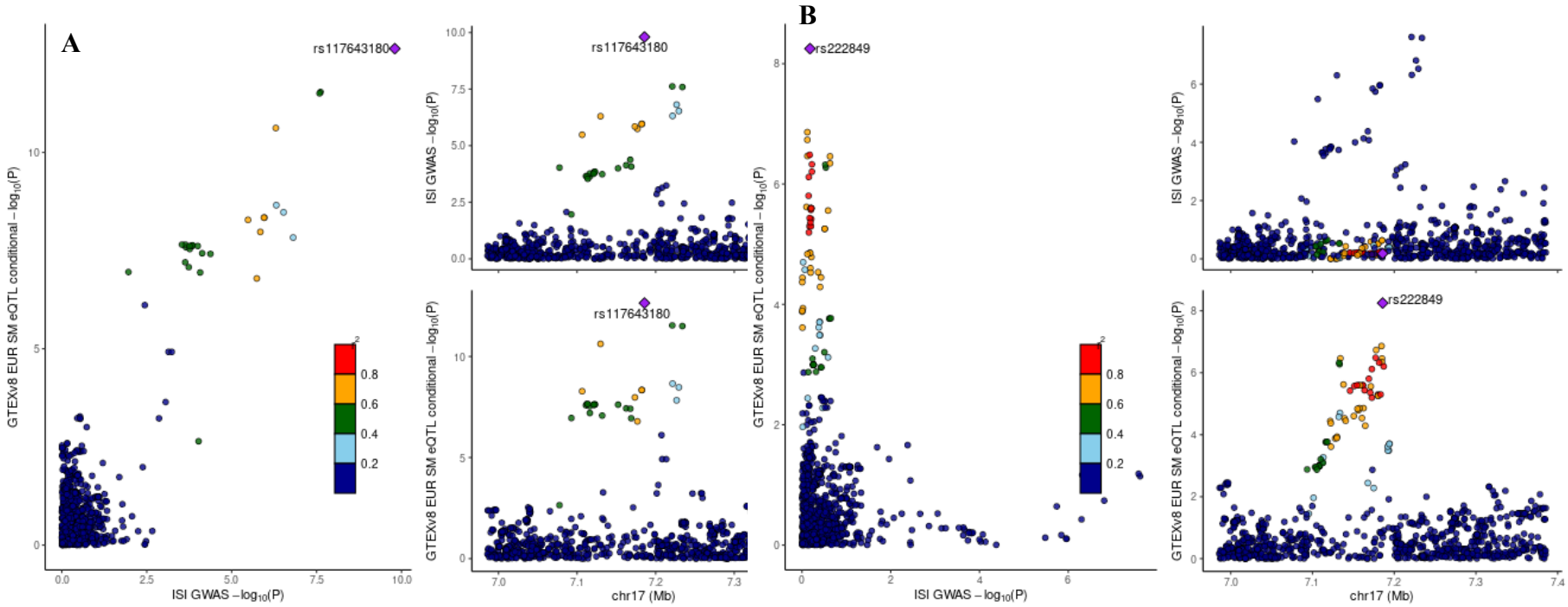


Figure 2.15: statistical colocalisation of ISI with conditional summary statistics for eQTL in skeletal muscle at the SLC2A4 locus.
A) IFC and SLC2A4 eQTL statistics conditioned on the secondary signal rs222849. B) IFC and SLC2A4 eQTL statistics conditioned on the primary signal rs117643180. Locus was defined as +/- 200kb of rs117643180 for colocalisation analyses. The Colour of points, as indicated in the legend, represent R^2 values compared to the lead SNP for each signal annotated with a purple diamond.

Table 2.11: Statistical colocalisation of IFC and ISI with conditional statistics for eQTL for SLC2A4 in skeletal muscle (GTEx v8)

Lead SNP rsid	Trait 1	SLC2A4 eQTL in skeletal muscle GTEx EUR:	nsnps	H0	H1	H2	H3	H4
rs117643180	IFC	conditioned on secondary signal: rs222849	935	1.06E-08	3.29E-07	0.000121	0.00277	0.997109
rs117643180	IFC	conditioned on primary signal: rs117643180	962	6.61E-06	0.001351	0.004857	0.992998	0.000787
rs117643180	ISI	conditioned on secondary signal: rs222849	934	1.54E-08	4.80E-07	0.000119	0.002712	0.997168
rs117643180	ISI	conditioned on primary signal: rs117643180	961	3.82E-05	7.80E-03	0.004825	0.986515	0.000822

Results from collaboration: assessment of the transcriptional regulation of rs117643180

Experiments presented in this section were part of a collaboration and performed by Professor Karen Mohlke's group at University of North Carolina.

To determine whether rs117643180 exhibits allelic differences in transcriptional activity, we performed transcriptional reporter assays in murine C2C12 and human LHCN-M2 muscle cell lines using 229-bp fragments in forward and reverse orientations with respect to a minimal promoter. In all cell types, the fragment showed higher transcriptional activity than an empty vector control, especially in the forward orientation, consistent with a role as a promoter or enhancer (**Figure 2.13D-E, Supplementary Figure S2.5**). In mouse C2C12 myoblasts, the rs117643180-A effect allele showed half the transcriptional activity as the reference C allele in both orientations ($P < 0.002$) (**Figure 2.13D**). Similar results were observed in human LHCN-M2 myoblasts and myotubes (**Figure 2.13E, Supplementary Figure S2.5)**. The lower transcriptional activity observed for the rs117643180-A allele is consistent with the GTEx v8 eQTL analysis of *SLC2A4* in skeletal muscle tissue.[154] In addition, we carried out electrophoretic mobility shift assays using LHCN-M2 nuclear extract and probes spanning the rs117643180 alleles. These assays provided supporting evidence for allelic differences in transcription factor binding (**Figure 2.16**). Taken together, these data suggest that rs117643180 may contribute to allele-specific differences in *SLC2A4* transcriptional activity in skeletal muscle.

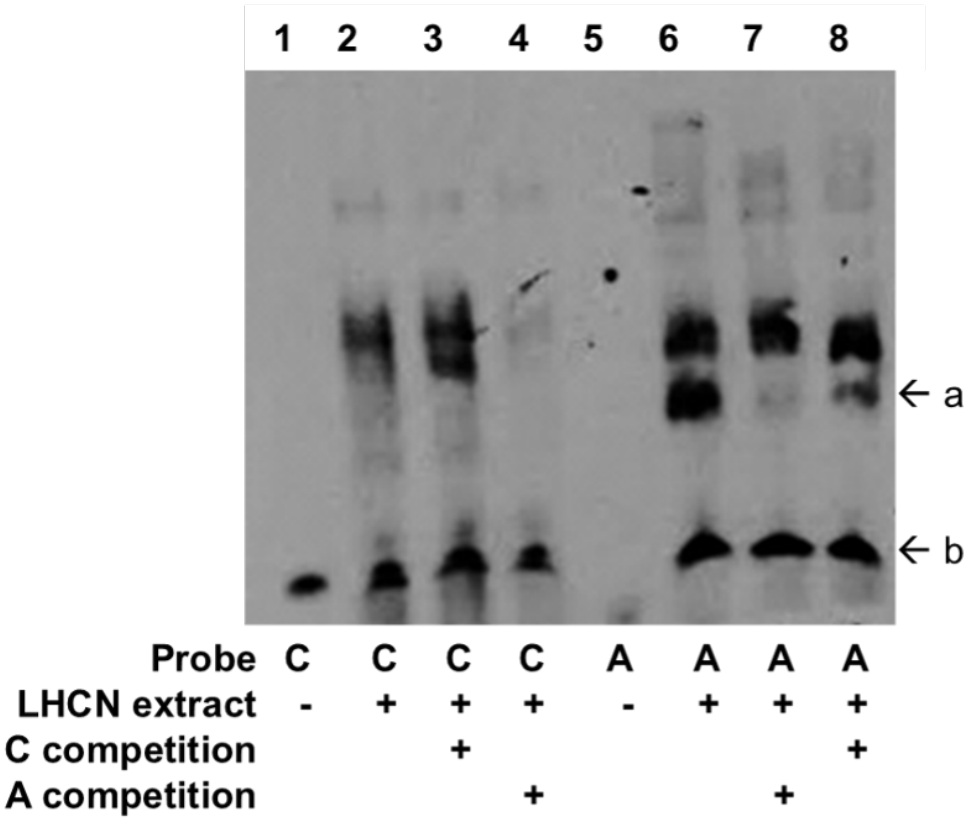


Figure 2.16: rs117643180 exhibits allelic differences in transcription factor binding.
An EMSA using 6µg per lane of nuclear extract from undifferentiated LHCN-M2 cells shows protein-DNA interactions for probes centred around each both alleles of rs117643180. The probe containing rs117643180-A shows allele-specific protein binding (Arrow A, lane 6), relative to the probe containing rs117643180-C (lane 2). A 25-fold excess unlabelled probe containing the A allele competed away A-specific bands more effectively (lane 7) than 25-fold excess unlabelled probe containing the C allele (lane 8). Arrow B shows a biotinylated free probe (200 fmol per lane).

2.4.6 Integration of additional phenotypic layers to identify additional loci for post-challenge insulin resistance

Previous research has shown the utility of integrating information on association with related traits when prioritising genetic loci associated with insulin resistance.[257] Here, I employed this approach to ISI and IFC with the aim to identify biologically relevant loci, with a post-challenge specific insulin resistance association pattern, which do not yet reach genome-wide significance. This approach allows us to identify additional loci of interest without sufficient sample size to detect at a stringent genome-wide significance threshold. To do this, variants that were suggestively associated with IFC and/or ISI were prioritised ($P < 5x10^{-5}$). These were further filtered these variants based on their association with 2 h plasma glucose levels[87] ($P < 5x10^{-4}$), to prioritise those variants with additional evidence of being associated with post-challenge insulin resistance state. I additionally employed a second strategy where I filtered the variants suggestively associated with ISI and/or IFC based on their lack ($P < 0.05$) of association with fasting insulin, to remove those likely associated with liver insulin resistance in the fasting state. These variants were further filtered based on evidence of being eQTLs in skeletal muscle ($P < 5x10^{-4}$).[155] This allowed us to prioritise those variants with evidence of regulating expression in skeletal muscle, a tissue of high biological relevance to postprandial insulin action. Loci which met both prioritisation criteria were considered to have a post-challenge insulin resistance specific association signature, with further evidence of regulatory effects in skeletal muscle, the key tissue in post-prandial glucose uptake.

A total of 10 loci where prioritised using this method. Two loci showed a post-challenge specific association signature, being suggestively associated with IFC/ISI, 2 h glucose and showing no evidence of being associated with fasting insulin in studies of European ancestry – *ERAP2, SLC2A4* (**Table 2.12 and 2.13**). This strategy prioritised 6 additional loci that were below genome-wide significance for ISI and IFC ($5x10^{-5} < P > 5x10^{-8}$; **Table 2.12 and 2.13**). A total of 8 loci where prioritised for IFC, 5 additional ($5x10^{-5} < P > 5x10^{-8}$), and 6 loci for ISI, 3 additional ($5x10^{-5} < P > 5x10^{-8}$).
[300].

Table 2.12 - Integration of additional phenotypic layers identifies additional post-challenge insulin resistance loci of interest

Integrated GWAS Annotation				Identified in Genetic Discovery?		Locus Annotation					
Lead Trait	**(1) 2hr Glucose**	**(2) Not Fasting Insulin + eQTL**	**Integrated traits**	**IFC ($P < 5x10^{-8}$)**	**ISI ($P<5x10^{-8}$)**	**locus**	**rsid**	**MarkerName**	**EA**	**OA**	**EAF IFC**
IFC	N	Y	Not Fasting Insulin + eQTL	N	N	FOXD2	rs116025863	chr1_47930641_C_T	T	C	0.9436
ISI	N	Y	Not Fasting Insulin + eQTL	N	N	LRP5	rs314779	chr11_68097826_G_T	T	G	0.7247
ISI	N	Y	Not Fasting Insulin + eQTL	N	N	CCND2	rs76895963	chr12_4384844_G_T	T	G	0.9769
ISI/IFC	Y	N	2hr Glucose	Y	N	C2CD4A	rs7167878	chr15_62396189_A_C	A	C	0.4606
ISI/IFC	Y	Y	2hr Glucose, Not Fasting Insulin + eQTL	Y	Y	SLC2A4	rs117643180	chr17_7185779_A_C	A	C	0.0231
ISI/IFC	Y	Y	Not Fasting Insulin + eQTL	N	N	SHMT1	rs7212808	chr17_18096846_A_G	A	G	0.4147
IFC	Y	N	2hr Glucose	N	Y	IRS1	rs2972144	chr2_227101411_A_G	A	G	0.3667
ISI	Y	N	2hr Glucose	N	N	PDGFC	rs10009166	chr4_157619733_A_C	A	C	0.322
IFC	Y	N	2hr Glucose, Not Fasting Insulin + eQTL	N	N	ERAP2	rs1216570	chr5_96261717_C_T	T	C	0.5156
IFC	Y	N	2hr Glucose	Y	N	PPP1R3B	rs7012814	chr8_9173358_A_G	A	G	0.4796

Table 2.13 – Annotation of loci identified in the integrated GWAS approach with their association with other traits.

Locus Annotation		IFC adjBMI EUR				ISI adjBMI EUR				2hr Glucose adjBMI EUR (Chen et al, 2021)[87]				Fasting Insulin adjBMI EUR (Chen et al, 2021) [87]				Skeletal muscle eQTL annotation	
locus	**rsid**	**Beta**	**SE**	**P-value**	**N**	**Beta**	**SE**	**P-value**	**N**	**Beta**	**SE**	**P-value**	**N**	**Beta**	**SE**	**P-value**	**N**	**GTEX (P < 5x10^{-4})[154]**	**FUSION (P<0.05)[2]58**
FOXD2	rs116025863	-0.067	0.014	1.74E-06	50671	0.0329	0.0141	0.01941	51187	-0.0251	0.015	0.09426	45269	0.0113	0.0094	0.2274	124952	-	CMPK
LRP5	rs314779	-0.0084	0.0073	0.2506	50671	0.0334	0.0073	5.41E-06	51171	-0.0128	0.0077	0.09645	45249.8	-0.0068	0.0048	0.1551	125378	C11orf24	-
CCND2	rs76895963	0.1027	0.0243	2.32E-05	49703	-0.1192	0.0243	9.08E-07	50231	0.0914	0.0357	0.01046	32043.9	0.0152	0.0253	0.5486	66305.8	CCND2	-
C2CD4A	rs7167878	-0.0426	0.0064	2.76E-11	53287	-0.0078	0.0064	0.2248	53657	0.0324	0.0058	2.32E-08	63395.9	0.0018	0.0039	0.6416	151013	-	-
SLC2A4	rs117643180	0.1428	0.0221	1.11E-10	50671	-0.1416	0.0221	1.58E-10	51187	0.1893	0.0253	7.31E-14	38302	0.018	0.0147	0.2193	112805	SLC2A4, DGL4	SLC2A4, YBX2,KCTD11
SHMT1	rs7212808	-0.0273	0.0066	3.54E-05	50671	0.033	0.0066	6.43E-07	51187	-0.0215	0.007	0.00213	45268.8	-0.0085	0.0044	0.05205	125397	SHMT1,LGL1,ALKBH5,USP32P2	ALKBH5, TOP3A, SHMT1,LGL1, MIEF2,SMCR8
IRS1	rs2972144	-0.0337	0.0066	3.30E-07	52458	0.0644	0.0067	1.10E-21	50511	-0.0312	0.0059	1.24E-07	63391	-0.0514	0.004	2.28E-38	150996	-	-
PDGFC	rs10009166	-0.0059	0.0069	0.3929	50671	0.3225	0.0069	9.47E-07	51187	-0.0251	0.0072	0.00049	45269	-0.0224	0.0045	5.64E-07	125397	-	
ERAP2	rs1216570	-0.0287	0.0065	1.01E-05	50671	0.024	0.0065	0.0002362	51187	-0.029	0.0069	2.64E-05	45268.8	-0.002	0.0044	0.6476	125397	ERAP2,LNPEP	ERAP2,ERAP1, LNPEP
PPP1R3B	rs7012814	0.0481	0.0065	1.35E-13	52458	3.00E-04	0.0064	0.9628	52828	0.0354	0.0058	1.04E-09	63395.8	-0.045	0.004	8.34E-30	148942	-	-

The *ERAP2* locus (rs1216570) was associated with higher IFC and 2 h glucose[87], and lower ISI, suggestive of post-challenge insulin resistance (**Figure 2.17a-b**. rs1216570 is an eQTL associated with higher *ERAP2* expression in skeletal muscle (GTEX v8 EUR: NES = 1.01, SE = 0.029, P = 4.15×10^{-138}). This eQTL in both muscle and adipose tissues, as well as 2hr glucose association has also been identified in other populations including in a cohort of African American ancestry[268], as well as a cohort from Sweden.[269] A muscle *ERAP2* eQTL at this locus has additionally identified in the Pima Indian population.[270] However, there is limited evidence of a shared genetic signal for this eQTL signal (GTEX v8 EUR) and IFC (posterior probability of colocalisation = 0.32).

The lead variant at this locus, rs1216570, is additionally an eQTL in skeletal muscle for *LNPEP (*GTEx v8 EUR: NES = -0.13, SE = 0.026, P = 3.92×10^{-7})[154] and there is strong evidence of colocalisation of this eQTL and IFC signal (posterior probability colocalisation = 0.85; **Table 2.14**). *LNPEP* encodes IRAP, an insulin responsive amino peptidase that collocates with GLUT4 on GLUT4 secretory vesicles, and is well established to be involved in GLUT4 trafficking.[271] This demonstrates that this integrated approach identifies loci that are biologically relevant to post-challenge insulin resistance.

SHMT1 (rs7212808) was associated with a favourable glycaemic signature (**Figure 2.17c-d**) and is additionally an eQTL for increased *LLGL1* expression in skeletal muscle (GTEx v8 EUR: NES = 0.15, SE = 0.028, P = 1.00×10^{-7} ; posterior probability of colocalisation IFC and LLGL1 eQTL = 0.91[154]). LLGL1 is a Rab GTPase activating protein that interacts with Rab10, a key player in insulin-stimulated GLUT4 translocation.[272] Although LLGL1 has not previously been implicated in post-challenge insulin response, it is plausible that this may play a role in regulating glucose uptake in the postprandial state.

All phenotypic signatures of identified loci were additionally assessed using formal statistical colocalisation method HyPrColoc[176] (**Figure 2.18**). This provided additional evidence for some of the identified loci, and further identified additional loci of interest below genome-wide significance for post-challenge insulin resistance that are shared with other traits including T2D.

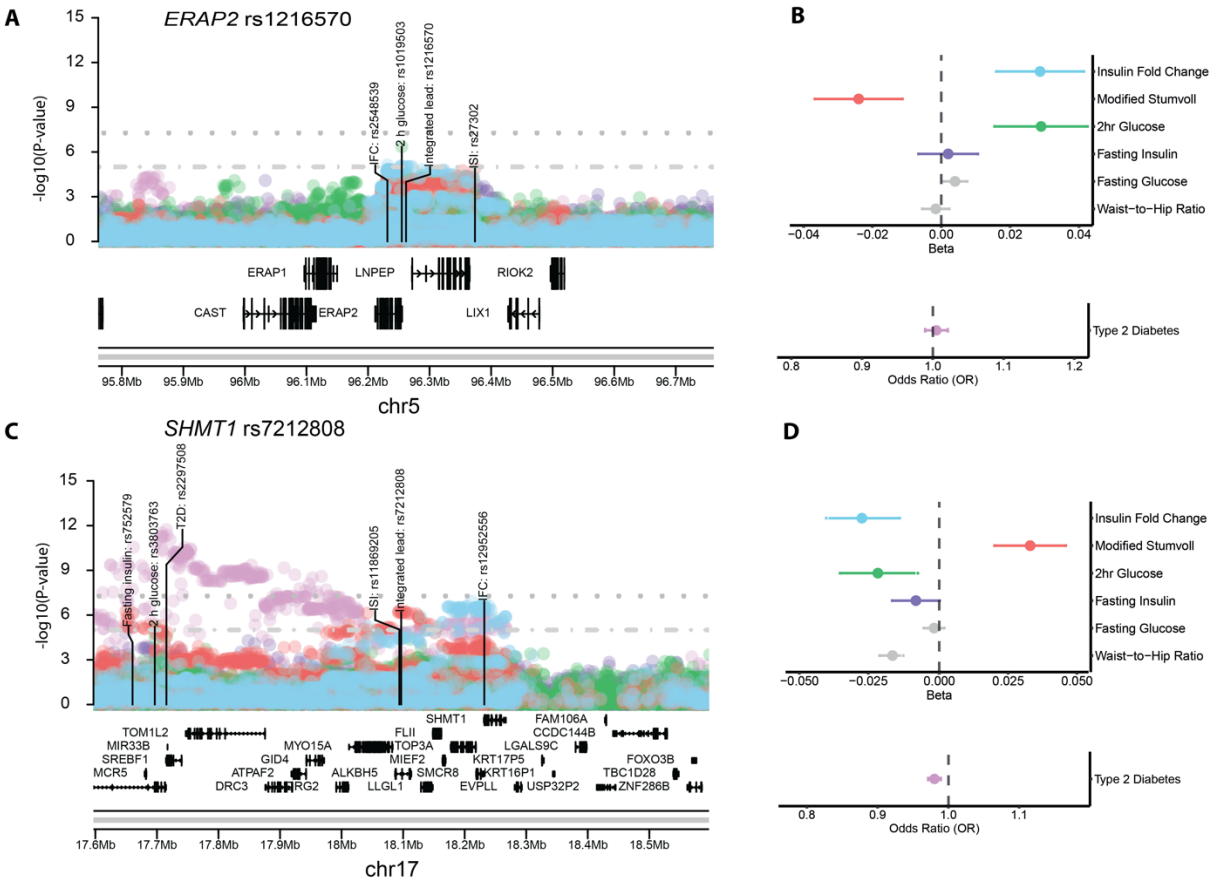


Figure 2.17: Integrated GWAS approach highlights additional loci including those implicating glucose transport

A) *ERAP2* locus: regional association plot for relevant traits. B) *ERAP2* locus: association of rs1216570 with relevant traits. **C)** *SHMT1* locus: regional association plot for relevant traits, with colours of points indicating traits of interest. **D**) *SHMT1* locus: association of rs7212808 with relevant traits. **A and C)** Regional association plot showing ± 500kb around the lead IFC-associated variant, with points indicating the association and location of individual variants. Colours of points indicating traits of interest: insulin fold change (IFC) - blue, Modified Stumvoll (ISI) – red, and 2 h glucose[87] – green, fasting insulin[87] – purple, type 2 diabetes[251] – pink . Y-axis denotes the $-\log_{10}$(p-value) of association. The dashed lines and dotted line indicate genome-wide ($P = 5x10^{-8}$) and suggestive ($P = 1x10^{-5}$) significance threshold, respectively. X-axis denotes the genomic position. All association statistics are from BMI adjusted analyses in studies of European ancestry, except T2D which is unadjusted for BMI**.** The dashed lines indicate suggestive ($P = 1x10^{-5}$) significance threshold and dotted line indicates genome-wide significance ($P = 5x10^{-8}$), respectively. Labels indicate the lead variant for each trait, with location indicating position. Integrated lead is the lead variant defined for this locus using the integrated GWAS approach. **B and D)** Plot shows beta estimate or odds ratio of association for lead variant with insulin fold change (IFC) - blue, Modified Stumvoll (ISI) – red, and 2 h glucose[87] – green, fasting insulin[87] – purple, type 2 diabetes[251] – pink, all included in a) as well as fasting glucose[87] and waist-to-hip ratio[256] plotted in grey. All association statistics are from BMI adjusted analyses, except T2D, in studies of European ancestry. Error bars indicate 95% CI.

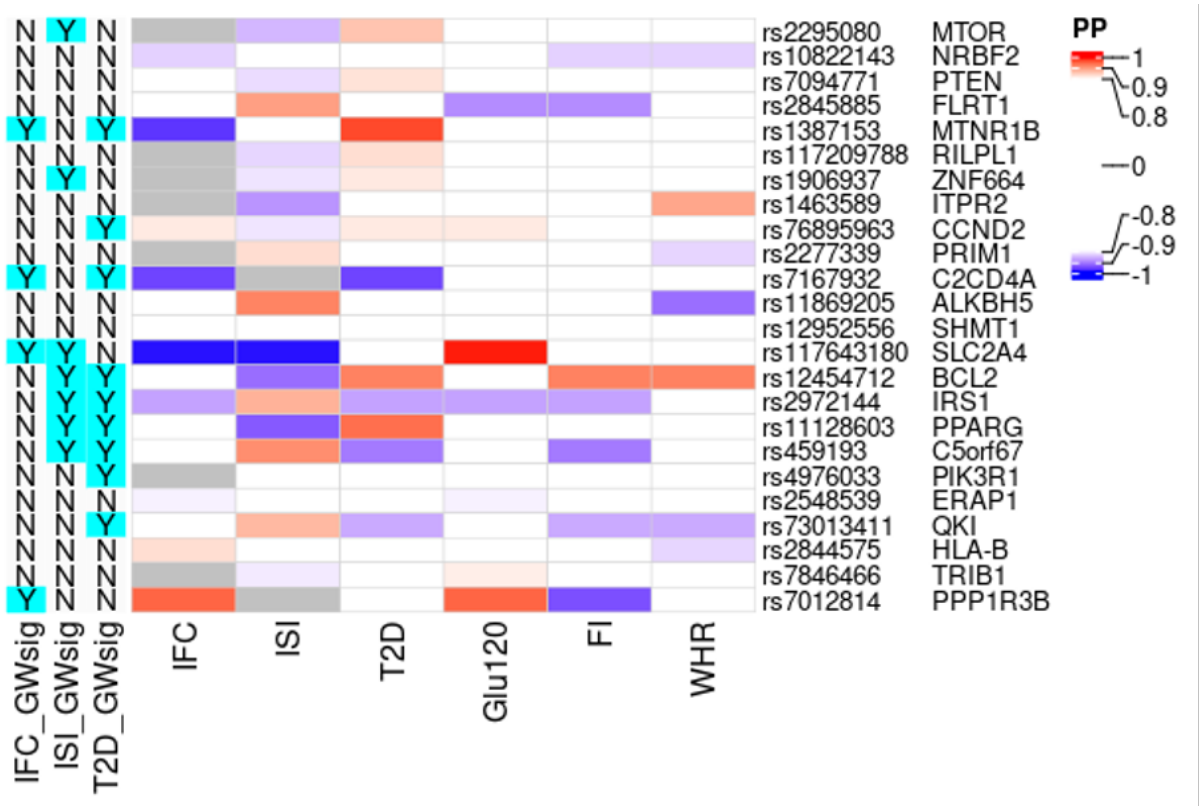


Figure 2.18: Joint colocalisation at loci associated with IFC or ISI with related traits
Fasting insulin (FI), 2hr glucose (Glu120) type 2 diabetes (T2D) and waist-to-hip ratio (WHR). All summary statistics used as input are from European ancestry individuals and adjusted for BMI, except for T2D which is unadjusted for BMI (Methods). GWsig denotes whether the IFC/ISI lead SNP at a locus (indicated by the rsid) meets genome-wide significance for the indicated trait (unadjusted P < $5x10^{-8}$). White denotes posterior probability < 0.8, with no evidence of colocalisation. Red positive beta for the indicated trait, the blue negative beta for the trait. Shade denotes the posterior probability of colocalisation between 0.8 and 1. Grey denotes that the trait was not included in HyPrColoc due to the locus (lead SNP +/- 500kb) not meeting the inclusion threshold of P < **$5x10^{-4}$** for that trait. All effects are aligned to the IFC effect allele.

2.4.7 Tissue of action

To delineate potential tissue of action for post challenge loci I first employed LDSC-SEG to identify if heritability of IFC or ISI is enriched in regions surrounding genes with the highest specific tissue or cell type expression. ISI showed nominal genome-wide enrichment of active chromatin state in adipose, liver, skeletal muscle, and pancreatic tissue, as well as nominal enrichment (unadjusted p value < 0.05) in tissue specific gene expression in liver, muscle tissues and other T2D unspecific tissues. Insulin fold change showed nominal enrichment in active chromatin marks in liver, pancreas and other tissues, as well as nominal enrichment in liver specific gene expression. However, none of these annotations survive multiple testing correction.

I further assessed potential tissue of action at individual associated loci by assessing colocalisation of loci identified to be associated with IFC and ISI in either the genetic discovery or integrated analyses with eQTLs in T2D relevant tissues – adipose, liver, pancreatic islets and skeletal muscle. There was colocalisation (posterior probability of colocalisation > 0.7) at 3 loci associated IFC at *ERAP2*, *SHMT1* and *IRS1* with eQTLs identified in skeletal muscle and subcutaneous and visceral adipose tissues, key tissues in the post-prandial insulin response (**Table 2.14**).

Table 2.14 – eQTL colocalisation at loci identified in the integrated GWAS approach
marginal summary statistics were used for colocalisation for both ISI and IFC as well as eQTLs. Results outline colocalisations with a posterior probability > 0.7 for PP.H4.abf. Posterior probability of colocalisation (PP) was calculated using approximate bayes factor (ABF) analysis from Wallace et al (2013) was assessed under the following hypotheses. All analyses were executed using R package coloc. H4 > 0.7 was used to assess evidence of colocalisation in this analysis. H0: neither trait has a genetic association in the region. H1: only trait 1 has a genetic association in the region. H2: only trait 2 has a genetic association in the region. H3: both traits are associated, but with different causal variants. H4: both traits are associated and share a single causal variant

locus	**trait1**	**trait2**	**ENSG_gene_eQTL**	eQTL **genename**	**eQTL_tissue**	**nsnps**	**PP.H0.abf**	**PP.H1.abf**	**PP.H2.abf**	**PP.H3.abf**	**PP.H4.abf**
CCND2	ISI	eQTL	ENSG00000118971.7	CCND2	Muscle_Skeletal	1117	2.027E-04	5.420E-04	4.420E-04	0.000	0.999
SHMT1	ISI	eQTL	ENSG00000091542.8	ALKBH5	Muscle_Skeletal	779	1.880E-42	1.087E-03	3.870E-41	0.021	0.978
SHMT1	ISI	eQTL	ENSG00000091542.8	ALKBH5	Adipose_Subcutaneous	779	2.340E-27	1.132E-03	4.820E-26	0.022	0.977
SHMT1	ISI	eQTL	ENSG00000091542.8	ALKBH5	Adipose_Visceral_Omentum	779	1.250E-11	1.529E-03	2.570E-10	0.030	0.968
IRS1	IFC	eQTL	ENSG00000169047.5	IRS1	Adipose_Subcutaneous	1237	7.240E-16	2.267E-03	1.080E-14	0.033	0.965
SHMT1	ISI	eQTL	ENSG00000131899.10	LLGL1	Liver	779	1.830E-05	1.831E-03	3.760E-04	0.037	0.961
IRS1	IFC	eQTL	ENSG00000261379.1	AC010735.1	Adipose_Subcutaneous	1237	2.630E-06	3.094E-03	3.930E-05	0.045	0.952
IRS1	IFC	eQTL	ENSG00000272622.1	AC010735.2	Adipose_Subcutaneous	1237	3.910E-08	3.317E-03	5.840E-07	0.049	0.948
SHMT1	ISI	eQTL	ENSG00000091542.8	ALKBH5	Pancreas	779	4.470E-04	2.820E-03	9.210E-03	0.057	0.930
CCND2	ISI	eQTL	ENSG00000118971.7	CCND2	Adipose_Subcutaneous	1117	2.267E-02	4.891E-03	4.940E-02	0.010	0.913
MTOR	ISI	eQTL	ENSG00000271895.2	AL109811.3	Adipose_Subcutaneous	1054	2.880E-05	1.560E-04	1.421E-02	0.076	0.910
SHMT1	IFC	eQTL	ENSG00000176974.19	SHMT1	Pancreas	683	4.480E-13	7.790E-04	5.310E-11	0.091	0.908
SHMT1	IFC	eQTL	ENSG00000131899.10	LLGL1	Muscle_Skeletal	683	1.770E-07	7.850E-04	2.100E-05	0.092	0.907
SHMT1	ISI	eQTL	ENSG00000240445.3	FOXO3B	Adipose_Subcutaneous	779	6.970E-04	3.661E-03	1.430E-02	0.074	0.907
IRS1	IFC	eQTL	ENSG00000272622.1	AC010735.2	Adipose_Visceral_Omentum	1237	2.296E-04	6.449E-03	3.429E-03	0.095	0.894

MTOR	ISI	eQTL	ENSG00000198793.12	MTOR	Muscle_Skeletal	1054	3.540E-06	2.320E-04	1.750E-03	0.113	0.885
MTOR	ISI	eQTL	ENSG00000171824.13	EXOSC10	Adipose_Visceral_Omentum	1054	3.570E-05	2.030E-04	1.760E-02	0.099	0.883
SHMT1	ISI	eQTL	ENSG00000131899.10	LLGL1	Muscle_Skeletal	779	2.250E-06	6.305E-03	4.640E-05	0.129	0.865
ERAP2	IFC	eQTL	ENSG00000113441.15	LNPEP	Pancreas	1625	5.350E-07	2.764E-02	2.300E-06	0.118	0.855
ERAP2	IFC	eQTL	ENSG00000113441.15	LNPEP	Muscle_Skeletal	1625	3.410E-05	2.829E-02	1.460E-04	0.121	0.851
SHMT1	ISI	eQTL	ENSG00000108448.20	TRIM16L	Adipose_Visceral_Omentum	779	3.080E-03	3.905E-03	6.328E-02	0.079	0.850
ERAP2	IFC	eQTL	ENSG00000247121.6	AC009126.1	Muscle_Skeletal	1625	7.140E-05	2.955E-02	3.070E-04	0.126	0.844
SHMT1	ISI	eQTL	ENSG00000176994.10	SMCR8	Liver	779	2.685E-03	6.626E-03	5.524E-02	0.136	0.800
MTOR	ISI	eQTL	ENSG00000226849.1	AL109811.1	Adipose_Subcutaneous	1054	5.060E-14	4.490E-04	2.490E-11	0.220	0.779
SHMT1	ISI	eQTL	ENSG00000279428.1	AC087164.2	Liver	779	3.330E-14	1.169E-02	6.850E-13	0.240	0.749

2.5 Discussion

The impairment of insulin's ability to control glucose metabolism, or insulin resistance, is a major contributor to a range of metabolic and endocrine disorders including T2D. Previous studies of the genetic determinants of insulin resistance have predominantly focused on the fasting state, where increased glucose and insulin more likely reflect liver insulin resistance than muscle or fat tissue insulin resistance. Here, I have attempted to use measures, IFC and ISI, that reflect, at least in part, insulin's important postprandial action on glucose uptake into skeletal muscle and adipose tissue. Using this approach, 9 novel loci associated with post-challenge insulin resistance in genetic studies of European ancestry were identified, including an association at *SLC2A4*, which encodes the key glucose transporter implicated in insulin-stimulated glucose disposal. *BICC1* was additionally identified to be associated with IFC in meta-analysis of cohorts of Hispanic American and East Asian ancestries.

The glucose transporter GLUT4 (encoded by *SLC2A4*) is highly enriched in muscle and fat but not present in liver. In response to insulin (and exercise in skeletal muscle), GLUT4 rapidly translocates from an intracellular pool to the cell surface to facilitate glucose uptake.[273] Genetic ablation of GLUT4 in mice results in severe postprandial insulin resistance, hyperinsulinemia, and hyperglycaemia.[274] Incorporation of measures of post challenge insulin has demonstrated that genetic variation which directly influences *SLC2A4* expression does indeed have a significant impact on the effectiveness of human postprandial insulin action.

Further evidence of loci directly implicating the biology of GLUT4 translocation in post-challenge insulin resistance comes from our findings at the *ERAP2* locus (rs1216570), suggestively associated ($P = 1.01 \times 10^{-5}$) with increased IFC and reduced ISI ($P = 2.4 \times 10^{-4}$), and suggestively associated with 2 h glucose ($P = 2.64 \times 10^{-5}$) but not fasting glycaemic measures. There is evidence of a shared genetic signal for IFC and eQTL in skeletal muscle for *LNPEP* which encodes IRAP, the insulin responsive amino peptidase. IRAP is found on GLUT4 secretory vesicles and binds TBC1D4, to prevent translocation of GLUT4 to the plasma membrane in the fasting state (low insulin).[201] IRAP is therefore a negative regulator of plasma membrane GLUT4[271]. The lead variant at this locus is further in perfect LD ($R^2=1$) in Europeans with a previously reported missense variant significantly associated with 2hr glucose (rs2549782)[275]. However, colocalisation of our lead variant rs1216570 with an *LNPEP* skeletal muscle

eQTL, no evidence of an impact of *ERAP2* knockdown on GLUT4 trafficking *in vitro*, and *in silico* algorithms including SIFT[276], PolyPhen2[277] and CADD[278] predicting the missense variant to be benign and to be unlikely to have an impact on the function of the encoding protein, suggest that *LNPEP* rather than *ERAP2* may be the causal gene at this locus.

We also identified loci near genes directly implicated in insulin signalling including *IRS1* (reviewed in[279]) and others near genes known to be involved more broadly in insulin sensitivity including *PPARG*[280], *PPP1R3B*[281,282], and *MTOR*[283]. The observation that 10 of the 11 loci identified in Europeans are additionally colocalised with T2D provides evidence of mechanisms behind these disease associations and importantly highlights the utility of studying the role of increased genetic risk of insulin resistance, and not only beta-cell dysfunction, in T2D pathogenesis.

Notably, loci near genes such as *MTNR1B* and *C2CD4A* that have been implicated in T2D through an impact on pancreatic beta-cell function were identified.[196,284] As we used indirect measures of post-challenge insulin resistance that include measurements of circulating insulin, it is not surprising that some genetic variants regulating insulin secretion are also captured by this approach. Interestingly the literature around *MTNR1B* and its role in T2D pathogenesis is unclear with conflicting results reported in the literatue.[285]. While this does not discount a role of MTNR1B effecting insulin secretion in beta cells at this locus, future studies should also consider the role of alternative candidate genes at this locus in other arms of T2D pathogenesis and whether other candidate causal genes could play additional roles in T2D risk at this locus, in the case, post-challenge insulin resistance.

Assessment of the relationship of ISI and IFC both genetically and observationally with other traits allowed assessment of the specificity of these traits to post-prandial insulin resistance. This demonstrated that ISI and IFC are both able to capture post-challenge insulin resistance with these being highlight correlated with post-challenge measures of glucose and insulin. However, the additional correlation of ISI with fasting insulin demonstrates, as expected that this is a measure of whole-body insulin sensitivity. On the other hand, IFC was not correlated with fasting insulin, suggesting that this is more specific to the postprandial state and therefore perhaps a better proxy when trying to capture post-challenge specific impacts on insulin sensitivity.

Strengths and limitations

This study represents the largest and most comprehensive assessment of the genetic architecture of post-challenge insulin resistance to date and includes systematic characterisation of post-challenge insulin resistance traits, as well as systematic annotation of identified genetic loci. The sample size for ISI is approximately 3.5x larger than the previous effort with an additional 6 loci identified over and replications of 2 loci identified previously. To our knowledge the genetic architecture of IFC has not been assessed previously, and comparison of IFC with other traits demonstrate that this is more specific to post-prandial insulin sensitivity than other OGTT traits studied previously thus providing novel aetiological insight.

There are several limitations that need to be considered in this study. First, ISI and IFC do not represent the gold standard measures of postprandial insulin resistance. Whilst, as discussed above, these measures are not available at sufficient scale for population genetic studies, it is likely that some non-specific effects, for example on insulin secretion rather than insulin sensitivity are captured using these traits. Tissue-specific expression annotation using eQTLs aimed to account for this, but it is possible there are still some false positive findings in this work.

Like other GWAS at present, our sample size is dominated by cohorts of European ancestry (97%), and this analysis was underpowered to detect ancestry-specific effects in Hispanic American and East Asian populations. Though to our knowledge this meta-analysis includes almost all studies that have OGTT insulin/glucose measures and genotyping data, future discovery efforts should aim to increase the diversity of such studies. In light of this, no replication analyses in an independent study were performed as part of this genetic discovery. Increased sample size of future efforts including additional studies will be needed to ensure the stability and replicability of the associations identified.

Though candidate causal genes and variants have been suggested at some loci, like other GWAS studies, we cannot be certain of the causality of these without *in vitro* functional follow-up to assess the regulatory architecture at each locus and how this is affected by candidate causal variants. Future work examining the regulatory architecture of skeletal muscle and adipose tissue in the context of post-challenge

insulin resistance will hopefully help to address this, along with focused *in vitro* characterisation of specific loci.

Finally, potential sex specific genetic effects were not determined in this study and sex stratified GWAS analyses in the future would be needed to determine whether there are sex specific role of genetic risk to post-challenge insulin resistance as has been done for other glycaemic traits.[214]

Conclusion

To conclude, 10 novel genome-wide loci not previously implicated in post-challenge insulin resistance were identified across analyses including individuals European and Non-European ancestry. This work shows that 8 of these share their genetic architecture with T2D and hence identify the possible mechanism of action that leads to T2D pathogenesis for these loci. The *BICC1* locus was specific and only identified in our meta-analysis of cohorts of Hispanic American and East Asian ancestries, despite the much smaller sample size.

I further demonstrate the utility of an integrative approach to prioritisation of loci below genome-wide significance that have biological relevance to traits of interest, with the identification of additional loci that are implicated in post-challenge insulin resistance. For the first time, an association of a common intronic variant in *SLC2A4* which encodes the insulin-regulated glucose transporter GLUT4, a critical player in postprandial glucose disposal, with insulin resistance is reported. We demonstrate *in silico* and *in vitro* that the lead variant at this locus (rs117643180) affects *SLC2A4* expression in skeletal muscle, suggesting potential mechanisms behind the post-glucose challenge-specific signature associated with this locus.

This provides novel insights into the control of post-challenge glucose disposal in humans and highlights several candidate loci and genes for future investigation.

Chapter 3: Investigating candidate regulators of insulin-stimulated glucose transport and GLUT4 trafficking

Chapter 3 Overview

The workhorse of insulin-stimulated glucose uptake in the postprandial state is the insulin-responsive glucose transporter GLUT4, encoded by *SLC2A4*. Predominantly expressed in muscle and adipose tissue, GLUT4 is responsible for the majority of glucose uptake in these insulin-responsive tissues and therefore thought to be a key player in post-prandial insulin resistance and type 2 diabetes.

Here, I aimed to identify novel regulators of GLUT4 and therefore insulin stimulated glucose uptake using a mid-to-high throughput assay in 3T3-L1 adipocytes. I identified genes located at loci associated with post-challenge insulin resistance which were expressed in muscle and/or adipose tissue in humans. The 36 prioritised genes were assessed for their role in GLUT4 trafficking through their systematic knockdown using short interfering RNAs (siRNA) pools in 3T3-L1 adipocytes to assess the impact of depletion of these genes on the trafficking of GLUT4 to the cell surface upon insulin stimulation or overall expression.

siRNA knockdown of 14 of the 36 genes resulted in significant changes in GLUT4 trafficking and/or regulation in response to insulin stimulation compared to controls. 9 of these genes have not previously been reported to be involved in GLUT4 trafficking, and therefore are novel candidates involved in this pathway that warrant further mechanistic investigation.

Publications relating to this chapter

Williamson A, [100 authors], Mohlke K.L., Wheeler E, O'Rahilly S, Fazakerley D.J., Langenberg C. Genome-wide association study of postprandial glucose metabolism and functional characterisation identifies candidate genes for insulin-stimulated glucose uptake. **Nature Genetics** (2023, accepted).

Collaboration and contributions

This project was done in collaboration with Drs Daniel Fazakerley and Dougall Norris. The paragraph below outlines relative contributions.

I designed the gene selection strategy and selected the genes using this as input for the experimental work presented in this chapter. I ran all the GLUT4 screen experimental work alongside Dr Norris, sharing culturing, siRNA knockdown and assay steps during data collection. Due to the scale of the screen this is not feasible to carry out as just one person. I carried out all data analysis of the results for the GLUT4 screen presented in this thesis as well as the interpretation of the results. Dr Fazakerley conducted the confirmatory radiolabelled glucose transport experiments. I conducted the qPCR experiments to confirm level of gene knockdown and the literature reviews to interpret the possible roles of the genes identified.

How this work fits in with other chapters in this thesis:

Chapter 2 summarised the results of a genetic discovery of post-challenge insulin resistance (N > 55,000) through two OGTT-derived traits that approximate, at least in part, postprandial insulin resistance – insulin fold change during an OGTT, and Modified Stumvoll ISI. Here I identified 11 genome-wide significant loci ($P < 5x10^{-8}$) in the meta-analysis of cohorts of European ancestry, and 1 genome-wide significant locus in a meta-analysis of 3 cohorts of Hispanic American and East Asian ancestry across both traits.

I additionally employed an integrative GWAS approach, where I further prioritised, loci associated with post-challenge insulin resistance at a more permissive threshold ($P < 5x10^{-5}$), which were:

1) at least suggestively associated with 2 h glucose[87] ($P < 5x10^{-4}$), and/or,
2) not associated with fasting insulin[87] ($P > 0.05$) and an eQTL in skeletal muscle[154].

The 5 loci identified in the genome-wide discovery stage for insulin fold change, as well as loci prioritised for this trait in the integrative GWAS approach were used as input to the gene selection outlined in this chapter.

3.1 Aims and Objectives

The genetic analyses presented in Chapter 2 identified novel loci associated with insulin resistance post-glucose challenge through their association with insulin Fold Change (IFC) during an oral glucose tolerance test (OGTT).

The project presented in this chapter aimed to follow up genes located at these genetic loci to investigate whether these genes play a role in insulin-stimulated glucose uptake, and therefore GLUT4 trafficking.

The following objectives were addressed:

1. Identify and refine a list of genes at genetic loci associated with post-challenge insulin resistance for further follow up.
2. Assess the role of these genes in GLUT4 trafficking and/or regulation in 3T3-L1 adipocytes.
3. Use the literature to hypothesise the potential mechanisms by which these genes are implicated in insulin-stimulated glucose transport.

3.2 Introduction

The disposal of glucose in the postprandial state is a critical process involving glucose uptake primarily into muscle and adipose tissues[27,286]. Defects in this process are established to be one of the initiating factors on the path towards T2D[7,47,287]. The workhorse of this process is the insulin-dependent glucose transporter Glucose transporter type 4 (GLUT4), encoded by *SLC2A4*, which is predominantly expressed in muscle and adipose tissues[236,273]. Unlike other tissues which are involved in glucose disposal such as the liver where glucose transport is constitutively active and does not rely on upregulation, in muscle and adipose tissue upregulation of plasma membrane GLUT4 by insulin stimulation, and exercise in the case of muscle, is critical for the uptake of glucose into these tissues. Approximately 50-80% of insulin stimulated glucose uptake in healthy lean individuals without diabetes occurs in skeletal muscle, with a further ~10-25% in white adipose tissue under hyperinsulinemic-euglycaemic clamp conditions[271,288–296].

Whilst there is a relatively lower proportion of glucose uptake in adipose tissue, this is a key metabolic organ, serving as a lipid store, an indicator of energy status via the secretion of adipokines, and regulation of glucose homeostasis.[297] These 3 key roles, and the fact that adipokines allow the regulation of other organs including the brain, liver and muscle mean adipose is considered a rheostatic organ, and therefore despite being responsible for a small fraction of glucose disposal, it plays an essential role in maintenance of whole-body glucose homeostasis.[297]

Whilst the key players of the GLUT4 trafficking pathway have been identified[298], there are still questions about how this process is regulated and how GLUT4 transport to the cell surface is coordinated[298]. Identification of additional players in this pathway may identify novel therapeutic targets for T2D. Further this pathway is thought to be equivalent in muscle and adipose tissue.[273] Therefore, despite being responsible for less glucose uptake, studying this process in adipocytes still allows insights to be gained across GLUT4 expressing tissues.

GLUT4, insulin resistance, and type 2 diabetes

The link between dysregulation of GLUT4 and insulin resistance has been demonstrated in mouse models where GLUT4 has been ablated either in the whole

organism or in specific tissues. Whole body knockout mice are surprisingly not diabetic but do display postprandial hyperinsulinemia.[299,300] This suggests that complete absence of GLUT4 could result in upregulation of compensatory mechanisms to maintain glycaemia. These mice did however have several abnormalities including reduction in adipose tissue mass, reduced longevity and growth retardation.[299,300] In heterozygous null mice, there is post-prandial insulin resistance and reduced glucose uptake in muscle and adipose tissues, as expected with a disruption of GLUT4.[274,301]

In rodent models, GLUT4 has been both specifically depleted and upregulated in muscle and adipose tissue.[274,301–303] These models show the complex interplay of muscle, adipose, and liver in maintaining whole body insulin sensitivity. In muscle-specific GLUT4 deficient mice, as well as loss of insulin responsiveness in muscle, these mice have decreased insulin responsiveness in adipose and liver.[302] A similar phenotypic effect is seen in mice with GLUT4 deficiency specifically in adipose tissue, despite this being responsible for a much smaller percentage of insulin stimulated glucose uptake.[304,305] Additional experiments shows that overexpression of GLUT4 in adipose rescues the insulin resistance phenotype in muscle-specific GLUT4 knockout animals.[303]

Unlike in rodent studies, the effect on expression of GLUT4 in humans with diabetes is debated. It has been demonstrated that protein levels of GLUT4 are reduced in skeletal muscle in individuals with T2D, as well as of several proteins involved in the trafficking of GLUT4[297], however this is not supported by other studies which reported normal GLUT4 expression in individuals with T2D.[306,307] Some studies have also identified changes in the regulation of trafficking and translocation of GLUT4 in individuals with diabetes, rather than effects on overall levels.[306,307]

In humans, no pathogenic coding mutations in GLUT4 have been reported to our knowledge. However GWAS reported previously a intronic variant associated with 2 h glucose[87], which was identified through work in this thesis (Chapter 2) to additionally be associated with insulin resistance post-glucose challenge and affect the expression levels of GLUT4 in skeletal muscle. However as discussed in previous chapters, other rare coding mutations in genes upstream of GLUT4 in the insulin-stimulated glucose uptake pathway have been identified including *AKT2* and *TBC1D4.*[106,108,111,113,236] Truncating mutations in *TBC1D4* have been identified in a patient with severe insulin

resistance as well as in the isolated Greenlandic Inuit population (*TBC1D4* p.Arg684*Ter*) – that is found in ~17% of this population.[109] It was identified that homozygous carriers of this variant had elevated circulating 2 h glucose and insulin levels but decreased fasting glucose levels. The same trend of phenotype was also observed in heterozygous carriers. This mutation is associated with reduced GLUT4 expression in skeletal muscle of carriers as well as reduction in expression of the muscle-specific long isoform *TBC1D4* both at the mRNA and protein levels.[236] The variant *TBC1D4* p.Arg684*Ter* was later additionally identified in Canadian and Alaskan Inuit populations but it is extremely rare in other outbred populations globally.[303] In the Canadian and Alaskan Inuit populations it was identified that homozygous carriers also had elevated 2 h glucose and insulin concentrations and reduced peripheral insulin sensitivity, but no discernible effect was identified for fasting glucose or insulin levels.[303] Due to the post-prandial specific effect of this variant, this means that there is an underdiagnosis of T2D in these carriers, with defects in glucose metabolism not apparent in fasting tests regularly done in the clinic.[303] Overall, such examples of mutations in humans of genes upstream of GLUT4 show that GLUT4 dysregulation can result in insulin resistance and T2D.

Insulin stimulated glucose uptake pathway

The PI3K/AKT pathway downstream of insulin signalling is responsible for the activation of GLUT4 trafficking in both muscle and adipose tissue[273,308,309]. As discussed in Chapter 1, AKT acts as a "hub" to the downstream actions of insulin signalling in insulin responsive tissues. Concerning GLUT4 trafficking, AKT mediates phosphorylation of TBC1D4 (also known as AS160) which associates with cargo of the GLUT4 secretory vesicles (GSVs), including insulin responsive amino peptidase (IRAP) which is encoded by *LNPEP*[272,310,311] (**Figure 3.1**). Inactivation of TBC1D4 further allows the maintenance of Rab10 activation which is essential for the promotion of trafficking and docking of GSVs to the cell surface.[272,312,313] Glucose uptake via GLUT4 in muscle tissue is activated by both exercise and insulin stimulation.[314] In exercise, muscle contraction activates GLUT4 translocation by a distinct mechanism.[314] Here, AMPK is activated which is able to phosphorylate TBC1D4.[314] This phosphorylation promotes GLUT4 translocation to the plasma membrane.[272,310,311]

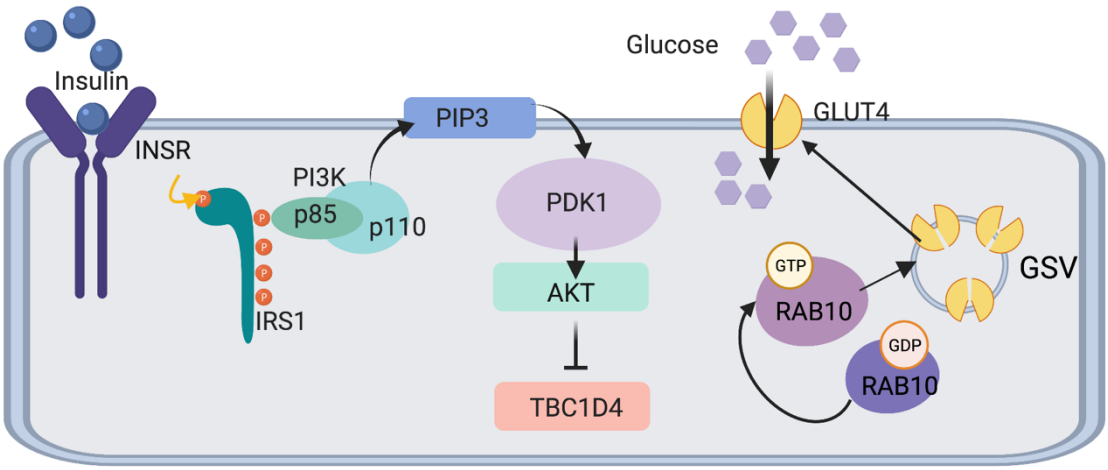


Figure 3.1: Schematic overview of the insulin-stimulated glucose uptake pathway in muscle and adipose tissue.
INSR – insulin receptor, IRS1 – insulin receptor substrate 1, Orange circles with P – phosphorylated, PI3K - Phosphoinositide 3-kinase, made up of P85 – the regulatory subunit and P110 – the catalytic subunit[315], AKT - AKT serine/threonine kinase 1/2 , TBC1D4 - TBC1 domain family member 4 also known as AKT substrate of 160 kDa (AS160), RAB10 Ras-Related GTP-Binding Protein 10, GTP – guanosine triphosphate, GDP - guanosine diphosphate, GLUT4 – glucose transporter type 4, GSV – GLUT4 secretory vesicle. Blue circles represent insulin, purple hexagons represent glucose. Figure generated using Biorender.com

In the absence of insulin/exercise stimulation the majority of GLUT4 (~80%) is sequestered intracellularly in these GSVs, and in the trans-Golgi network.[316] GSVs are only transported to the cell surface in the presence of insulin stimulation, this is the primary way that insulin promotes the amount of GLUT4 at the cell surface (**Figure 3.1 and 3.2**). These GSVs contain GLUT4 as well as additional cargo such as IRAP and Rab10.[317–319]

The remaining intracellular GLUT4 (~20%) is trafficked to the plasma membrane through endosomes and their constant dynamic recycling between intracellular compartments and the plasma membrane (**Figure 3.2**).[317,320] In adipocytes and muscle cells the major pathway in which this occurs is through clathrin-mediated endocytosis. Once internalised GLUT4 is sorted into endosomes where it can be quickly returned to the plasma membrane via exocytosis, or recycled and sorted into GSVs where it will be rapidly trafficked to the cell surface upon insulin stimulation.[317,320–323] In the absence of insulin only up to ~5% of GLUT4 is present at the cell surface with this being constantly recycled intracellularly.[324]

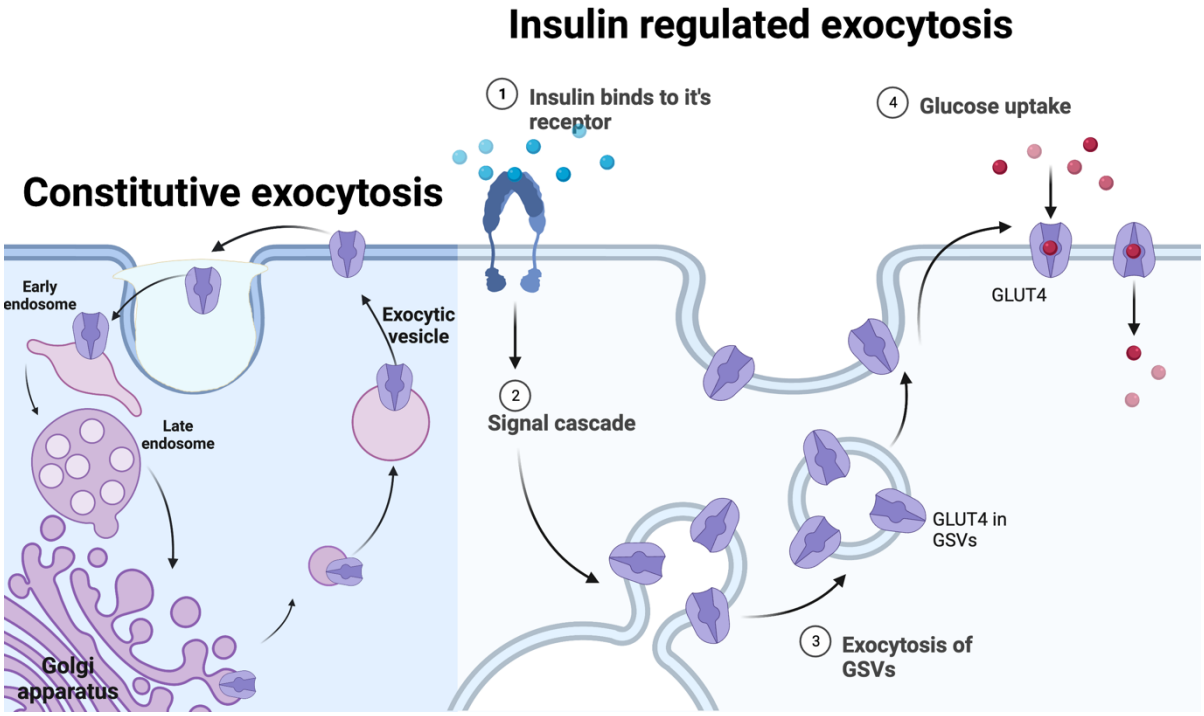


Figure 3.2: Transport and recycling of GLUT4 from the plasma membrane.
GLUT4 is trafficked to the cell surface through two routes. The majority of this is transported in the presence of insulin through insulin regulated exocytosis (right hand side) which results in the trafficking of GLUT4 secretory vesicles (GSVs) to the cell surface. The left-hand side of the figure outlines the constitutive exocytosis of GLUT4 through which ~20% of the intracellular pool is trafficked to the cell surface. This process happens in the absence of insulin. Figure generated using Biorender.com

Whilst it has been shown that dysregulation of GLUT4 results in insulin resistance and subsequently contributes to T2D pathogenesis, there is still much to be understood about the regulation of this pathway. Therefore, studies to identify novel players in GLUT4 regulation and trafficking may aid greater understanding into the regulation of this pathway and thus candidates for therapeutic targets for T2D.

3.3 Methods

3.3.1 Gene selection:

The genes assessed in this chapter were selected from genetic loci associated with insulin fold change (IFC) between 120 min and fasting during an oral glucose tolerance test (OGTT) in a meta-analysis of GWAS (see chapter 2).

IFC is correlated with post-challenge insulin and glucose to a greater extent than fasting measures (**See Chapter 2**). Therefore, following the identification of the *SLC2A4* locus directly implicating the encoding protein, GLUT4, it seems justifiable that studying genes at genetic loci associated with IFC may allow identification of additional novel regulators of insulin-stimulated glucose disposal. These loci included 4 genome-wide significant ($P < 5x10^{-8}$) loci identified in cohorts of European ancestry, and 1 of meta-analysis of cohorts of East Asian and Hispanic American ancestry. Additional 5 loci were selected below genome-wide significance ($5x10^{-8} < P < 5x10^{-5}$) which show evidence of also being associated with 2 h glucose levels, and/or not associated ($P > 0.05$) with fasting insulin but evidence of being an expression quantitate trait locus in skeletal muscle (GTEX v8). All protein-coding genes defined in UCSC genome browser[325] located ±500kb of the lead IFC variant at a given locus were initially prioritised. Genes were subsequently filtered based on evidence of expression in human skeletal muscle and adipose tissue (GTEx v8) – the primary tissues where insulin-stimulated GLUT4 trafficking occurs.[273] Evidence of expression was considered where transcripts per million (TPM) greater than 1 was reported at least one of these tissues for a given gene.

The expression of these genes in differentiated 3T3-L1 adipocytes were further assessed in a publicly available RNA-sequencing dataset (GEO accession:GSE129957)[326], as well as an in house dataset from the Fazakerley Lab (Unpublished). A threshold of Fragments Per Kilobase of transcript per Million mapped reads (FPKM) >1 was used as evidence of expression and confirmed in the 3T3-L1 adipocyte line used in this assay using qPCR (see below).

36 genes at genetic loci associated with IFC were depleted using short interfering RNA (siRNA) and the effect of gene knockdown on insulin-stimulated GLUT4 translocation was assessed in 3T3-L1 adipocytes (**Supplementary Table S3.1**). These experiments

were conducted in 3T3-L1 adipocytes stably expressing a GLUT4 construct (HA-GLUT4-mRuby3), and complementary experiments were conducted in wild type 3T3-L1 adipocytes to ensure findings were relevant to endogenous GLUT4.

3.3.2 Assay Overview

In order to assess the role of candidate novel genes in GLUT4 trafficking I used an assay based in 3T3-L1 adipocytes, a murine cell line. Whilst this cell line is not in muscle, this cellular model has several advantages of available muscle cell lines for measuring GLUT4 levels. 3T3-L1 adipocytes are highly insulin responsive and endogenously express GLUT4 and have required machinery for GLUT4 trafficking.[324] 3T3-L1 adipocytes are a long-established cell line for assessing glucose disposal at the cellular level.[327–329]

Whilst not a myocyte cell line, despite muscle being the primary tissue of glucose disposal, there is low expression of GLUT4 in available cell lines and these do not reflect biology of muscle cells *in vivo*. Further, while a number of human adipocyte lines are now becoming available, these are currently less reliable in terms of adipocyte differentiation and less characterised than the 3T3-L1 model, which has been the workhorse cell line for studies into GLUT4 traffic for a number of years.

3.3.3 Cell culture

3T3-L1 preadipocytes were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% foetal bovine serum (FBS) and GlutaMAX in a humidified incubator at 37 °C with 10% CO_2. For 3T3-L1 stable cell lines, preadipocytes were transduced with pBABE-puro retrovirus expressing HA-GLUT4-mRuby3. Puromycin (2 µg/mL) was used to select transduced cells. 3T3-L1 preadipocytes were seeded into 6 well plates and were differentiated into adipocytes. Differentiation was initiated at 100% confluence through addition of DMEM/FBS/GlutaMAX supplemented with 350 nM insulin, 0.5 mM 3-isobutyl-1-methylxanthine (IBMX), 250 nM dexamethasone, and 400 nM biotin. After 3 d, cells were moved to DMEM/FBS/GlutaMAX supplemented with 350 nM insulin for a further 3 d. Cells were reverse transfected with control or targeting siRNAs 6 d after the initiation of differentiation.

3.3.4 siRNA-mediated gene knockdown in differentiated adipocytes

Dharmacon siGenome siRNA pools (Horizon Discovery) to genes-of-interest (see **Supplementary Table S3.2** for details);*Akt1 and Akt2, Tbc1d4, Rab10* as positive controls; and non-targeting control (NT) siRNA were used. siRNAs were resuspended in RNAse-free water in a laminar flow hood in RNAse free environment following manufacturer specifications. The concentration of resuspended siRNA was measured using Nanodrop and standardized to 25 µM. Aliquots of diluted siRNA were stored at -80°C and thawed on ice prior to use.

For reverse transfection, 2.5 µL of TransIT-X2 transfection reagent and 1.25 µL of siRNA (25 µM) was added to 63 µL Opti-MEM and incubated at room temperature for 30 min. 3T3-L1 adipocytes 6 days post-differentiation were trypsinised in 10x trypsin-EDTA at 37°C, resuspended in DMEM/FBS media and centrifuged at 200 x *g* for 5 min at room temperature. Pelleted cells were resuspended in DMEM/FBS/GlutaMAX (13.5 mL per 6-well plate) and 563 µL of resuspended cells were added to each tube of transfection reagent. Media containing cells and siRNA complexes were reseeded into Matrigel-coated 96-well plates (95 µL per well). For seeding of cells into 24- or 48-well plates, the amounts of OptiMEM, transfection reagent, siRNA and cells were adjusted appropriately based on well area. After 24 h, cells were washed with DMEM/FBS/GlutaMAX and media was replaced every 24 h. Cells were assayed 96 h post-knockdown.

Knockdown of positive control genes, where involvement in GLUT4 trafficking is well established, was validated at the protein level by western blot (**Figure 3.3**). Knockdown of genes that were identified to be of interest following GLUT4 translocation assay were validated at the mRNA level by qPCR. See below for further details.

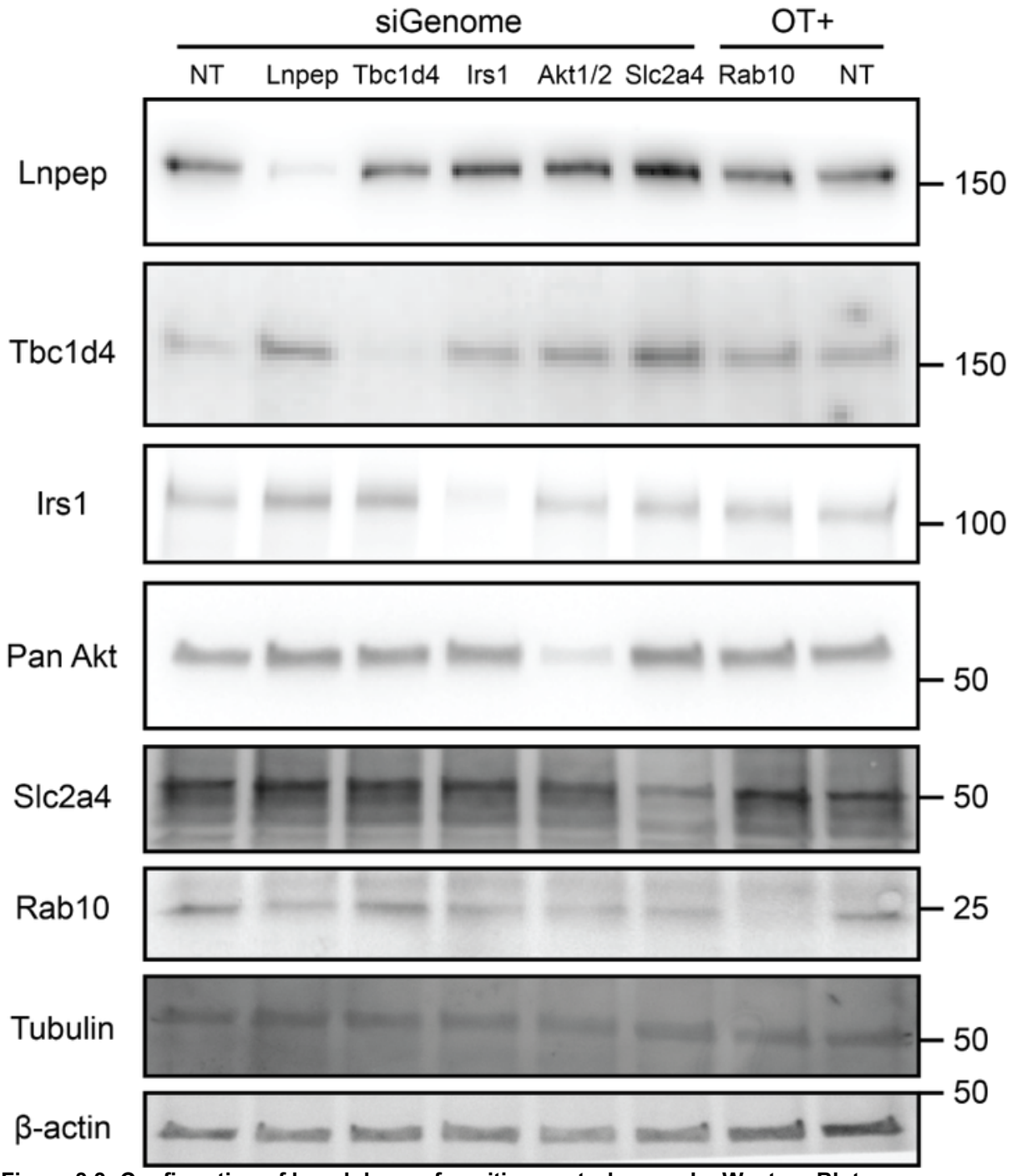


Figure 3.3: Confirmation of knockdown of positive control genes by Western Blot.
Representative blot from N=2. Protein size in kDa is outlined on the right-hand side of the blot. siGenome and OT+ represent siRNA pools with their corresponding targets indicated below (see methods) and NT denotes non-targeting control. Antibodies are outlined on the left-hand side of the blot with Tubulin and B-actin used as loading controls. Western blotting was performed by Dr Dougall Norris.

3.3.5 GLUT4 translocation assay

Assay design

siRNA knockdown in each experiment was performed in technical duplicates. Non-targeting (NT) siRNA was used as the control condition in all experiments, with three sets of NT control per assay plate. Positive control siRNA targeting *Akt1 and Akt2, Tbc1d4, Rab10* mRNA were included in all experiments.

Insulin stimulation and immunofluorescence staining

96 h after transfection cells were washed 3x in serum-free DMEM and incubated in DMEM/GlutaMAX/0.2% BSA for 2 h. Cells were stimulated with 0.5 nM or 100 nM insulin for 20 min. Cells were washed by gently dunking the 96-well plates 12 times in beakers containing ice-cold PBS (6 washes in each of two 1 L beakers containing PBS; all the subsequent PBS washes were performed using this method). Plates were placed on ice and residual PBS was removed with a multichannel pipette. Cells were fixed with 4% paraformaldehyde (PFA) for 5 min on ice, and 15 min at room temperature. PFA was quenched with 50 mM glycine in PBS for 10 min. Cells were washed in PBS and blocked with 5% Normal Swine Serum (NSS, Dako, X0901) in PBS for 20 min. Cells were then incubated with primary antibody solution in 2% NSS in PBS for 1 h at room temperature. For detecting endogenous surface GLUT4 in wild type 3T3-L1 adipocytes, cells were incubated with human anti-GLUT4 antibody (LM048; antibody kindly provided by Joseph Rucker, Integral Molecular, PA, USA)[330]. For detecting surface HA-GLUT4-mRuby3, cells were incubated with mouse anti-HA (1:500, Covance, MMS-101R). For TfR surface labelling, primary antibody against TfR (CD71, Invitrogen) and secondary antibody Alexa-647 conjugated Goat anti-rat IgG (Life Technologies; 1:500) were used.

After PBS washes, cells were incubated with secondary antibody solution in 2% NSS in PBS containing Hoechst 33342 (Life Technologies) for 1 h at room temperature. For detecting endogenous GLUT4, we used Alexa-488-conjugted anti-human IgG antibody (Life Technologies). For HA-GLUT4-mRuby3 expressing cells, we used Alexa-488-conjugted anti-mouse IgG antibody. (Life Technologies). Cells were washed 12 times in RT PBS and stored in PBS containing 2.5% DABCO, 10% glycerol, pH 8.5, sealed and kept at 4°C in the dark prior to imaging. Plates were equilibrated to room temperature for 30 min before imaging.

Plates were imaged on the Perkin Elmer Opera Phenix High Content Screening System and mid-section confocal images were obtained using a 20x water objective (N.A 1.0), with 2-pixel binning. Excitation wavelengths and emission filters used were as follows: endogenous surface GLUT4 and HA-tagged surface GLUT4: 488 nm, 500-550 nm; endogenous total GLUT4 and mRuby-tagged total GLUT4: 561 nm, 570-630 nm; Hoechst: 405 nm, 435-480 nm, TfR: 640 nm, 650-790 nm. Lipid droplets were visualised using direct phase contrast (DPC).

To quantify total endogenous GLUT4 expression, in WT 3T3-L1 adipocytes, cells were permeabilised and stained with anti-GLUT4 antibody after plates had been imaged for surface GLUT4. Cells were permeabilised in 5% NSS in PBS containing 0.1% saponin for 20 min, followed by incubation with rabbit anti-GLUT4 antibody (Courtesy of Professor David James, University of Sydney)[331]. After washing with PBS, cells were incubated with 2% NSS in PBS containing Alexa-568-conjugated goat anti-rabbit IgG antibody (Life Technologies) for 1 h at room temperature. Plates were then re-imaged on the Perkin Elmer Opera Phenix High Content Screening System as described above.

Imaging data analysis

Imaging data were analysed using Harmony phenoLOGIC Software (v4.9; Perkin Elmer). For endogenous GLUT4 indicated by the Alexa 488 signal was quantified for surface GLUT4 prior to permeabilisation and total GLUT4 following permeabilisation and re-staining with Alexa-568 (see above). Surface GLUT4 was quantified in HA-GLUT4-mRuby cells by Alexa-488 antibody signal and total GLUT4 by the mRuby signal. Nuclei were quantified using the Hoechst signal, and nuclei were defined and quantified using pre-set method A. The surface GLUT4 signal was subsequently normalised to nuclei number. The ratio of surface to total GLUT4 levels were calculated from the surface GLUT4 signal normalised to the total GLUT4 signal. The same protocol was followed for TfR with the Alexa-647 antibody signal.

Surface and total GLUT4 and TfR were quantified by taking the mean intensity across the entire field of view – the fluorescence intensity across 8 fields of view were then averaged to provide a final fluorescence intensity for each well. The intensity across two technical replicates (wells on separate 96-well plates) were then averaged providing a mean intensity for 16 fields of view, per biological replicate. The surface

GLUT4 levels were subsequently normalised to nuclei number. The ratio of surface to total GLUT4 levels were calculated from the surface GLUT4 levels normalised to the total GLUT4 signal.

Lipid droplets were quantified using digital phase contrast (DPC) signal which was quantified using the intensity of the signal across number of pixels per image then averaged across mean of the 6-8 images per well.

Data processing and statistical analysis

Background florescence signal, without antibody stain, was subtracted from each measure and experimental measures per plate were normalised to the average value of all conditions across each plate. Differences in mean measures for a given knockdown were compared to the non-targeting control values using a two-sided *t*-test using t.test() function in R (v4.0.4) across all replicates (N biological endogenous GLUT4 = 3; N biological replicates HA-GLUT4-mRuby = 5, 2 technical replicates per N). Significance was considered at a false discovery rate of 5% (FDR (q) < 0.05). Candidate genes which were significantly different to control (FDR < 0.05) for surface/nuclei number and/or surface/total GLUT4 in at least 1 of the assayed cell lines and directionally consistent effects in the other compared to NT, were considered as robust findings in the assay. All genes identified to have robust effects were taken forward for Radiolabelled Glucose Transport Assays.

3.3.6 Radiolabelled glucose transport assay

Assay conducted by Dr Daniel Fazakerley

Following siRNA-knockdown, 3T3-L1 adipocytes were serum-starved for 2 h in DMEM containing 0.2% BSA and GlutMAX media in a humidified incubator, 37°C 10% CO_2. Cells were washed three times in pre-warmed Krebs-Ringer buffer (KRP; 0.6 mM Na_2HPO_4, 0.4 mM NaH_2PO_4, 120 mM NaCl, 6 mM KCl, 1 mM $CaCl_2$, 1.2 mM $MgSO_4$ and 12.5 mM HEPES (pH 7.4)) at 37°C. Cells were stimulated with 0.5 nM insulin for 20 min where indicated. To determine the non-specific glucose uptake 25 µM cytochalasin B (Sigma Aldrich) was added to the wells before the addition of H^3-2-deoxyglucose (2DOG). During the final 5 min of insulin stimulation, 2DOG (50 µM unlabelled 2DOG) was added to wells to measure glucose uptake rates. Following three washes in ice-cold PBS, cells were solubilized in 1% (w/v) Triton X-100 in PBS

on a shaker for 1 h. Tracer uptake was assessed by liquid scintillation counting using a β-scintillation counter. Values for all samples were normalized to protein content.

3.3.7 Confirmation of gene knockdown

RNA extraction

RNA was extracted from 3T3-L1 adipocytes cultured in 24-well plates 4 d after siRNA knockdown (see above). Cells were washed 2x in PBS and TRI reagent (Zymo Research) was added to each well. Cells were then flash frozen in the plate on dry ice and stored at -80°C. RNA extraction was conducted in an RNAse free environment using the Zymo Research Directzol RNA Extraction Kit following standard manufacturer's protocol. Briefly, plates were thawed on ice and cells were lysed by manual cell scraping and incubation at room temperature for 10 min. Samples were mixed with an equal volume of 100% and added directly to an RNA extraction column. All centrifugation steps were conducted at 12,000 xg in a microcentrifuge. RNA was eluted in 30 µL of RNAse and DNAse free water and RNA concentration was assessed by UV spectrophotometry at 260nm using NanoDrop Spectrophotometer ND-1000 (Thermo Scientific).

Reverse transcription

RNA was diluted to a consistent concentration per biological replicate. Reverse transcription was conducted in a volume of 50 µL for each sample. Random primers (Promega) were annealed in a total volume of 30 µL at 70°C for 5 min, and immediately rapidly cooled on ice, in line with manufacturers' specifications. Reverse transcription was conducted in a total volume of 50 µL using MMLV Reverse Transcriptase (Promega) and RNasin RNA inhibitor (Promega). Reverse transcription was conducted using BioRad Tetrad 2 Thermal Cycler at 25°C for 10 minutes, 37°C for 1 h and 85°C for 10 min before being stored at -20°C.

qPCR RNA quantification

Quantitative PCR was conducted using Thermo Fisher TaqMan gene expression assays following the standard manufacturer's protocol. Probe specifications for targets of interest are outlined in **Table 3.1.** Briefly, the pre-mixed primer sets and probes were added to master mix containing TaqMan 2x Universal PCR Master Mix (ThermoFisher) and nuclease-free water. Polymerase chain reaction was carried out using QuantStudio 7 or QuantStudio 5 Real-Time PCR system (Thermo Fisher Scientific).

ActB (Mm04394036_g1) and *Canx* (Mm00500330_m1) were used as housekeeping controls for all samples. The standard pre-set cycling conditions for TaqMan reagents were used – 10 mins at 95ºC for polymerase activation followed by 40 PCR cycles (95ºC for 15 seconds and 60ºC for 1 minute).

Table 3.1 - TaqMan probes (Thermo Fisher) used to quantify mRNA expression

Gene	TaqMan Probe ID
ActB	Mm04394036_g1
Canx	Mm00500330_m1
Slc2a4	Mm00436615_m1
Rab10	Mm00489481_m1
Akt1	Mm01331626_m1
Akt2	Mm00545827_m1
Tbc1d4	Mm01231395_m1
Llgl1	Mm00839427_g1
Flii	Mm01234001_g1
Drg2	Mm00840194_g1
Pdzk1ip1	Mm00522487_m1
Slc36a4	Mm00552274_m1
Vps13c	Mm01169975_m1
Alkbh5	Mm00841140_m1
Cmpk1	Mm00727806_s1
Tal1	Mm00441665_m1
Tnks	Mm00558256_m1
Tvp23b	Mm00509270_m1
Tom1l2	Mm00462049_m1
Bicc1	Mm00473345_m1

qPCR data analysis

Double delta Cycle Threshold values (ddCT) were calculated and the fold change relative to the non-targeting (NT) control was used to assess knockdown efficiency. The mean CT value of the housekeeping genes *ActB* and *Canx* were used as housekeeping controls for normalisation for a given sample. One-way ANOVA was used to compare delta CT (dCT) values for a given knockdown target compared to NT control, and significance was considered at FDR < 0.05.

SDS-PAGE and immunoblotting

Typically, 10 µg of protein was resolved by SDS-PAGE and transferred to PVDF or nitrocellulose membranes. Membranes were blocked in 5% BSA or skim milk powder in TBS-T (0.1% v/v Tween-20 in Tris-buffered saline) for 1 h followed by overnight incubation at 4°C with specific primary antibodies (listed below). Following incubation with horseradish peroxidase (HRP)-conjugated anti-rabbit or mouse immunoglobulin G (IgG) or Alexa-647-conjugated anti-mouse IgG secondary antibodies for 1 h. Protein bands were visualised using ECL (Thermo Scientific) or 647-fluorescence intensity on

the Chemidoc MP (Bio-Rad). The following antibodies were used: Lnpep/IRAP (Cell Signaling Technologies, #6918); TBC1D4 (Cell Signaling Technologies, #2670); IRS1 (Cell Signaling Technologies, #3407); PanAkt (Cell Signaling Technologies, #2920); GLUT4 (polyclonal antibody recognising the C-terminus of GLUT4 provided by Professor David James, University of Sydney)[331]; Rab10 (Thermo Fisher; # MA5-15670); alpha-tubulin (Merck; #T9026) and beta-actin (Cell Signaling Technologies; #8457).

3.4 Results

3.4.1 Rationale for studying genes at loci associated with insulin fold change (IFC)

IFC was associated with a total of 10 loci from both the genetic discovery and integrated GWAS approach associated with post-challenge insulin resistance in cohorts of European ancestry (**see Chapter 2**). This included the *SLC2A4* locus directly implicating the encoded protein GLUT4. IFC further appears more specific to the post-challenge state observationally with IFC being more strongly correlated than ISI with 2 h measures of insulin and glucose in the Fenland Study (**see Chapter 2**). Therefore, it was hypothesised that characterising genes at genetic loci associated with IFC, associated at genome-wide significance or identified in the integrated approach, may identify additional novel regulators of insulin-stimulated glucose disposal. A total of 36 genes were selected following this strategy and assessed for the impact of gene knockdown on insulin-stimulated GLUT4 translocation in two 3T3-L1 adipocyte cell lines. Wildtype 3T3-L1 adipocytes allowed the study of endogenous GLUT4 and the second cell line where GLUT4 reporter construct (HA-GLUT4-mRuby) was introduced allowed the study of total and surface GLUT4 simultaneously without further permeabilization **(Methods).**

The results from wildtype (WT) 3T3-L1 adipocytes, at 0.5nM insulin – more reflective of the endogenous GLUT4 responses to a physiological insulin dose, were used as the primary results at a threshold of FDR < 0.05. Effects that that were also directionally replicated for surface GLUT4/nuclei number or surface/total GLUT4 in HA-GLUT4-mRuby cells were taken forward as significant and robust findings.

3.4.2 Evaluation of the system used

Using this approach, knockdown of the kinases *Akt1* and *Akt2* as well as the Rab protein *Rab10* impaired insulin-stimulated HA-GLUT4-mRuby and endogenous GLUT4 translocation, as expected (**Supplementary Table S3.3 Supplementary Table S3.5, Figure 3.4**).[327,332] *Tbc1d4* knockdown suggested an increase in cell surface GLUT4, consistent with its role as a negative regulator of GLUT4 translocation[237,288,333] but this did not reach FDR threshold FDR < 0.05.

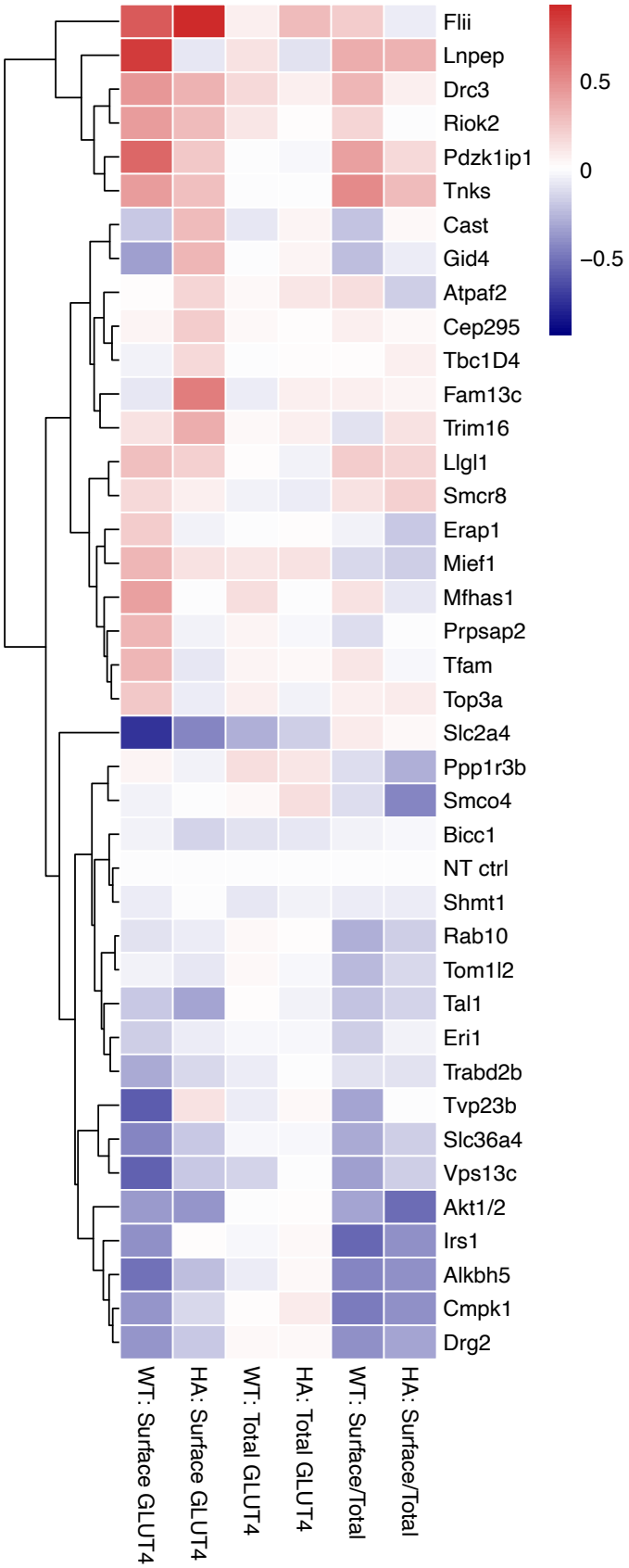


Figure 3.4: GLUT4 screen results for 3T3-L1 WT and 3T3-L1 HA-GLUT4-mRuby cells for all targets

All values are normalised to NT control (0) for the relevant measure/cell line combination. Targets with values lower than NT control are shown in blue and those with higher values in red. Values are scaled within a measure/cell line pair and therefore scales are not directly comparable. Clustering of targets was done using hierarchical clustering. Values are those measured at 0.5nM insulin.

Based on testing a subset of 20 siRNAs by qPCR, including those for genes where depletion had no effect on GLUT4, the average knockdown efficiency was 85% (**Figure 3.5; Table 3.2**). These results provide confirmation that the system we have utilised to assess impact on insulin stimulated GLUT4 trafficking works and provides robust results.

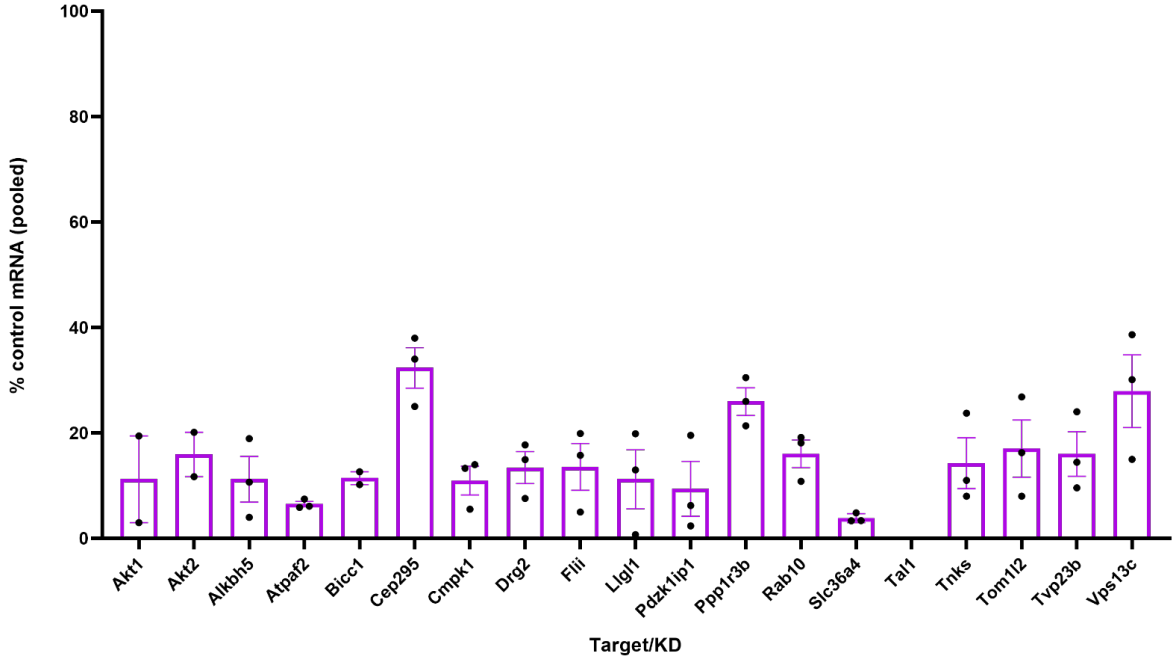


Figure 3.5: Average percentage of mRNA detected in knockdown samples compared to pooled control RNA
Bars represent average knockdown of genes of interest in knockdown samples compared to pooled control RNA. Individual points represent individual replicates, and error bars represent +/- SE.

Table 3.2: quantitative PCR (qPCR) assessment of knockdown efficiency of targets in 3T3-L1 adipocytes

						one-way ANOVA - comparison of dCT		
KD sample	**Probe**	**N**	**Mean expression (%)***	**SD**	**SEM**	**Mean fold change in expression: pooled sample vs target**	**P value**	**q value (FDR adjusted P value)**
Akt1/2	Akt1	2	11.21	11.61	8.21	3.709	<0.0001	0.0002
Akt1/2	Akt2	2	15.93	5.946	4.204	2.842	0.0008	0.001
Alkbh5	Alkbh5	3	11.2	7.475	4.316	3.44	<0.0001	<0.0001
Atpaf2	Atpaf2	3	6.49	0.8646	0.4992	3.688	<0.0001	<0.0001
Bicc1	Bicc1	2	11.42	1.699	1.201	3.138	0.0007	0.001
Cep295	Cep295	3	32.33	6.658	3.844	1.657	0.0251	0.0263
Cmpk1	Cmpk1	3	10.95	4.697	2.712	3.305	<0.0001	<0.0001
Drg2	Drg2	3	13.42	5.249	3.031	2.986	<0.0001	0.0002
Flii	Flii	3	13.55	7.688	4.439	3.105	<0.0001	0.0002
Llgl1	Llgl1	3	11.2	9.694	5.597	4.142	<0.0001	<0.0001
Pdzk1ip1	Pdzk1ip1	3	9.395	9.005	5.199	3.914	<0.0001	<0.0001
Ppp1r3b	Ppp1r3b	3	25.94	4.566	2.636	1.962	0.0084	0.0098
Rab10	Rab10	3	16.03	4.537	2.62	2.685	0.0004	0.0006
Tal1	Tal1	3	not detected - expression seen in MGI mouse adipose tissue data (http://www.informatics.jax.org)					
Slc2a4	Slc2a4	3	19.05	11.73	6.775	2.48	0.0011	0.0013
Slc36a4	Slc36a4	3	3.845	0.8467	0.4889	2.767	0.0013	0.0011
Tnks	Tnks	3	14.25	8.355	4.824	2.978	0.0001	0.0002
Tom1l2	Tom1l2	3	17.03	9.44	5.45	2.706	0.0004	0.0006
Tvp23b	Tvp23b	3	16.02	7.328	4.231	2.751	0.0003	0.0006
Vps13c	Vps13c	3	27.92	11.97	6.91	1.94	0.0091	0.0101

* % mRNA expression compared to pooled control sample

3.4.3 Known regulators of GLUT4 traffic were genetically prioritised and shown effect GLUT4 trafficking

Four of the genes from the list of 36 genes-of-interest already implicated in insulin signalling and GLUT4 trafficking, as well as GLUT4 itself (*Slc2a4*), were identified as significantly (FDR < 0.05) altering cell surface GLUT4 abundance: *Irs1*[334], *Lnpep*[335], *Tnks*[336,337] and *Vps13c*[338]. These five genes acted as internal controls on top of the positive control genes described above, being selected due to their location at loci associated with IFC despite their role in GLUT4 trafficking already being reported. We were blinded to the identity of all genes during the experiments, with their identity only revealed upon data analysis.

Irs1, encoding insulin receptor substrate 1, lies upstream of *Akt1/2* and is involved in transmitting signals from insulin receptor to the PI3K/AKT pathway. Depletion/deletion of Irs1 is established to result in a reduction in GLUT4 at the cell surface[334]. This observation is replicated in this experiment, both in wildtype 3T3-L1 and HA-GLUT4-mRuby cell lines (**Figure 3.4, 8, Table 3.3, Supplementary Tables 3.4 and 3.5**), with a significant reduction in surface GLUT4 levels being identified at 0.5nM insulin compared to non-targeting control. No effect for total GLUT4 was determined. Although not discernible for surface GLUT4 alone, a trafficking effect of Irs1 knockdown on reducing the ratio of surface/total GLUT4 is also discernible when stimulated with 100nM insulin.

LNPEP, encoding IRAP, colocalises with GLUT4 on GSVs and interacts with TBC1D4, a negative regulator of GLUT4 traffic.[271,288] When TBC1D4 is phosphorylated by AKT, it dissociates from LNPEP and allows translocation of GLUT4 to the cell surface. Consistent with previous reports[271], knockdown of *Lnpep* in 3T3-L1 adipocytes results in a significant increase in surface GLUT4 upon insulin simulation (**Table 3.3, Supplementary Table S3.4**).

TNKS is also thought to co-localise with GSVs, being shown to directly interact with LNPEP, and therefore it has been suggested that Tnks effects GLUT4 trafficking through it's regulation of LNPEP[336,339]. It has previously been shown that double knockdown of *Tnks* (encoding TNKS1) and *Tnks2* (encoding TNKS2) in 3T3-L1 adipocytes results in a reduction of GLUT4 translocation to the cell surface compared to control upon insulin stimulation.[340] siRNA knockdown of *TNKS* in a skeletal muscle

cell line also identified a reduction of overall GLUT4 levels upon insulin stimulation.[340] However, in these experiments, where only siRNA knockdown of Tnks (encoding TNKS1) alone was done, an increase in GLUT4 at the cell surface at both 0.5nM and 100nM insulin was observed, inconsistent with this result (**Figure 3.6**, **Table 3.3, Supplementary Table S3.4)**. In mice where only *Tnks* was depleted, increased glucose uptake and elevated GLUT4 protein expression in white adipose tissue has been reported[340]. In primary adipocyte cultures from these deficient mice increased GLUT4 was identified at the plasma membrane and an increase in glucose uptake compared to wildtype adipocyte cultures.[340] Consistent effects were also seen in an independent radio-labelled glucose transport assay (see below).

VPS13C, vacuolar protein sorting 13 homolog c, has been associated to T2D through robust association of a genetic locus near *VPS13C* with T2D and glycaemic traits[72,80,87]. Previous studies where *Vps13c* has been depleted in C2C12 myoblasts showed an increase in GLUT4 at the cell surface compared to control upon stimulation with AICAR, an AMPK activator that is able to stimulate translocation of GLUT4 to the cell surface[340]. VPS13C was further shown to be a interactor of TBC1D1, a protein closely related to TBC1D4 in this experiment[340].

This is in contrast with what this study identified, with knockdown of *Vps13c* resulting in a consistent decrease in both total and surface GLUT4 levels in WT 3T3-L1 adipocytes compared to control when stimulated with insulin (**Figure 3.6**, **Table 3.3, Supplementary Table S3.4)**. This difference in effect seen may be due to differences both in cell type and stimulation. These results are further consistent with the effects seen in the GWAS for this locus on insulin fold change (**see Chapter 2**) and other glycaemic traits such as 2 h glucose[87]. Consistent effects were also seen in an independent glucose transport assay (see below).

3.4.4 Novel genes involved in GLUT4 trafficking

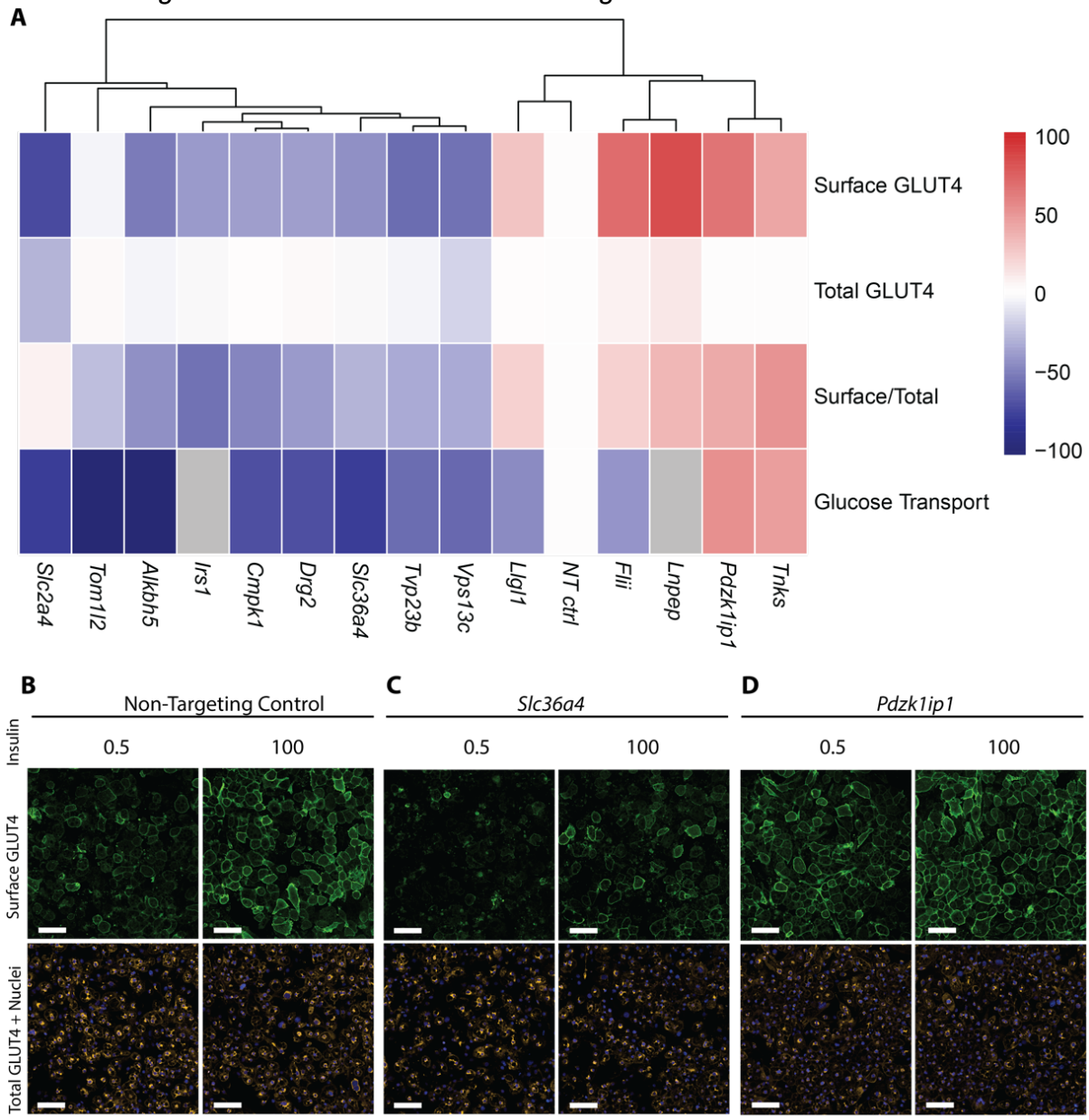


Figure 3.6: 14 genes at loci associated with insulin fold change were identified to significantly impact GLUT4 trafficking in adipocytes

a) Effect of gene knockdown on GLUT4 trafficking and glucose transport at 0.5 nM insulin stimulation in WT 3T3-L1 adipocytes. Genes are clustered using hierarchical clustering. All results are normalised to non-targeting (NT) control for a given measure, with scale representing those where there is an increase compared to NT control and blue a decrease, as indicated by figure legend on right. Scale represents the relative change in percentage compared to NT (NT scaled to 0) for each condition. with this subsequently scaled to the % of the absolute maximum response across all measures to allow direct comparison across these. Grey represents genes for which glucose transport data was not collected as their role is established in the literature. b) Fluorescence microscopy images of selected example control knockdown using non-targeting control siRNA in WT 3T3-L1 adipocytes. This was used as the comparator group for gene-specific knockdown c) Fluorescence microscopy images of selected example knockdown of *Slc36a4* in WT 3T3-L1 adipocytes compared to non-targeting (NT) control (Figure 4b). d) Fluorescence microscopy images of selected example knockdown of *Pdzk1ip1* in WT 3T3-L1 adipocytes compared to NT control (Figure 4b). In Figures 4B-D green represents the surface GLUT4 signal, orange the total GLUT4 signal and blue indicates individual nuclei.

A further 9 genes (*Alkbh5, Cmpk1, Drg2, Flii, Llgl1, Pdzk1ip1, Slc36a4, Toml12, Tvp23b*) whose knockdown significantly (FDR < 0.05) altered absolute cell surface GLUT4 abundance and/or the ratio of cell surface GLUT4-to-total GLUT4 protein upon insulin stimulation, indicative of dysfunctional insulin-stimulated GLUT4 translocation, were identified **Figure 3.6**, **Table 3.3, Supplementary Tables 3.4**. To our knowledge, these genes have not been previously implicated in insulin action or GLUT4 trafficking in metabolically relevant cell types and therefore warrant future mechanistic investigation.

Table 3.3 - GLUT4 screen results in 3T3-L1 WT cells: Prioritised and Control Genes

		Surface/Total GLUT4				Surface GLUT4/ Nuclei number				totalGLUT4/Nuclei number				
name	**Insulin (nM)**	**Mean**	**SE**	**P**	**FDR**	**Mean**	**SE**	**P**	**FDR_adj_P**	**Mean**	**SE**	**P**	**FDR**	**Prioritised?**
Akt1/2	0.5	0.717	0.129	1.36E-05	1.54E-04	0.629	0.119	3.66E-03	1.68E-02	0.974	0.039	9.75E-01	9.94E-01	Control
Alkbh5	0.5	0.608	0.070	1.67E-05	1.54E-04	0.473	0.082	1.98E-03	1.14E-02	0.952	0.034	6.68E-01	8.59E-01	Y
Cmpk1	0.5	0.572	0.038	3.31E-06	5.08E-05	0.611	0.150	2.35E-02	5.68E-02	1.014	0.055	4.05E-01	6.97E-01	Y
Drg2	0.5	0.640	0.014	6.37E-05	4.19E-04	0.619	0.139	2.66E-02	6.11E-02	0.988	0.009	7.66E-01	9.30E-01	Y
Flii	0.5	1.264	0.083	2.35E-02	5.68E-02	1.691	0.134	2.41E-05	3.70E-04	1.055	0.050	9.47E-02	2.56E-01	Y
Irs1	0.5	0.487	0.046	4.56E-08	2.10E-06	0.588	0.102	1.63E-02	4.41E-02	0.927	0.073	3.45E-01	6.61E-01	Y
Llgl1	0.5	1.273	0.240	1.84E-02	4.71E-02	1.277	0.237	7.76E-02	1.37E-01	1.015	0.036	3.96E-01	6.97E-01	Y
Lnpep	0.5	1.399	0.081	3.09E-04	1.58E-03	1.838	0.487	3.99E-07	1.84E-05	1.118	0.121	3.11E-03	2.74E-02	Y
NT ctrl	0.5	1.040	0.024	NA	NA	0.986	0.057	NA	NA	0.973	0.017	NA	NA	Control
OT_Rab10	0.5	0.635	0.036	7.53E-04	3.07E-03	0.268	0.221	3.42E-04	3.01E-03	0.957	0.008	7.88E-01	9.30E-01	Control
Pdzk1ip1	0.5	1.445	0.177	5.12E-05	3.93E-04	1.649	0.160	6.98E-05	8.03E-04	0.983	0.044	8.42E-01	9.44E-01	Y
Rab10	0.5	0.752	0.072	9.72E-05	5.59E-04	0.888	0.099	4.24E-01	5.74E-01	1.034	0.048	9.33E-02	2.56E-01	Control
Slc2a4	0.5	1.133	0.079	3.45E-01	4.17E-01	0.253	0.073	1.17E-05	2.70E-04	0.679	0.066	4.20E-09	1.93E-07	Y
Slc36a4	0.5	0.742	0.038	2.79E-03	8.01E-03	0.555	0.225	9.33E-03	3.30E-02	0.942	0.035	5.17E-01	7.86E-01	Y
Tbc1d4	0.5	1.065	0.073	7.09E-01	7.37E-01	0.944	0.119	7.19E-01	8.35E-01	0.973	0.032	9.94E-01	9.94E-01	Control
Tnks	0.5	1.553	0.149	3.56E-07	8.18E-06	1.427	0.094	7.65E-03	2.93E-02	0.940	0.060	4.90E-01	7.77E-01	Y
Tom1l2	0.5	0.788	0.059	1.11E-02	3.00E-02	0.952	0.156	8.39E-01	8.85E-01	1.025	0.073	2.87E-01	5.92E-01	Y
Tvp23b	0.5	0.715	0.075	1.12E-03	3.97E-03	0.396	0.177	3.93E-04	3.01E-03	0.903	0.041	1.52E-01	3.69E-01	Y
Vps13c	0.5	0.705	0.079	7.77E-04	3.07E-03	0.425	0.237	7.46E-04	4.90E-03	0.808	0.086	7.81E-04	1.72E-02	Y

Mean values denote the mean of normalised values relative to background control per experiment at the respective insulin dose
* T-test P value for given target at indicated insulin dose, comparing target to NT ctrl was calculated at 0.5nM. Full results at for all genes at all insulin doses are outlined in **Supplementary Table S3.4**.

One clue to how each gene might influence insulin-stimulated glucose uptake comes from comparison of data on total GLUT4 and cell surface GLUT4, alongside. For example, knockdown of *Alkbh5* significantly lowered cell surface endogenous GLUT4 abundance as well as a trend in total GLUT4 (**Figure 3.6**, **Table 3.3, Supplementary Tables 3.4)**. This is also the case for *Flii* where knockdown resulted in a significant increase of total levels and surface levels of GLUT4 compared to control cells (**Figure 3.6**, **Table 3.3, Supplementary Tables 3.4)**.

Alternatively, knockdown conditions that altered cell surface GLUT4 abundance without overt changes in total cellular GLUT4 suggest that genes-of-interest may be more directly involved in regulated GLUT4 traffic. Examples here include Llgl1, which is a known interactor of the GLUT4 regulator *Rab10*, as well-known regulators of membrane traffic via the Golgi (*Drg2, Tvp23b)* and endosomes (*Pdzk1ip1, Toml12*), both key sites of GLUT4 traffic as discussed in the introduction to this chapter. The experimental and human genetic evidence supporting a role for these genes in glucose disposal provides impetus for further investigation into their mechanism-of-action and whether they also affect insulin-stimulated glucose uptake in muscle. The literature evidence for these genes in light of these observations are summarised below.

3.4.5 Glucose transport

The results in this paragraph were generated by Dr Daniel Fazakerley

Knockdown of 7 of the 9 novel genes also altered insulin-stimulated 2-deoxyglucose (2DOG) uptake in a directionally consistent manner to their observed effects on cell surface GLUT4 (**Figure 3.6 and 3.8; Supplementary Table S3.3 and S3.7)**. In the case of *Pdzk1ip1* knockdown, both cell surface GLUT4 and 2DOG uptake were increased (**Figure 3.6a, c**, fold change surface GLUT4 Pdzk1ip1 vs NT control = 1.39, *t*-test P = 5.12×10^{-5}; fold change 2DOG Pdzk1ip1 vs NT control = 1.34, *t*-test P = 0.017; **Supplementary Table S3.3-3.7)**. Conversely, knockdown of *Alkbh5, Cmpk1, Drg2, Ppp1r11, Slc36a4, Toml12,* and *Tvp23b* decreased cell surface GLUT4 and 2DOG uptake (**Figure 3.6a, b, Figure 3.7**). Only *Flii* and *Llgl1* knockdown resulted in inconsistent effects on cell surface GLUT4 and 2DOG

uptake, with increased cell surface GLUT4 but lower 2DOG uptake (**Figure 3.6a, Figure 3.7**).

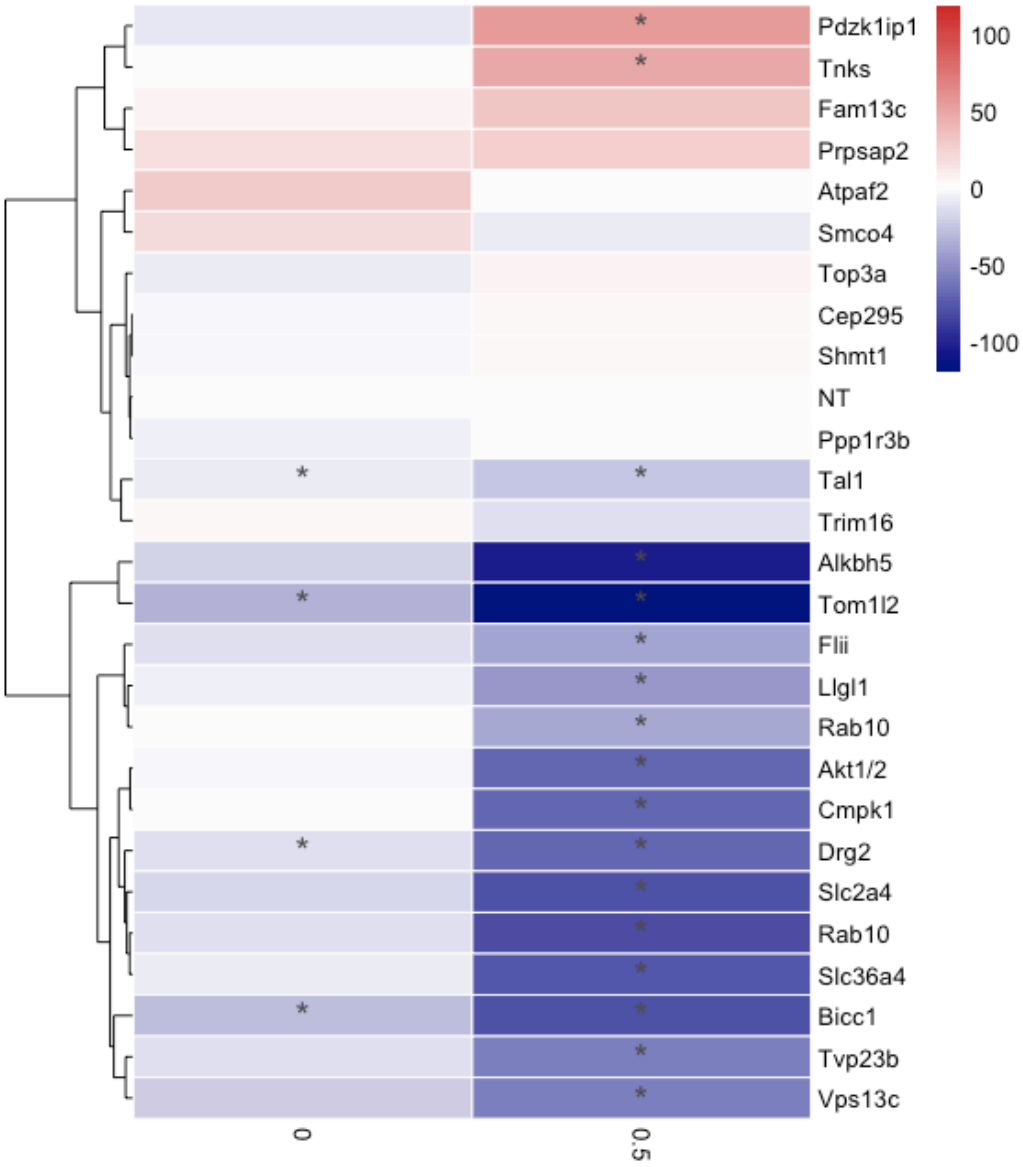


Figure 3.7: Glucose Transport relative to non-targeting control
Values are normalised glucose response to NT control per dose of insulin (nM)* Indicates targets that are significantly different from the NT control (NT) at each dose ($P < 0.05$; t-test). Scale represents the relative change in percentage compared to NT (NT scaled to 0) for each dose, with this subsequently scaled to the % of the absolute maximum response across both doses to allow comparison across these.

3.4.6 Other impacts of knockdown of the 9 novel GLUT4 regulators of interest

As well as assessing the impact of knockdown of the genes of interest on the regulation and trafficking on GLUT4, we additionally assessed additional parameters relating to adipocyte proliferation (number of cells), differentiation (e.g., lipid droplet size and number) in wildtype 3T3-L1. In addition, we assessed the

impact of knockdown on transferrin receptor, which is trafficked and recycled to the plasma membrane via similar processes as GLUT4, with the aim to identify the specificity of the effect to GLUT4 specifically.

For control gene *Tbc1d4* nominal impact on reduced lipid droplets normalised for nuclei number (LDs/Nuclei#) was identified. There was also nominal impact on lipid droplet of novel genes implicated in GLUT4 trafficking/expression: *Alkbh5, Tom1l2, Llgl1* and *Tvp23b* (**Supplementary Table S3.6**), though none of these survive multiple testing. It should be noted that this method is just one of many to assess changes in adipocyte differentiation and future work should examine this further using different metrics and methodology.

3 of the 9 novel genes knockdown appear to also significantly (FDR < 0.05) impact TfR at 0.5nM insulin, giving the initial suggestion that these genes are more generally involved in trafficking of cargo to the cell surface, rather than a specific effect on GLUT4 (*Flii*, *Slc36a4*, *Llgl1*; **Supplementary Table S3.4**) pathway. 6 of the 9 novel genes did not significantly impact TfR suggesting these effects may be more specific to GLUT4 trafficking, although future more specific mechanistic work will be needed to confirm this.

3.4.7 Possible mechanisms of action for the novel regulators of GLUT4 traffic

Through literature review and the effects observed on knockdown of the 9 novel regulators of GLUT4 trafficking, I aimed to assess possible links to processes known to be important in the regulation of GLUT4.

ALKBH5

As mentioned above, Alkbh5 knockdown resulted in a reduction of surface GLUT4 with a trend in reduction in total GLUT4 suggesting that this has an impact on regulation of the overall levels of GLUT4 in the cell. *ALKBH5* encodes AlkB Homolog 5, an m6A RNA demethylase in the same protein family as FTO and is involved in the regulation of gene expression, suggesting this gene affects GLUT4 through longer-term control of GLUT4 mRNA translation. These findings in a cellular model directly relevant to post-challenge insulin action complement findings from a recent

study in a breast cancer model that identified *GLUT4* mRNA as a ALKHB5 substrate, and reported that knockdown of *ALKBH5* decreased GLUT4 mRNA stability and protein levels[341].

FLII

Similarly, *Flii* knockdown resulted in an impact on transcription, with knockdown of Flii resulting in an upregulation of GLUT4 total and surface levels. *Flii*, encoding Flightless I, has been established to play key roles in cell migration and wound healing[342].

Further FLII has been reported to play a number of additional roles including actin remodelling, and coactivation of nuclear receptors playing a role in transcriptional activation[342,343]. For example, Flii has been identified to be a repressor of PPARy, as it binds to LRR (LXXLL motif) of PPARy in absence of ligand preventing ligand binding, preceding PPARy repression of transcriptional activity[271]. PPARy is a key transcription factor well established to be involved in T2D[271]. Flies with mutant Flii additionally show alternations in metabolic genes including those involved in lipid and carbohydrate metabolism[344].

CMPK1

Knockdown of *Cmpk1* resulted in an impact on trafficking of GLUT4 upon insulin stimulation, but not overall levels in both WT and HA-GLUT-mRuby cells. *CMPK1*, encoding cytidine/uridine monophosphate kinase 1, is involved in nucleic acid metabolism, more specifically, the formation of UMP, CMP and dCMP which are required for nucleic acid synthesis and bind P2Y receptors[345,346]. Dysfunction of nucleic acid regulation has been shown to result in insulin resistance through promotion of inflammation in T2D-relevent tissues[347]. The P2Y family of G-protein coupled receptors which are responsive to endogenous mono- and dinuceotides nucleotides including UMP, CMP and dCMP[347]. For example, the administration of $P2Y_6R$ agonists, increases glucose uptake in C2C12 myotubes and 3T3-L1 adipocytes, independent of insulin. Increased recruitment of GLUT4 to the plasma membrane was also seen in these cell lines[348]. This is directionally consistent with the observations with siRNA KD of *Cmpk1* in adipocytes, which we would expect to result in reduction of nucleotides that activate these receptors.

DRG2

DRG2 is a GTP-binding protein known to play a role in the regulation of cell growth and differentiation[349,350]. Drg2 has also identified to play a role in adipocyte differentiation, with overexpression of Drg2 in 3T3-L1 adipocytes resulting in increased lipid deposition[351]. Further, mice overexpressing Drg2 have increased weight and adiposity on high fat diet compared to controls, as well as impaired glucose tolerance and insulin resistance[351]. Drg2 depletion has been identified to result in defects in transferrin receptor recycling, with a reduction of transferrin receptor at the cell surface, which is reflective of the observations for GLUT4 in this study[351].

LLGL1

LLGL1 is a cytoskeletal protein that has been identified to interact and activate Rab10 and play a role in membrane protein trafficking[352,353]. Rab10 is known to activate the translocation of GLUT4 secretory vesicles suggesting that it is plausible that *LLGL1* may also play a role in this process[354].

PDZK1IP1

Limited evidence of the role of PDZK1IP1 is reported in the literature. Knockdown of Pdz1ip1 resulted in an increase in GLUT4 trafficking to the cell surface, suggesting this may be a negative regulator of GLUT4. A large-scale proteomics screen identified localisation of Pdz1ip1 in urinary exosomes[354] . Also known as MAP17, PDZK1IP1 has been reported to be localised to the plasma membrane and Golgi in proximal kidney tubules, and overexpression has been reported to be associated with accelerated tumour progression[354]. PDZK1IP1 has been shown to be essential for the activity of SGLT2 (SLC5A2) in kidney, a target of key T2D medication, showing that the encoding protein plays a key role in the facilitation of transport across the plasma membrane by an independent family of glucose transporters[355,356]. The Golgi, exocytic vesicles, and plasma membrane are important sites of GLUT4 within the cell, however further investigation is needed to determine the role of *PDZK1IP1* in GLUT4-expressing tissues.

SLC36A4

SLC36A4, also known as PAT4, is a ubiquitously expressed non-proton-coupled amino acid transporter, which has been shown to play a role in the transport and/or sensing of amino acid including proline, tryptophan and alanine[355,356]. It has been shown that Slc36a4 is involved in the control of the amino acid pool in the lysosomes and regulates mTORC1 signalling[357]. mTORC1 is a negative regulator of insulin signalling, playing a role in blocking AKT activation, as well as other proteins upstream such as Grb10[357]. Slc36a4 has further been identified to be localise to the Golgi where it can also activate mTOCRC1[358,359].

TOM1L2

TOM1L2 encodes TOM1-like 2, a protein that has been suggested to be involved in membrane trafficking, therefore a role in GLUT4 trafficking is plausible. TOM1L2 has been shown to be involved in endosomal recruitment and sorting, through interaction with Tollip [360]. Further TOM1L2 has been shown to interact with clathrin and ubiquitin, important in protein trafficking[361]. TOM1L2 has been shown to be involved in the regulation of growth factor induced mitogenesis, which is of note since both mitogenesis and glucose uptake via GLUT4 are both downstream of insulin receptor activation[362,363].

TVP23B

Little has been reported for the role of *TVP23B*, encoding Trans-Golgi Network Vesicle Protein 23 Homolog B. It has been reported to be expressed in the retina with a decrease in expression of *TVP23B* (also known as *FAM18B*) in individuals with diabetic retinopathy, compared to individuals with diabetes without retinopathy[364]. TVP23B has further been identified to be involved in late golgi transport and in the trans golgi network in yeast, plants and mammalian cells[365–367]. This is an important site of intracellular GLUT4 transport.

3.4.8 Evaluation of gene selection strategy

Of the 36 genes taken forward, we identified 14 genes located at 8 loci to be significantly implicated in GLUT4 trafficking through *in vitro* assessment (**Table 3.4**).

Whilst 3 of these 8 loci from which candidate genes involved in GLUT4 trafficking were identified are associated with IFC at genome-wide significance, the remaining 5 loci are those that were prioritised below genome-wide significance, through their association with additional relevant traits (see methods). This suggests the potential utility of prioritising genes from loci that do not yet meet the stringent genome-wide significant threshold for *in vitro* follow up. 6 of the loci had a single gene implicated in GLUT4 trafficking identified at this locus, suggesting these are potential candidates for further examination as candidate causal genes at these loci. The remaining 2 loci, *FOXD2* and *SHMT1,* have 2 (*Cmpk1, Pdzk1ip1*) and 6 genes (*Alkbh5, Drg2, Flii, Llgl1, Toml12, Tvp23b*) implicated in GLUT4 trafficking, respectively.

We further were inclusive in the physical distance at which a locus window was defined for gene selection, ±500kb of the lead variant. The start position of the 14 genes where siRNA knockdown significantly impacted GLUT4 was on average 59.5 kb away from the lead variant at a given locus with distance ranging from 0.79 kb (SLC2A4; rs117643180) to 494.62 kb away (IRS1; rs2972144). This suggests that there is value in including a wide window of gene selection in loci where protein coding genes are sparsely distributed, such as IRS1. However, with the average distance of significant genes from the lead variant of the locus they were selected from being much closer, this suggested that a more restrictive window could be used to prioritise genes in future *in vitro* follow up.

Table 3.4: Mapping of genes where knockdown significantly impacted GLUT4 trafficking/regulation to insulin fold change associated loci

Locus category IFC	IFC Locus	rsid of lead SNP at locus	Lead SNP b37 CHR	Lead SNP b37 POS	Ensembl Gene Name Human	ENSG ID	Relative lead SNP distance to gene start (KB)	Relative lead SNP distance to gene end (KB)
integrated	*SHMT1*	rs12452987	17	18272449	*ALKBH5*	ENSG00000091542	186.057	159.181
integrated	*FOXD2*	rs116025863	1	47930641	*CMPK1*	ENSG00000162368	131.172	86.13
integrated	*SHMT1*	rs12452987	17	18272449	*DRG2*	ENSG00000108591	281.249	261.164
integrated	*SHMT1*	rs12452987	17	18272449	*FLII*	ENSG00000177731	124.299	110.219
integrated	*IRS1*	rs2972144	2	227101411	*IRS1*	ENSG00000169047	-494.622	-563.064
integrated	*SHMT1*	rs12452987	17	18272449	*LLGL1*	ENSG00000131899	143.548	124.26
integrated	*ERAP2*	rs1216570	5	96261717	*LNPEP*	ENSG00000113441	-9.381	-111.502
integrated	*FOXD2*	rs116025863	1	47930641	*PDZK1IP1*	ENSG00000162366	281.376	273.925
Genome-wide	*SLC2A4*	rs117643180	17	7185779	*SLC2A4*	ENSG00000181856	0.793	-5.797
Genome-wide	*MTNR1B*	rs1387153	11	92940662	*SLC36A4*	ENSG00000180773	63.321	9.532
Genome-wide	*PPP1R3B*	rs7012814	8	9173258	*TNKS*	ENSG00000173273	-240.166	-466.598
integrated	*SHMT1*	rs12452987	17	18272449	*TOM1L2*	ENSG00000175662	525.621	396.713
integrated	*SHMT1*	rs12452987	17	18272449	*TVP23B*	ENSG00000171928	-411.859	-437.578
Genome-wide	*C2CD4A*	rs7167878	15	62396189	*VPS13C*	ENSG00000129003	251.601	43.517

3.5 Discussion

Using technology that incorporated two complementary methods of detecting GLUT4 at the cell surface (see **Methods**), as well as direct measures of radiolabelled glucose analogue uptake, I examined the impact of reducing the expression of 36 genes identified at genetic loci associated with IFC. This revealed 9 novel regulators of insulin-stimulated glucose uptake and GLUT4 translocation at 4 loci. Although muscle is the major site of insulin-stimulated glucose uptake *in vivo,* the molecular mechanisms of insulin-stimulated glucose uptake are highly shared with adipocytes.[271] To date, no skeletal muscle-based model to assess GLUT4 trafficking with the responsiveness to insulin comparable to that of the mouse 3T3-L1 adipocyte has been developed, thereby limiting scope for screening in a muscle cell line.

Of the genes implicated in insulin-stimulated glucose uptake, a subset appeared to be important for the maintenance of the total amount of GLUT4 in the cell. For example, the knockdown of *Alkbh5* in 3T3-L1 adipocytes lowered both total and cell surface endogenous GLUT4 abundance. *ALKBH5* encodes AlkB Homolog 5, an m6A RNA demethylase, suggesting this gene affects GLUT4 through longer-term control of *SLC2A4* mRNA translation. This is consistent with observations for siRNA knockdown of ALKBH5 in a breast cancer cell line, and demonstrates mechanism in a more relevant to metabolically-relevant cell type.[341] Additional processes involved in the maintenance of the cellular GLUT4 pool, including transcription, stable intracellular localisation, or protein degradation may be involved in the case of other genes.

Our studies implicated other genes likely more directly in the process of GLUT4 trafficking, including translocation to the cell surface, since they did not alter total cellular GLUT4. Examples here include *Llgl1*, known to interact with the GLUT4 trafficking regulator Rab10, and established regulators of membrane traffic via the Golgi (*Drg2, Tvp23b*) and endosomes (*Pdzk1ip1, Tom1l2*), both key sites of GLUT4 traffic.

Evaluation of gene selection strategy

The screening strategy was inclusive in terms of locus with genetic distance around a lead variant for a given locus for gene selection, as well as utilisation of a permissive approach in terms of loci included in screening -with the loci below genome-wide significance included for gene selection. This shows that an inclusive approach is

beneficial for identifying genes involved in relevant biology if feasible provides optimal number of hits. From 5 loci which were identified at genome-wide significance, 4 in cohorts of European ancestry, and 1 from the meta-analysis of cohorts of East Asian and Hispanic American ancestry, 3 genes significantly implicated in GLUT4 trafficking were identified, 1 of which has not been identified previously.

Whilst the effects of the genes identified to be implicated in GLUT4 trafficking cannot unequivocally be linked as causal genes at the loci from which they were selected, there are some examples which may be interest to characterise at the locus level in the future. For example, at the *MTNR1B* locus, a neighbouring gene *SLC36A4* was identified, encoding a non-ionic amino acid transporter, the depletion of which resulted in a marked impairment of GLUT4 trafficking in 3T3-L1 adipocytes. There are some conflicting observations concerning the mechanisms whereby variation near/in *MTNR1B,* one of two human melatonin receptors*,* might contribute to T2D pathogenesis (summarised in[320] and discussed in[285]), and therefore it will be interesting to investigate the role of SLC36A4 dysfunction contributing to the T2D association at this locus. The T2D risk allele at the *MTN1RB* locus (rs10830963) has been associated with increased MTNR1B mRNA in human islets[368] and associated with increased fasting glucose and reduced beta-cell function.[195–199] However, in contrast, the characterisation of rare, coding, *MTNR1B* variants have identified decreased melatonin signalling due to dysfunction of MTNR1B results increased T2D risk.[200] Similarly, in human clinical studies there are contradictory results with regards to benefit of melatonin administration with acute treatment of melatonin resulting in impaired glucose tolerance in two studies of non-diabetic[105] and diabetic individuals[201]. However larger studies identified over time that baseline lower levels of melatonin secretion were associated with a higher risk of developing T2D.[199] Implication *SLC36A4*, a gene located at the *MTNR1B* locus, in the regulation of GLUT4 trafficking *in vitro* suggesting that alternative explanations for the impact of this locus on metabolic traits are possible alongside influence of MTNR1B, through influences on insulin sensitivity, which warrant further investigation.

Genes from prioritised loci below genome-wide significance ($5x10^{-8} < P < 5x10^{-5}$) in this screening strategy were included. Whilst the confidence of the association of these loci with post-challenge insulin resistance is less confident, 11 of the 14 genes screened with a significant effect in this assay were located at these loci, suggesting it

is possible that these loci are implicated in post-challenge insulin resistance. Whilst the causative link of the genes we identified at these loci cannot be determined; there is suggestive evidence at some of these loci. For the SHMT1 locus for which 6 of 14 genes implicated in GLUT4 trafficking were implicated. This suggests it is possible that there is a complex regulatory architecture at this locus resulting in the co-regulation of multiple genes that play a role in this process. Further investigation is needed to identify whether the GWAS locus identified at SHMT1 commonly regulates all these genes to identify genetic evidence for these genes in post-challenge insulin resistance.

Use of a murine adipocyte cell line to assess GLUT4 trafficking

Whilst 3T3-L1 adipocytes are not a human nor muscle cell line they provide great utility and are established as a model to assess GLUT4 trafficking. Whilst there are mammalian muscle cell lines available, these were not used in this assay as these cells express low levels of GLUT4[369], and therefore may be less sensitive to pick up genes which when knockdown result in more subtle changes in GLUT4 traffic/regulation. These cells further have been shown to not express the correct machinery necessary to reflect all components of the GLUT4 pathway and for responses to reflect the endogenous response in mature tissues in the whole organism.[369] However, it should be noted there are differences in human and mouse adipose biology[370,371], therefore further validation will be needed to ensure the roles for genes identified in 3T3-L1 adipocytes are translatable to humans when improved cell lines are available for both human muscle and adipose.

Strengths and limitations

These results provide initial evidence of the role of 9 genes in GLUT4 trafficking and/or regulation. This work as several strengths including that the findings reported are robust to replication in two independent 3T3-L1 adipocyte cell lines with directionally consistent effects observed for endogenous GLUT4 in the WT 3T3-L1 adipocytes and 3T3-L1 adipocytes stably expressing HA-GLUT4-mRuby. We were blinded to the identity of all genes, including “internal” positive controls during the experiments, prior to data analysis, to ensure experiments were as unbiased as possible. The effects for most of the genes identified in this siRNA GLUT4 screen were replicated in an independent radiolabelled glucose transport assay which was conducted independently from the screen. We did identify some inconsistent effects for some genes which could be potentially explained by the fact that this assay does not

measure glucose uptake specifically by GLUT4 and therefore can capture effects of glucose uptake independent of this such as via GLUT1.

There are some limitations to this work which need to be considered. Firstly, the results shared are using a single pool of siRNAs per gene for gene knockdown. Whilst these pools contain 4 independent siRNAs, replication with independent set of siRNAs is needed to ensure robust results. Further causal genes to these loci of interest cannot be assigned through this screen alone, and further experiments which aim to characterise the allelic effect of candidate causal variants at a locus will be needed to assign a causal gene. Further research is also needed to see if complex regulatory architecture at a locus, may implicate a group of genes which effect GLUT4. Further, whilst this assay provides evidence that these genes are implicated in GLUT4 trafficking, the exact mechanism by which these genes are involved remains to be proven and further research focused on each gene of interest will be needed to elucidate the functional relevance to GLUT4 trafficking and whether these effects are directly on GLUT4 or through a cascade of effects on other genes. Finally, as described above, the cell line used in this assay is not of human origin and therefore further research is needed to see the translatability of these results in the human context when more optimal cell lines are available sourced from disease relevant tissues.

Conclusions and future directions

This set of work provides initial evidence of 9 genes, which were prioritised for screening using genetically informed strategies, are implicated in GLUT4 trafficking. Further work will need to be done to disentangle the exact mechanisms by which these genes result in effect GLUT4 trafficking. For example, genes where knockdown effected both total and surface GLUT4 may be implicated in wider transcriptional programmes that disrupt a broader range of genes. Whereas genes that are implicated in surface GLUT4 only may be involved more directly in the control of GLUT4 trafficking directly through direct interaction with GLUT4 or through an effect on other proteins involved in GLUT4 trafficking pathway, for example in the regulation of GLUT4 secretory vesicle transport and docking to the cell surface.

To conclude, through the employment of cell-based functional screening methods, 9 putative novel regulators of GLUT4 were identified. This provides novel insights into the control of post-challenge glucose disposal in humans and highlights several

candidates for future investigation to fully elucidate their role in GLUT4 trafficking and their potential as therapeutic targets for T2D.

Chapter 4: Dissecting the genetic heterogeneity of type 2 diabetes using untargeted plasma metabolomics.

Chapter 4 overview

Background: Type 2 diabetes (T2D) is a major global health burden and patients present with a high heterogeneity in symptoms and aetiology. Part of the heterogeneity has been attributed to distinct genetic endotypes for T2D, but the fine scale aetiological characteristics of such subtypes are still unclear.

Methods: Here I combined genetic information using (partial) polygenic risk scores (PGS) for T2D with comprehensive untargeted metabolomic profiling in >10,000 individuals. The potential role of metabolites as causal mediators from genetic risk to T2D incidence was assessed using a causal mediation framework and bespoke follow-up including genetic colocalisation and Mendelian randomisation analysis.

Results: Metabolites significantly associated with T2D incidence, as identified in a systematic literature review and prospective observational associations, span almost all biological pathways, suggesting a lack of specificity and potential for confounding. Of 134 metabolites associated with a PGS for T2D, 71 were associated with incident T2D with significant and directionally consistent associations. These metabolites included well established candidates covering known pathways including glucose, lipid and amino acid metabolites, as well as less established pathways such as steroid metabolism. Genetic T2D endotypes showed distinct metabolomic signatures, representing specific metabolic pathways. The causal mediation framework identified a distinct set of established and novel candidate causal mediators. Bespoke follow up of a single example with a well-defined genetic architecture, glycine, demonstrated the utility of genetically driven mediation assessment to further aetiological understanding.

Conclusions: Here I demonstrate the utility of integrating genetic risk with large scale metabolomics profiling to allow the refinement of key aetiological pathways in T2D. Further the association of genetic T2D endotypes with distinct sets of metabolites highlights the distinct aetiological contribution to such subtypes, which could potentially be utilised as biomarkers for these. The genetically guided causal mediation strategy proposed further provides a general framework to reveal the biological mechanisms that underlie confounded observational associations. Such approach, when applied at scale, may facilitate the prioritisation of interventions targeting metabolite levels in the prevention of T2D, compared to other targets.

Collaboration and contributions:

The systematic literature review search terms and strategy used was originally used by Dr Laura Wittemans, a former group member, who included the results of this up to April 2018 in her PhD thesis. I updated this review independently, to allow an up-to-date systematic view of the literature around this topic. I independently conducted all analyses presented in this chapter.

4.1 Aims and Objectives:

The overall aim of this chapter was to identify and refine metabolomic signature of T2D through the integration of genetics:

- Assess the evidence base of associations between metabolites and T2D incidence through systematic literature review and large-scale observational association testing.
- Assess metabolomic signatures of T2D and its endophenotypes using (partial) genetic risk scores for T2D.
- Apply a causal framework including genetic and mediation analyses to prioritise potential causal pathways from the plethora of statistically significant observational associations.

4.2 Introduction

T2D is highly heterogeneous, and is contributed to by multiple pathophysiologies, with several efforts having attempted to understand the relative contribution of these to T2D, and whether there are disease subtypes, utilising both phenotypic and genetic data[132,134,137,372] (**see expanded discussion in Chapter 1**). Whilst reproducible identification of subclusters have been established, the underlying biological pathways associated with such clusters have not been systematically characterised. Detailed molecular phenotyping of plasma 'omics', including genomics and untargeted metabolomics measurement of circulating small molecules in large-scale cohorts of people followed up for T2D development – has the potential for agnostic testing and discovery of pathways underlying genetic endotypes and T2D risk.

Previous studies have highlighted a plethora of metabolites that are associated with T2D incidence, spanning almost all biological entities (see Results below), and therefore aetiological insight from these studies is limited. Further, these have not been shown to provide a clinically meaningful improvement for disease prediction. Whilst this is the case, such associations can be informative if they can highlight new causal pathways in T2D. However, this is a challenge in purely observational studies as there is strong influence of many factors that impact both the circulating metabolome and T2D. Therefore, confounding is extremely likely.

Such an overwhelming number of observational associations calls for novel strategies in biomarker selection, particularly if utilising this large-scale data to study T2D aetiology. The utilisation of genetic approaches is one way to overcome such challenges and this is what I have aimed to do in this chapter.

One such approach is the use of genetic risk scores subsuming the cumulative genetic risk of disease. Such genetic approaches have advantages over classical observational analyses as they are less prone to confounding, with genetic variants are randomly sorted during meiosis and are hence independent from major environmental confounders[373,374]. Further, genetic approaches such as genetic colocalisation and Mendelian randomisation allow the assessment of the potential

causality of such metabolites in mediating disease risk. These approaches therefore strongly increase the ability to detect candidate causal relationships.

Further, agnostic approaches for metabolite profiling that maximise the number of metabolites that can be assessed in a single study need to be employed, in large scale population cohorts. This will provide the most comprehensive view of the metabolomic landscape of T2D in a single study. This is particularly important as many of the current observational studies assessing the association of the metabolome with T2D rely on targeted metabolomics approaches and therefore are restricted to a limited number of metabolites. Larger study sample sizes will allow true assessment of the stability of identified associations.

Here, I integrated untargeted plasma metabolomic measures at baseline and information on T2D genetic risk, in a large study of individuals with extensive follow up data on T2D onset over more than 25 years. The utilisation of T2D genetic risk scores allowed me to refine metabolites associated with increased T2D risk, as well as identify metabolomic signatures for specific T2D endotypes. I finally assessed the potential causal role of these metabolites in T2D using a causal mediation framework and bespoke genetic follow up approaches. This genetically guided strategy provides a general framework to integrate genetics and other ‘omics layers to reveal the biological mechanisms and infer causality in disease.

4.3 Methods

4.3.1 Systematic literature review

A systematic literature search on PubMed was conducted up to 12th October 2022 to generate an overview of all metabolites reported for incident T2D. We searched for studies assessing the link between the serum or plasma metabolome and T2D, using a combination of three search term strings: for T2D, epidemiological studies, and metabolomics – focusing on two most used measurement techniques: proton nuclear magnetic resonance (^{1}H-NMR) and mass spectrometry (MS) (**Supplementary Table S4.1**). Results were filtered to include only articles with free full text availability, in English and listed as studies in humans prior to title screening. The phases of systematic review are outlined in **Figure 4.1**. Of the 2,754 titles initially found, 98 studies were retrieved: 45 cross-sectional and 53 longitudinal.

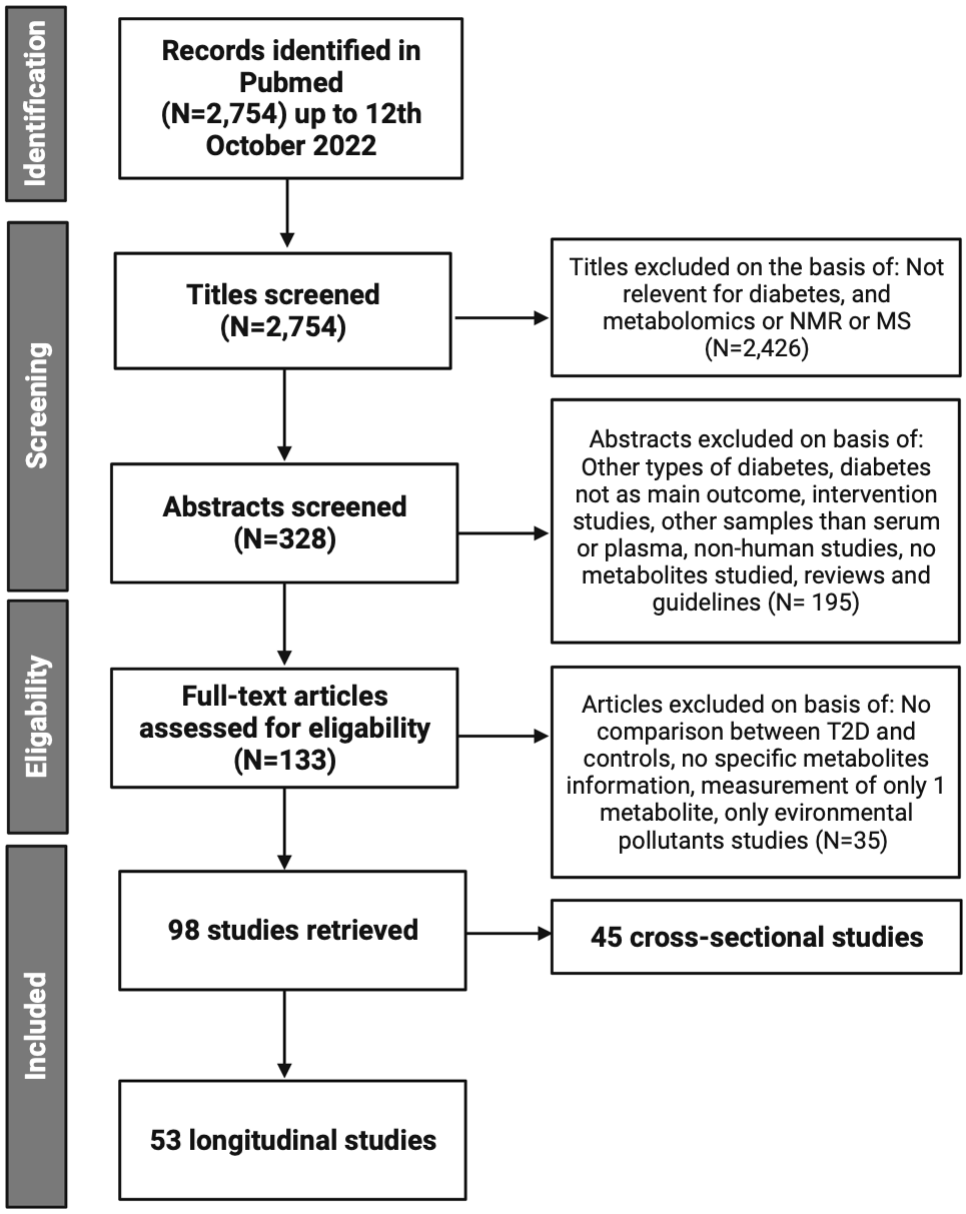


Figure 4.1: PRISMA diagram of systematic literature review on the metabolomic signature of type 2 diabetes.
NMR (Nuclear Magnetic Resonance): nuclear magnetic resonance, MS: mass spectrometry, T2D: type 2 diabetes. Created with BioRender.com.

4.3.2 The EPIC-Norfolk Study

The European Prospective Investigation of Cancer (EPIC)-Norfolk study is a longitudinal population-based study of 25,639 participants living in Norfolk and the surrounding area. Participants were aged 40-79 years old at baseline and were

recruited from 35 general practitioner practices between 1993 and 1997. Participants were invited to attend an initial health check, and subsequently invited to 4 further follow-up health checks between 1998 and 2018. The study was approved by the Norfolk Research Ethics Committee (ref. 98CN01) and all participants gave signed informed consent. The study presented in this chapter includes two subsets of EPIC-Norfolk participants, described in further detail below.

T2D case-cohort

The EPIC-Norfolk T2D case-cohort study (N = 1,501) consists of all incident T2D cases up to 31st December 2007 (672 participants, 42% female) and a selected sub-cohort of controls (829 participants, 59% female) from the full EPIC-Norfolk cohort who had a stored blood sample[375]. Incident T2D cases were defined using several sources of evidence including: self-report, hospital admissions, drug register data, mortality, and linkage to primary and secondary care.[376]

Quasi-random sub-cohort

The quasi-random (QR) sub-cohort consists of 11,996 individuals with metabolomics measurements, selected randomly from EPIC-Norfolk. All genetic analysis was done in 10,794 participants where both genotype and metabolomics data were available. Metabolomics profiling was completed in two measurement batches of 5094 (53% female) and 5698 (53% female) individuals.

Although there is an inherent overlap of incident T2D cases between these two sub-cohorts in EPIC-Norfolk due to study design, the cases in the T2D case-cohort in this study were separate from those in the QR sub-cohort.

Metabolomics profiling and processing

Untargeted metabolite profiling of baseline non-fasted serum was performed using the DiscoveryHD4 platform which has been previously described (Metabolon, Inc., Durham, USA)[377].

Metabolite measures were normalised by taking the natural log of each measurement, winsorizing to 5 standard deviations (SD) and scaled to mean of 0 and SD of 1. Metabolite annotation and pathway classifications are as reported by Metabolon, Inc. These metabolites are from a range of pathways including amino acid, carbohydrate,

lipid, and nucleotide metabolism along with metabolites that are of unknown identity. A total of 1,504 metabolites were detected using this platform in EPIC-Norfolk, with 1,379 remaining after data processing described above across all batches, 771 metabolites were present in all three measurement batches.

Genotyping

Individuals in EPIC-Norfolk were genotyped using Affymetrics UK Biobank Axiom Array. SNPs (single nucleotide polymorphisms) which were not directly genotyped were imputed using the Haplotype Reference Consortium (HRC).[378]

4.2.3 Polygenic Score Calculation

T2D polygenic risk score

A weighted polygenic risk score for T2D (T2D PGS) was calculated using 236 loci that were identified to be distinct primary association signals at genome-wide significance ($p< 5x10^{-8}$) by Mahajan *et al* (2018) [72]. The highly pleiotropic GCKR locus was excluded, which had strong effects on multiple metabolites, to minimise the influence of reverse causality and horizontal pleiotropy[165]. However, the exclusion of the *GCKR* variant from the T2D PGS had a negligible effect on the overall T2D PGS associations (Pearson coefficient = 0.99). Allele scores for each SNP were weighted for their association to T2D using beta estimates reported in both BMI-adjusted and BMI-unadjusted association analyses, generating two T2D PGS, that showed highly concordant plasma metabolite associations (**Supplementary Table S4.3, Supplementary Figure S4.1**).

T2D partial PGS

Using the same method described above, previously established partial polygenic scores for T2D (pPGS) were calculated as reported by Udler *et al*[134] who classified five clusters of T2D genetic loci representing different T2D pathophysiological pathways (**Supplementary Table S4.4**). Three of these represent different types of insulin resistance and two represent beta cell dysfunction and were assigned to distinct pathologies including beta cell dysfunction and obesity. The *GCKR* locus was excluded from the pPGS, in line with the overall T2D PGS. pPGS were scaled to mean 0 and SD of 1 to allow direct comparison of effects across scores.

An additional 93 SNP T2D genetic score for T2D, including all SNPs from which these pPGS were derived was also constructed, with allele scores weighted by the BMI unadjusted beta estimates for T2D reported by Morris *et al* (2012) (**Supplementary Table S4.5**).[146] The *GCKR* locus was also excluded from this PGS, as above. All SNPs within the 93 SNP T2D PGS are included in across the pPGS.

4.2.4 Statistical Analyses

Incident diabetes-metabolite associations

Analyses to assess the association of plasma metabolite levels with T2D incidence were performed in the EPIC-Norfolk T2D case-cohort. These analyses included 819 metabolites which were measured in at least 0.5% of all included participants, including at least 100 T2D cases. This stringent metabolite selection removes spurious associations of metabolites, such as drug metabolites which may only be present in a small number of cases.

Prentice weighted Cox proportional hazards models adjusted for age and sex were used to test for association between T2D incidence and plasma metabolite levels. This model was further adjusted for known confounders: body mass index (BMI), waist-to-hip ratio (WHR), educational attainment, smoking status, alcohol consumption and physical activity. Significant associations were considered at a false discovery rate (FDR) < 0.05 (N tests = 819).

T2D PGS association analysis

Final measurement batches taken forward for all genetic analyses contained 5094 and 5698 participants from the EPIC-Norfolk QR sub-cohort, respectively. Metabolites detected in less than 50 individuals in either batch were removed from the analysis. A total of 992 were taken forward for analysis.

Association of the T2D PGS with plasma metabolite levels were tested using linear regression models adjusted for age, sex, and population structure defined by the first four genetic principal components, separately by measurement batch, and meta-analysed using an inverse variance fixed effect model using R package metaphor (version 2.0.0). Significant associations were considered at an FDR < 0.05 (N = 992). Heterogeneity between the two batches was considered using the p-value corresponding to Cochran's Q Statistic.

All metabolites significantly associated with either the T2D PGS adjusted or unadjusted for BMI (FDR < 0.05; N=123) were tested for their association with each of the 236 individual component SNPs of the T2D PGS, plus *GCKR*. Mixed effects linear regression models adjusted for age, sex and population structure were used to test for association between these SNPs and metabolites. Significant associations were considered at an FDR adjusted p-value threshold of FDR < 0.05 (subsequently referred to as FDR < 0.05; N tests = 29,274).

T2D Partial PGS: Genetic T2D endotypes

I tested the association of pPGS with 992 plasma metabolites in the QR sub-cohort using linear regression models adjusted for age sex and population structure (1st 4 genetic principal components). These results were compared to that of a smaller 93 SNP PGS for T2D, which includes the SNPs from which these partial PGS were derived. The beta estimates of the associations with metabolites for either the 236 or 93 SNP T2D PGS were highly correlated (Pearson's r = 0.94; **Supplementary Figure S4.2**).

4.2.5 Causal mediation analyses

To determine which of the associated metabolites may mediate genetic risk to T2D onset a causal mediation framework in the EPIC-Norfolk T2D case-cohort was employed (N=1,501). This analysis considered causal mediators for T2D PGS and the 5 partial PGS to incident T2D. Analyses were restricted to metabolites which were significantly (FDR < 0.05) associated with both T2D incidence and each PGS. I used the *medflex* R package to estimate the natural indirect effect – the effect of T2D PGS on T2D through metabolites using an imputation-based approach[379]. The T2D PGS and pPGS were adjusted for population stratification. I ran these causal mediation analyses for the T2D PGS to T2D and the pPGS to T2D adjusted for age and sex only, and subsequently adjusted for BMI. Sensitivity analyses, to assess the role of insulin resistance was run adjusting for age, sex, BMI and WHR.

4.2.6 Assessing the causal role of circulating glycine in T2D

Genetic discovery of plasma glycine levels

A genome-wide association study (GWAS) for plasma glycine levels was conducted in European individuals in UK Biobank (UKBB). Glycine levels were obtained in the

subset of individuals for which ^{1}H-NMR metabolomics profiling in UKBB and quality control and standardisation was conducted using the *ukbnmr* R package[380]. GWAS was conducted using the HRC imputed genotype data using BOLT-LMM[381] adjusting for age, sex and the first 10 genetic PCs. A meta-analysis for plasma glycine levels was conducted using fixed effect meta-analysis in METAL[247]. Variant IDs were harmonised between the two studies that were included in the meta-analysis – the GWAS in UKBB (N = 108,287) described above and the largest GWAS previously reported (N = 80,003), with a total sample size of N=188,290 included in the meta-analysis [382,383]. Genome-wide significant signals were defined at $P < 5x10^{-8}$ and +/- 500kb of each lead variant was used to define independent loci.

Genetic colocalisation

I used genetic colocalisation to assess and identify shared genetic signals between glycine and T2D. For glycine, summary statistics from the meta-analysis described above were used as input. For T2D summary statistics were obtained from the largest available T2D genetic discovery, unadjusted for BMI for cohorts of European ancestry[251]. Genetic colocalisation analyses was conducted using R package *coloc* using the coloc.abf() function[177]. Loci were included in the analysis where the minimum p value was $P < 1x10^{-5}$ for each trait. Significant colocalisation was considered at a posterior probability > 0.75 for H4 – which assumes that both traits are associated and share a single causal variant. Lead variants at these loci for glycine were annotated using their nearest gene as well as their nearest protein coding gene.

Mendelian Ramdomisation

Mendelian randomisation (MR) analyses were conducted using TwoSampleMR package as has been previously described[179]. MR was conducted bidirectionally for pairs of traits where variants had been harmonised to exclude palindromic SNPs and using Steiger filtering using the *harmonise_data()* and *steiger_filtering()* functions, respectively. The following summary statistics were used as input, with genome-wide significant lead variants for the exposure used as input prior to downstream filtering: T2D: Vujkovic *et al* (2021) European ancestry cohorts only[251], Fasting Insulin: lead GWAS variants identified in European ancestry cohorts by Chen *et al* (2021)[87] were weighted by unadjusted BMI weights reported by Lagou *et al* (2021)[214], WHR adjBMI: Pullit *et al* (2019)[256], Glycine: GWAS meta-analysis described above. MR estimates were calculated using the *mr()* function, with MR Egger Intercept also calculated using

mr_pleiotropy_test(). Significant causal associations were considered at P < 0.05 for the IVW estimate.

Statistical analyses:

All statistical analyses were conducted using R (versions 3.6.3, 4.0.2) for Windows and Linux, and Stata 14.

4.3 Results

4.3.1 A systematic literature review reveals a plethora of metabolites associated with T2D risk

I systematically assessed the literature and identified more than 250 unique circulating metabolites that have been reported to be associated with T2D incidence up to October 2022 across 53 longitudinal studies (**Figure 4.1 and 4.2; Supplementary Table S4.1 and S4.2**). This literature review included 3 search strings for: T2D, epidemiological studies and metabolite measurement methods (**See Methods**).

Most of these studies (77.4%) relied on targeted approaches whose coverage is dominated by complex lipids and lipid load of different lipoproteins. The remaining 12 studies used non-targeted approaches allowing for a greater and more diverse coverage of the metabolome. Most consistently reported T2D-associated metabolites included, besides glycaemic measures, branched-chain amino acids, free fatty acids, or phospholipids.

However, even the most comprehensive untargeted mass spectrometry-based techniques currently only scratch the surface on the tens-of-thousands of small molecules predicted to circulate in blood[384], and constant technological developments will therefore likely lead to an ever-increasing number of T2D-associated molecules of unclear predictive and aetiological relevance. Data integration strategies that contextualize small molecule profiles in blood with the heterogeneous aetiology of T2D are urgently needed.

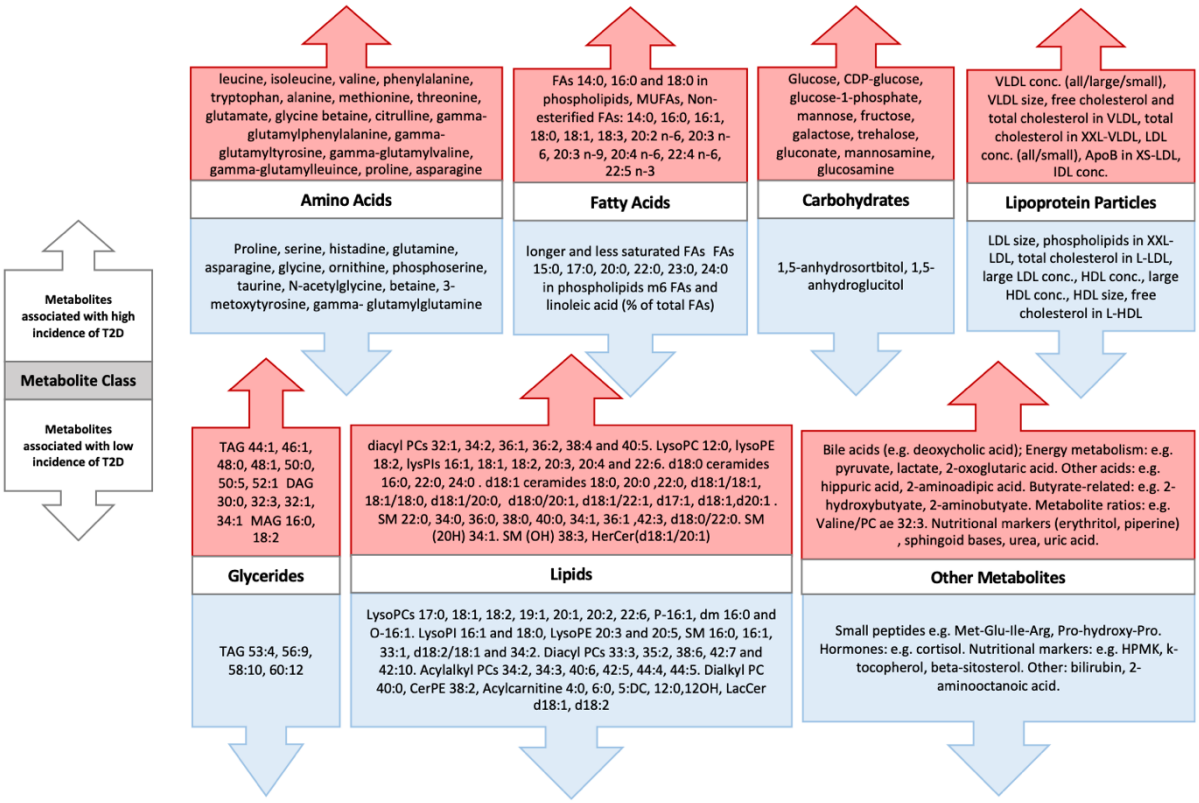


Figure 4.2: Metabolites associated with incident T2D identified by systematic literature review
Metabolites reported to be associated with incident T2D in a systematic review of the literature. Metabolites are sorted by pathway, with metabolites positively associated with incident T2D highlighted in red boxes, and those inversely associated with incident T2D shown in blue boxes.

4.3.2 More than 300 metabolites are significantly associated with incident T2D in the EPIC-Norfolk T2D case cohort

Almost half (~45%, 382) of all plasma metabolites tested (N = 819) were significantly associated (FDR <0.05) with incident T2D in the EPIC-Norfolk T2D case-cohort (N = 1501; 672 incident T2D cases, see **Methods**). Associated metabolites comprised diverse biochemical entities (**Figure 4.3; Supplementary Table S4.6**), and the top ranking positively associated metabolites included known proxies for glycaemic dysregulation such as mannose (HR = 3.66 (2.87, 4.67), P = $1.5x10^{-25}$) and glucose (HR = 2.46 (1.97,3.08), P = $2.66x10^{-15}$). Metabolites strongly inversely associated included glycine (HR = 0.49 (0.40,0.57); P = $2.99 x10^{-26}$) and related catabolites such as N-acetylglycine (HR = 0.42 (0.36,0.50), P = $3.1x10^{-24}$). More than two-thirds (69%) of the associated metabolites were likely result of confounding as adjustment for common risk factors strongly attenuated their association with incident T2D (**Supplementary Figure S4.3**).

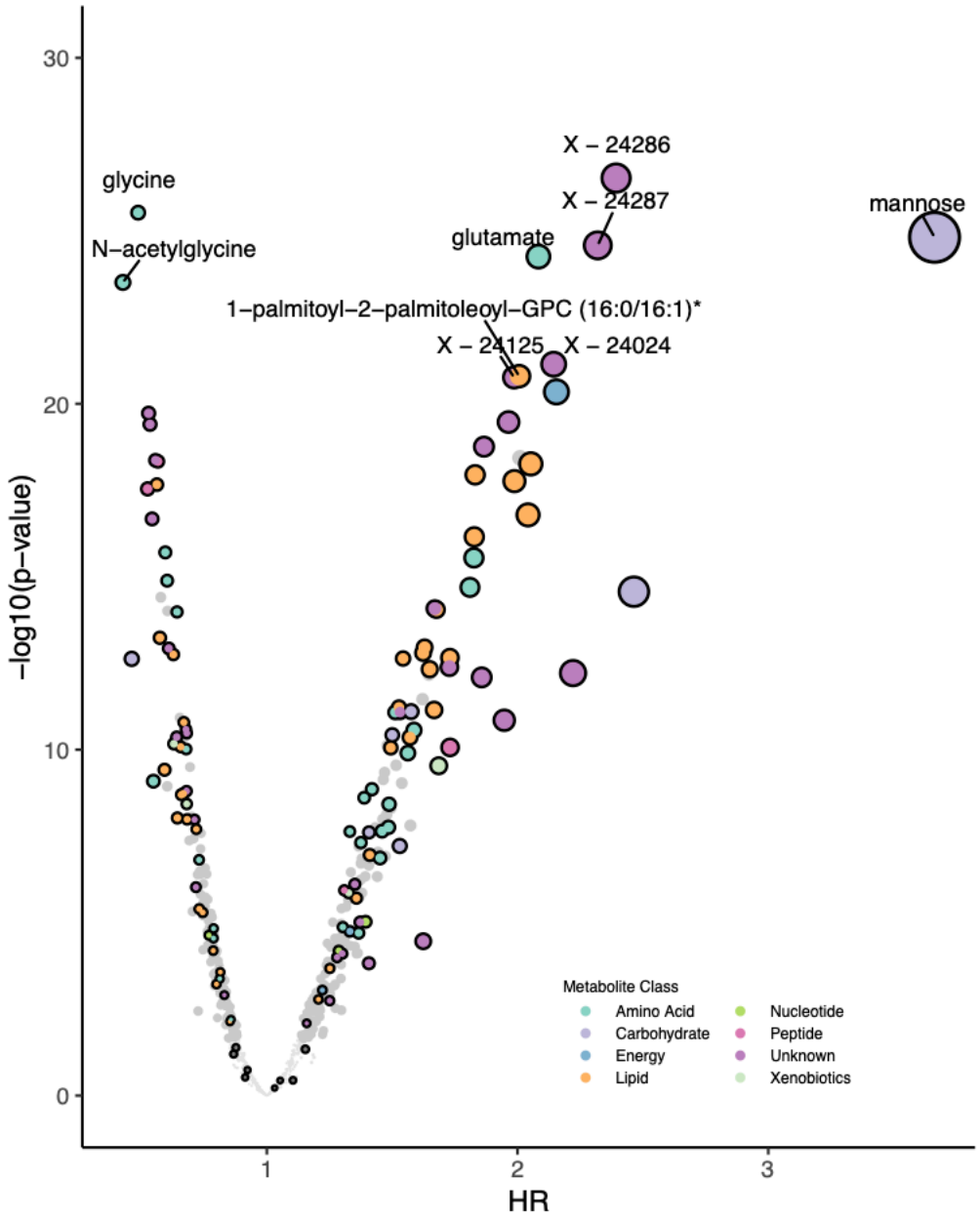


Figure 4.3: 382 metabolites are associated with incident Type 2 Diabetes in the Epic-Norfolk T2D case-cohort. Associations of metabolites with T2D incidence in the base model. Coloured points outlined in black represent the metabolite class of significant associations for metabolites significantly in the complete model, adjusted for additional T2D risk factors as well as age and sex (FDR < 0.05), dark grey represents metabolites that were significant in the base model (FDR < 0.05) but not the complete model, and non-significant association for either model (FDR >5%) are shown in light grey (**see methods, Supplementary Figure S4.3**). Point size is relative to the magnitude of effect of the metabolite on incident T2D the base model on a linear scale. Labels denote the top 10 associated metabolites.

4.3.3 Convergence of metabolite associations for genetic susceptibility to T2D and metabolic profiles associated with incident T2D

I next assessed the overlap between plasma metabolites associated with T2D risk and those associated with genetic liability towards T2D, to identify putative candidates for mediation analysis. Unless specified all genetic analyses described below were done in the larger quasi-random sub-cohort of EPIC-Norfolk (N = 10,972).

Among 134 metabolites associated with a polygenic score (PGS) for T2D (see **Methods**), 73 were significantly (FDR < 0.05) associated with both the T2D PGS and incident T2D (**Figure 4.4**). 71 were significant and directionally consistent in the longitudinal analysis, covering known pathways such as glucose, lipid, and amino acid metabolism, but also less-established pathways including steroid metabolism. In contrast, 14 metabolites associated with the T2D PGS but not with incident T2D likely representing pathways involved in sphingolipid metabolism, and amino acid metabolism (**Supplementary Table S4.6**). Effects from observational analyses and genetic analyses for each metabolite were highly correlated (Pearson's R = 0.91).

Sensitivity analyses using a T2D PGS adjusted for BMI showed little difference in associations, with effect sizes being highly correlated (Pearson's R = 0.995; **Supplementary Figure S4.1**).

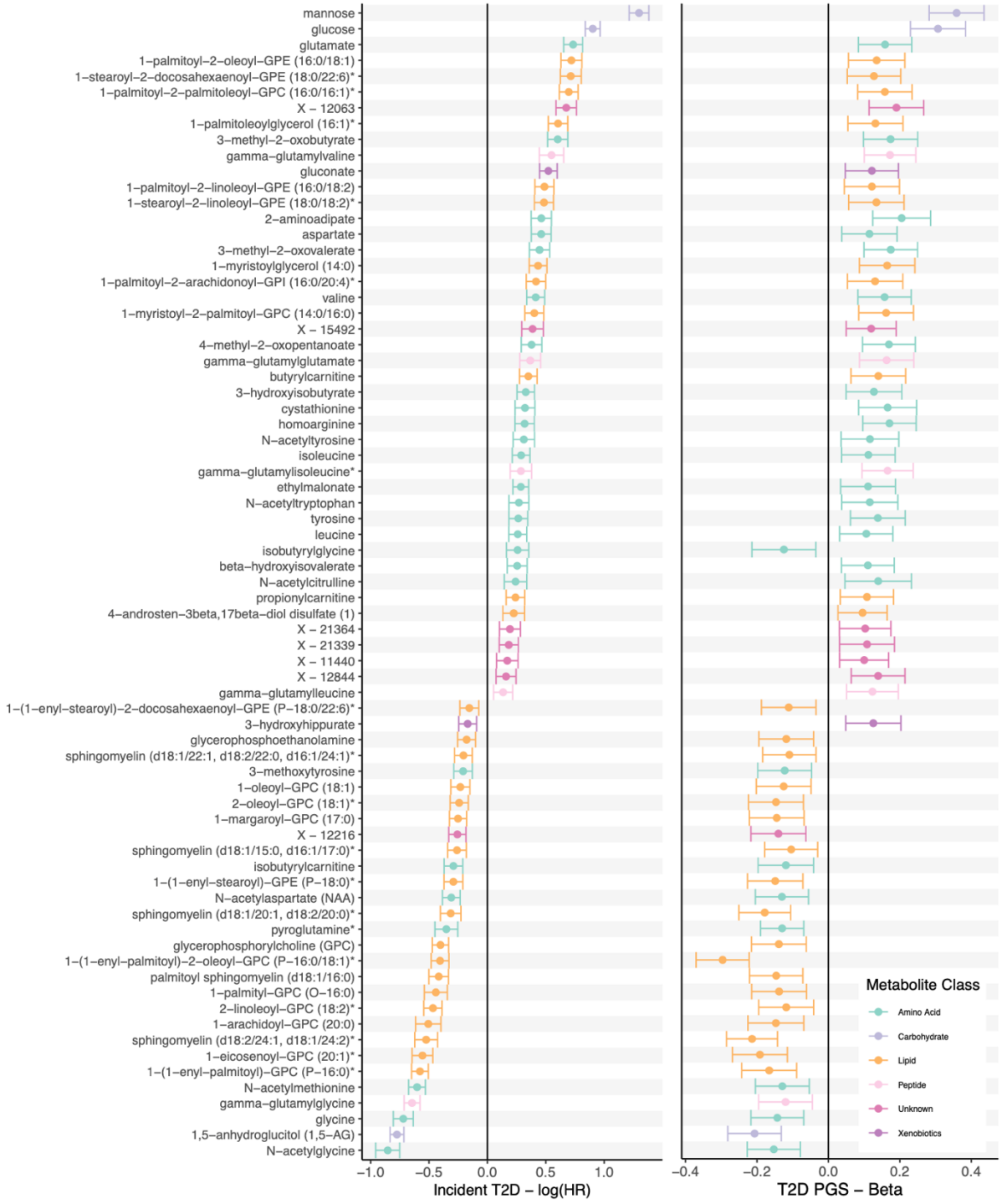


Figure 4.4: T2D PGS associations mirror those seen observationally.

Metabolites significantly associated with T2D PGS and incident T2D (FDR < 0.05). Hazard ratios of incident T2D associations (left) and Effect estimates of T2D PGS associations (right) for those that are significantly associated with BMI unadjusted T2D PGS. Colours of points indicate the metabolite class.

There was negligible impact of single variants on the overall T2D PGS results, with no individual variant appearing to drive individual metabolite associations with the T2D PGS and a total of 126 significant SNP-metabolite associations (FDR < 0.05) showing largely directionally concordant associations with the T2D PGS-metabolite

associations (**Figure 4.5 and 4.6; Supplementary Figure S4.4**). The known pleiotropic variant rs1260326 within *GCKR*[377], which was excluded from our PGS, was significantly associated with 40 metabolites (FDR < 0.05; **Figures 4.5 and 4.6.**).

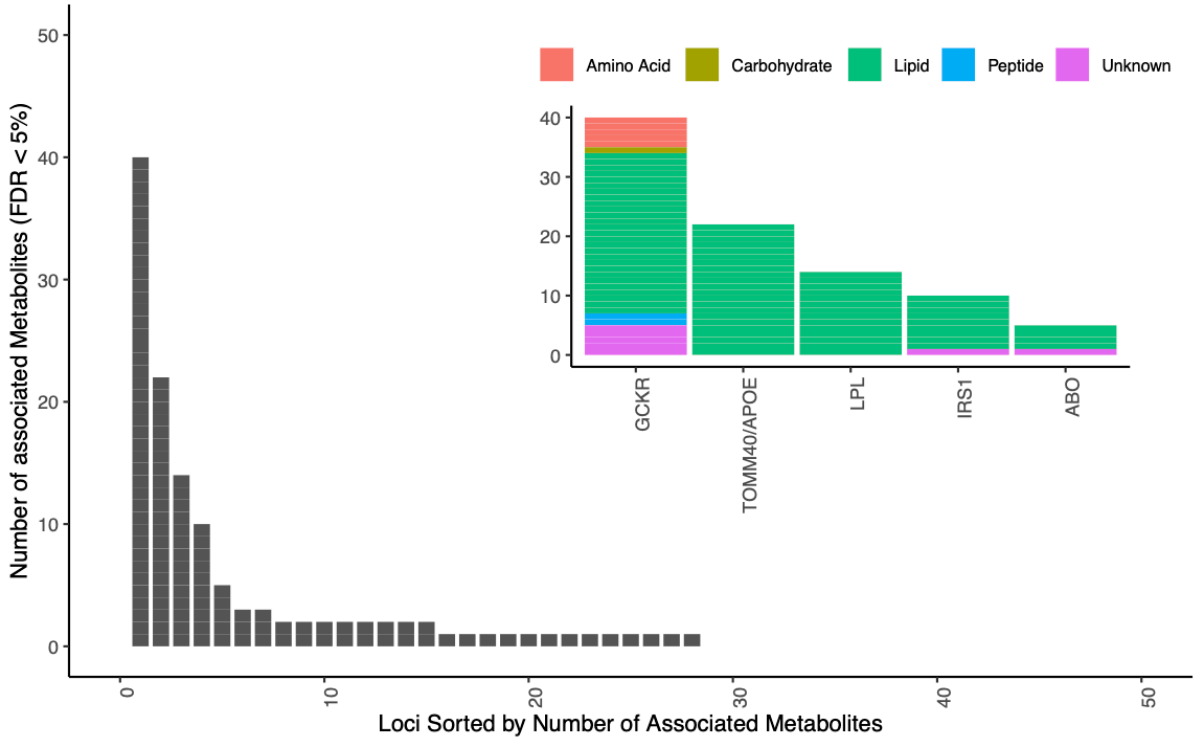


Figure 4.5: number of metabolites significantly associated (FDR < 0.05) with each SNP in the T2D PGS.

Inset figure shows the breakdown of these association by metabolite calls for the 5 SNPs with the most significant metabolite associations (FDR < 0.05). SNPs are named by their nearest gene as reported by Mahajan *et al* (2018).

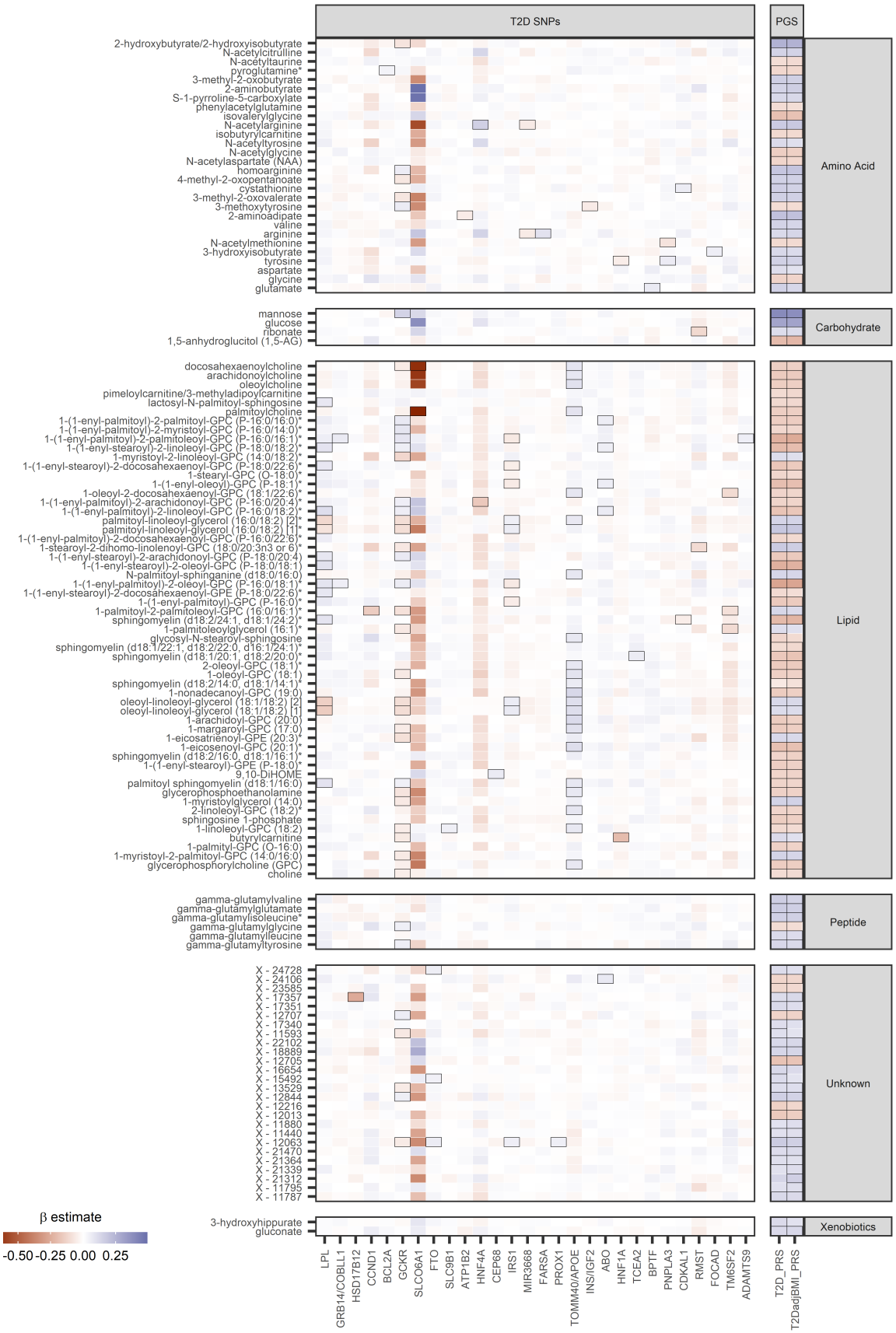


Figure 4.6: SNP-wise metabolite associations for the SNPs included in the T2D PGS, plus *GCKR*. Loci with at least 1 significant metabolite association. Significant associations are outlined in black (FDR < 0.05).

4.3.4 Genetic T2D endotypes have distinct metabolomic signatures

To understand the genetic contribution of T2D susceptibility to plasma metabolite profiles in greater detail, I utilised five aetiologically distinct pPGS, reflecting genetic T2D endotypes, and observed distinct metabolic profiles for each pPGS, and additional metabolite associations not previously seen for the overall T2D PGS. A 93-SNP overall T2D PGS was used for comparison as this contain sthe SNPs from which the pPGS were derived (**Supplementary Figure S4.2**).

A total of 100 metabolites that were significantly associated with at least one of the pPGS (FDR < 0.05) were identified, showing mutually exclusive association profiles across genetic T2D endotypes (**Figure 4.7 and Supplementary Table S4.7**). For example, the lipodystrophy-like pPGS, had most metabolite associations (n=44) of all pPGS that were also directionally concordant with the overall 93-SNP PGS, and showed a strong lipidomic signature with 39 associated lipid species (**Supplementary Table S4.7**). A further 22 metabolites were exclusively associated with the lipodystrophy pPGS including a cluster of four branched-chain amino acid (BCAA)-related metabolites, including valine.

A similar picture of shared metabolites emerged for the other genetic T2D endotypes, with a single lysolipid metabolite, 1-docosapentaenoyl-GPC (22:5n3)*, being associated with the liver pPGS and overall T2D PGS and a further 14 metabolites associated exclusively with the liver pGRS, predominantly other lipid species including polyunsaturated fatty acids, lysolipids, and phospholipids species (**Supplementary Table S4.7**). Plasma alanine specifically associated with the beta cell pPGS and overall T2D PGS, along with 6 other metabolites, including 3 glycaemic-related metabolites species (**Supplementary Table S4.7**). The obesity pPGS and overall T2D PGS specifically associated with the unknown compound X-12063 and diacylglycerol oleoyl-linoleoyl-glycerol (18:1/18:2) [2]. Thirteen further metabolites including 8 unknown metabolites and 3 metabolites involved in purine metabolism, associated with the obesity pPGS alone (**Supplementary Table S4.7**).

More than 80% (55 out of 67 covered) of the metabolites associated with any of the pPGS were also significantly associated with incident T2D.



Figure 4.7: Metabolite associations with partial genetic risk scores for T2D

Beta estimates of association are given for metabolites significantly associated with the 93 SNP T2D PGS, as well as the partial genetic risk scores of the liver-like, lipodystrophy-like, beta-cell-like and obesity-like clusters of variants. No metabolites were significantly associated with the proinsulin like score and therefore this is omitted form the plot. Error bars denote the 95% confidence interval for the significant associations. Significantly associated metabolites are coloured by metabolite class, with metabolites not associated with a given pPGS shown in light grey.

4.3.5 Mediation analysis highlights metabolites on a potentially causal pathway to T2D

I next formally tested whether plasma metabolites, prioritised by both significant (FDR < 0.05) observational and overall T2D PGS associations in the smaller EPIC-Norfolk T2D-case cohort (N metabolites = 25), represent pathways that might causally contribute to the aetiology of T2D using a mediation framework.[385] Because formal mediation analyses were performed in the smaller T2D case cohort, only the 25 metabolites that were significantly (FDR < 0.05) associated with T2D incidence and the T2D PGS were in this set considered. Of these, 18 metabolites show suggestive evidence of being causal, with 6 being robustly associated after additional adjustment for BMI (**Figure 4.8; Table 4.1).** Strongest evidence was seen for 'true positive' findings, such as carbohydrate metabolism metabolites, mannose and glucose, providing confidence in the methodological approach. Other potentially causal metabolites included glycine-related metabolites, with effect sizes comparable to well-established glycaemic metabolites.

The attenuation of other associations following adjustment for BMI does not preclude that altered metabolism of these metabolites contributes to T2D onset via obesity-related pathways, such as insulin resistance.

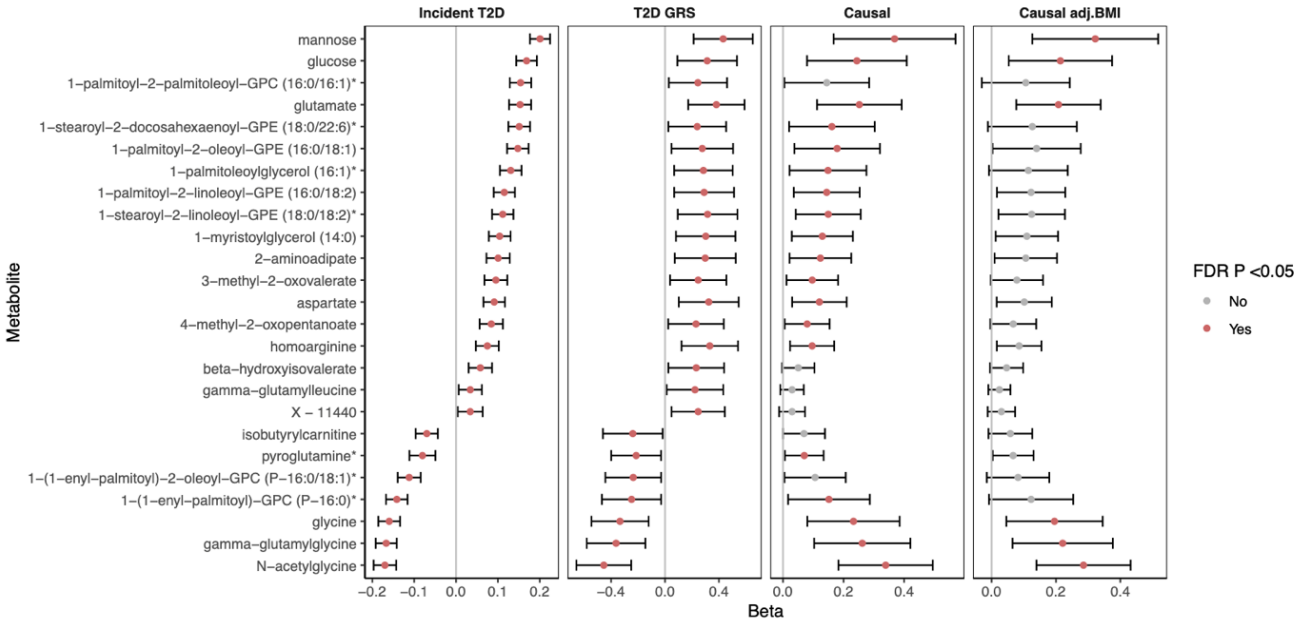


Figure 4.8: Candidate causal metabolites mediating type 2 diabetes genetic risk to incidence of Type 2 Diabetes.

Beta estimates for metabolites that are significantly associated with incidence of T2D (incident T2D) and T2D PGS in the Epic-Norfolk T2D case cohort. Causal beta estimates from mediation analyses are absolute values and therefore directionality cannot be inferred from these. Significantly associated metabolites are denoted by red points (FDR < 0.05), and those not significantly associated in each analysis are denoted by grey points (FDR > 0.05)

Table 4.1: Causal mediation results for metabolites mediating association of T2Dpgs to T2D. Natural Indirect effects.

ID	Beta	SE	pvalue	FDR	BIOCHEMICAL	SUPER_PATHWAY	SUB_PATHWAY
M48153	0.369	0.103	3.31E-04	3.39E-03	mannose	Fructose, Mannose and Galactose Metabolism	Carbohydrate
M27710	0.339	0.079	2.00E-05	5.01E-04	N-acetylglycine	Glycine, Serine and Threonine Metabolism	Amino Acid
M33949	0.262	0.081	1.25E-03	7.82E-03	gamma-glutamylglycine	Gamma-glutamyl Amino Acid	Peptide
M00057	0.252	0.071	4.07E-04	3.39E-03	glutamate	Glutamate Metabolism	Amino Acid
M48152	0.244	0.084	3.70E-03	1.54E-02	glucose	Glycolysis, Gluconeogenesis, and Pyruvate Metabolism	Carbohydrate
M00058	0.233	0.078	2.78E-03	1.39E-02	glycine	Glycine, Serine and Threonine Metabolism	Amino Acid
M19263	0.179	0.072	1.31E-02	2.73E-02	1-palmitoyl-2-oleoyl-GPE (16:0/18:1)	Phospholipid Metabolism	Lipid
M52466	0.161	0.072	2.52E-02	4.06E-02	1-stearoyl-2-docosahexaenoyl-GPE (18:0/22:6)*	Phospholipid Metabolism	Lipid
M52474	0.152	0.069	2.76E-02	4.06E-02	1-(1-enyl-palmitoyl)-GPC (P-16:0)*	Phospholipid Metabolism	Lipid
M52431	0.149	0.065	2.19E-02	3.92E-02	1-palmitoleoylglycerol (16:1)*	Monoacylglycerol	Lipid
M52446	0.149	0.055	6.35E-03	2.27E-02	1-stearoyl-2-linoleoyl-GPE (18:0/18:2)*	Phospholipid Metabolism	Lipid
M42449	0.144	0.056	9.38E-03	2.40E-02	1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	Phospholipid Metabolism	Lipid
M52470	0.144	0.071	4.30E-02	5.12E-02	1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*	Phospholipid Metabolism	Lipid
M35625	0.13	0.051	1.15E-02	2.62E-02	1-myristoylglycerol (14:0)	Monoacylglycerol	Lipid
M06146	0.123	0.052	1.77E-02	3.40E-02	2-aminoadipate	Lysine Metabolism	Amino Acid
M00443	0.12	0.046	8.93E-03	2.40E-02	aspartate	Alanine and Aspartate Metabolism	Amino Acid
M52478	0.106	0.052	4.01E-02	5.01E-02	1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*	Phospholipid Metabolism	Lipid
M15676	0.097	0.043	2.63E-02	4.06E-02	3-methyl-2-oxovalerate	Leucine, Isoleucine and Valine Metabolism	Amino Acid
M22137	0.096	0.037	9.59E-03	2.40E-02	homoarginine	Urea cycle; Arginine and Proline Metabolism	Amino Acid
M22116	0.08	0.038	3.52E-02	4.63E-02	4-methyl-2-oxopentanoate	Leucine, Isoleucine and Valine Metabolism	Amino Acid
M46225	0.07	0.033	3.12E-02	4.33E-02	pyroglutamine*	Glutamate Metabolism	Amino Acid
M33441	0.069	0.035	5.03E-02	5.72E-02	isobutyrylcarnitine	Leucine, Isoleucine and Valine Metabolism	Amino Acid
M12129	0.05	0.027	6.30E-02	6.85E-02	beta-hydroxyisovalerate	Leucine, Isoleucine and Valine Metabolism	Amino Acid
M18369	0.03	0.02	1.32E-01	1.37E-01	gamma-glutamylleucine	Gamma-glutamyl Amino Acid	Peptide
M46599	0.03	0.022	1.71E-01	1.71E-01	X - 11440	Unknown	Unknown

4.3.6 Changes in circulating glycine are likely a consequence rather than a cause of insulin resistance in T2D

To further refine and understand the possible causal link between glycine metabolism and T2D, bespoke follow-up analysis was done using genetic colocalisation and MR. Glycine was focused on specifically, as opposed to the other glycine-related metabolites discussed above, due to increased coverage across genetic studies of metabolites allowing increased power in statistical analyses. As glycine is covered both by Metabolon platform and the Nightingale platform which has been used at scale in UK Biobank, this metabolite is a good exemplar for the framework proposed, with well powered genetic discovery for circulating glycine possible. The overlap of such platforms is limited, with only 7 metabolites overlapping between these measurement platforms[382]. The other candidate causal glycine-related metabolites, which are not additionally covered by the Nightingale platform, N-acetylglycine and gamma-glutamylglycine, are only associated with 6 and 3 genetic loci, respectively in the largest available GWAS for these metabolites[165].

I updated the genetic discovery for glycine through meta-analysis of the largest publicly available GWAS summary statistics (N = 80,003)[383], with a GWAS I ran in UK biobank using NMR measure of glycine (N = 108,287). This identified a total of 54 loci at genome wide significance ($P < 5x10^{-8}$; MAF > 0.1%), a gain of 27 loci from the previous discovery effort[383]. This updated GWAS was used as input for downstream follow up of glycine using genetic colocalisation and MR.

A shared genetic signal between plasma glycine levels and T2D was identified (posterior probability >0.75) at multiple loci, including genes directly involved in insulin signalling (*INSR*, *PPARG*) as well as genes potentially involved in IR and related processes (*SCAP*, *ENPP1*, *PPP1R3B, COBLL1, C5orf67;* **Table 4.2**).

Bidirectional two-sample MR for IR-related traits, fasting insulin and waist-to-hip ratio adjusted for BMI (WHRadjBMI), identified a causal association from insulin resistance to glycine levels (e.g. fasting Insulin to glycine: beta = -0.028, SE = 0.007, P = 4.46e-5; IVW), as previously shown[383] , but not the reverse, that is, glycine on T2D or related phenotypes (e.g. glycine to fasting insulin: beta = -0.53, SE =0.43, P =0.22; IVW),

suggesting that the association of T2D with lower levels of glycine is mediated by IR (**Table 4.3; Supplementary Figures S4.5-S4.9**).

Table 4.2: Genetic colocalisation of T2D and glycine at glycine associated loci

Lead Glycine SNP (b37)	H4 posterior probability	rsid	Nearest Gene	Nearest protein coding gene
chr11_32927778_G_T	0.9980	rs145678014	*QSER1*	*QSER1*
chr19_7197931_A_T	0.7902	rs553564605	*INSR*	*INSR*
chr2_165528624_G_T	0.9960	rs1128249	*COBLL1*	*COBLL1*
chr3_12391583_A_G	0.9750	rs2197423	*PPARG*	*PPARG*
chr3_47478830_A_G	0.9780	rs11712590	*SCAP*	*SCAP*
chr5_55814890_A_G	0.9460	rs61587941	*C5orf67*	*C5orf67*
chr6_160769811_C_T	0.9730	rs668871	*SLC22A3 (OCT3)*	*SLC22A3*
chr8_9184231_A_G	0.9920	rs4240624	*LOC157273*	*PPP1R3B*
chr6_132111112_C_T	0.7772	rs3850252	*MIR548H5*	*ENPP1*

Further, when additionally adjusting for WHRadjBMI in the causal mediation analysis, as a proxy of insulin resistance in EPIC-Norfolk cohort, marked attenuation of all but one (N-acetylglycine) of the candidate causal associations was observed (**Supplementary Figure S4.10**). This further suggests that these findings were partially driven by missing adjustment of proxies for insulin resistance, mutually raising the risk for T2D and altering plasma metabolites levels but via unrelated mechanisms.

Table 4.3: Mendelian Randomisation Results for glycine, T2D and insulin resistance related traits - Fasting insulin (FI) and WHR adjusted for BMI (WHRadjBMI)

exposure	outcome	method	N SNPs	beta	se	pval
FI	Glycine	MR Egger (intercept)	53	-0.0003	0.000206	1.52E-01
FI	Glycine	Weighted mean	53	-0.0142	0.0063	2.57E-02
FI	Glycine	Inverse variance weighted	53	-0.0286	0.0070	4.46E-05
FI	Glycine	Simple mode	53	-0.0236	0.0099	2.09E-02
FI	Glycine	Weighted mode	53	-0.0131	0.0080	1.08E-01
WHRadjBMI	Glycine	MR Egger (intercept)	332	8.22E-05	5.05E-05	0.1044
WHRadjBMI	Glycine	Weighted median	332	-0.0119	0.0018	2.33E-11
WHRadjBMI	Glycine	Inverse variance weighted	332	-0.0116	0.0011	5.46E-25
WHRadjBMI	Glycine	Simple mode	332	-0.00415	0.0050	0.4020
WHRadjBMI	Glycine	Weighted mode	332	-0.00955	0.0028	6.35E-4
Glycine	Fasting insulin	MR Egger (intercept)	30	-0.0044	0.0021	0.045
Glycine	Fasting insulin	Weighted mean	30	0.2645	0.3594	4.62E-01
Glycine	Fasting insulin	Inverse variance weighted	30	0.7051	0.3718	5.79E-02
Glycine	Fasting insulin	Simple mode	30	0.6530	0.5867	2.75E-01
Glycine	Fasting insulin	Weighted mode	30	0.2572	0.3533	4.72E-01
Glycine	WHRadjBMI	MR Egger (intercept)	40	-0.0019	0.0025	0.4590
Glycine	WHRadjBMI	Weighted mean	40	0.1511	0.2341	0.5186
Glycine	WHRadjBMI	Inverse variance weighted	40	0.1127	0.3967	0.7763
Glycine	WHRadjBMI	Simple mode	40	0.3535	0.3605	0.3329
Glycine	WHRadjBMI	Weighted mode	40	0.1200	0.1637	0.4681
Glycine	T2D	MR Egger (intercept)	38	-0.0039	0.0059	5.09E-01
Glycine	T2D	Weighted median	38	-0.9034	0.6952	1.94E-01
Glycine	T2D	Inverse variance weighted	38	-1.0358	0.9332	2.67E-01
Glycine	T2D	Simple mode	38	-1.3176	1.2787	3.09E-01
Glycine	T2D	Weighted mode	38	-1.1980	0.6290	6.46E-02

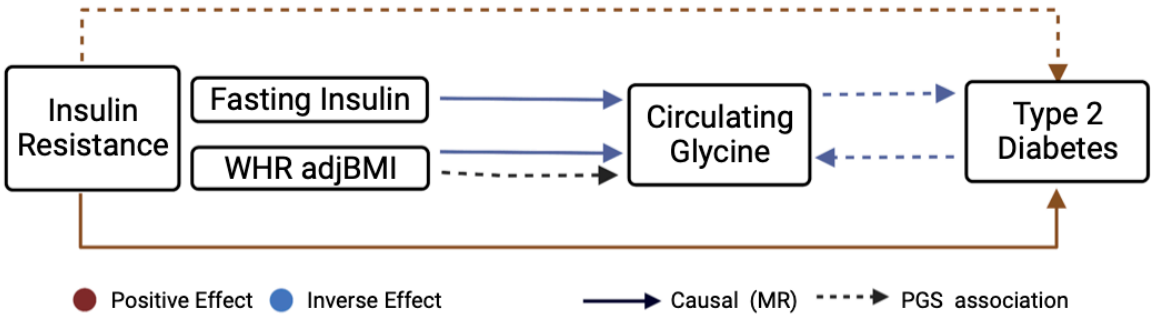


Figure 4.9: The inverse association of glycine with T2D and T2D genetic risk is a consequence of insulin resistance.

Arrows represent the direction of association from exposure to outcome with red indicated positive effects and blue indicating inverse effects of exposure on outcome. Dotted lines highlight significant associations of polygenic scores (PGS), and solid lines highlight significant causal association using Mendelian Randomisation (MR). Created with BioRender.com.

4.4 Discussion

Hundreds of metabolites have been associated with altered glycaemic control and T2D risk. Despite this, there is little evidence that these improve risk prediction in a clinically meaningful way over and above established glycaemic markers.[386–388] Associated metabolites may nevertheless help to identify novel pathways causally related to T2D, but this requires methodological approaches that overcome challenges of observational studies, specifically confounding, i.e. the influence of factors affecting the plasma metabolome and T2D risk. Genetic approaches to causal inference offer opportunities for this in an observational setting provided certain assumptions are met.[76,389] Integrating information on genetic predisposition to T2D and its endotypes, my work identifies distinct metabolomic signatures that reflect specific aspects of the genetic liability to T2D using partial genetic scores.

There is an increasing awareness that patients with T2D may fall into aetiologically distinct clusters, although there is little convergence of distinct genetic liability, as investigated in the present study, and clusters defined based on diagnosis at onset[135,390–394]. This work advances our understanding of how partial genetic liability translates into disturbed intermediate metabolism.

For example, the lipodystrophy-related PGS, likely representing a fat-distribution mediated form of insulin resistance T2D endotype[134], may mediate increased T2D risk via metabolites involved in glutathione (GSH) synthesis, like glutamate, glycine, or gamma-glutamyl glycine, which has been linked to redox signalling and oxidative stress.[395] Reduced levels of GSH have been reported in T2D patients and individuals with impaired glucose tolerance compared to individuals without diabetes with normal glucose control.[396] Further, reduced GSH levels have been reported to be associated with microvascular complications in diabetes.[396,397] Glutamate-serine-glycine (GSG) index has been identified to be positively correlated with hepatic insulin resistance and non-alcoholic fatty liver disease (NAFLD) severity, further suggesting that these metabolites may be useful biomarkers for insulin resistance, T2D as well as related complications.[398,399] While these results persisted in the mediation analysis also including BMI as a confounder, larger genetic studies on GSH levels are needed to

allow for a similar rigorous causal assessment as implemented for plasma glycine levels.

The analyses in this chapter add genetic support for the role of these pathways in T2D aetiology. Mannose was identified to be the strongest associated metabolite with T2D incidence and the T2D score and was identified to be causal candidate for the association of both the T2D PGS and beta-cell-related score. Circulating levels of mannose, a carbohydrate metabolite, are well established to be positively associated with T2D, and insulin resistance, across multiple cohorts[400–402]. Our replication of this association, as well as that of related glycaemic metabolite glucose, shows the strength of employing this approach to identify biologically meaningful and relevant associations.

Strong and replicable observational associations between plasma levels of BCAAs and other amino acids with incident T2D have stimulated interest in their pharmacological modulation, e.g., via activation of BCAA catabolism[403], to prevent T2D onset or to mitigate symptoms, like insulin resistance. Branched chain amino acids (BCAAs; isoleucine, leucine, and valine) are derived from the diet, with mammals being unable to synthesise these *de novo*. Higher levels of BCAAs have been robustly associated with T2D risk, as well as other metabolic disease such as obesity, and proposed to be a biomarker of future T2D onset.[404–409] Further associations of derivatives of BCAA intermediates have also been reported with T2D. A candidate causal role for BCAAs in development of T2D has been supported by genetic evidence through MR studies[410]. One randomised control trial has further shown that dietary restriction of BCAAs results in an improvement of insulin sensitivity in individuals with T2D.[410] Further, an additional trial where individuals with T2D were given sodium phenylbutyrate – a promoter of BCAA catabolism resulted in an improvement of peripheral sensitivity.[403] Consistently, infusion of amino acids, including BCAAs, in healthy volunteers impaired insulin action.[410]

In this study, no evidence was obtained that BCAAs lie on a direct causal pathway from genetic risk to T2D onset, although strongly associated with both. This therefore suggests BCAAs do not act as direct putative metabolic mediators for T2D, but rather

altered BCAA metabolism contributes to T2D via other pathways, for example insulin resistance. It is important to note, that this is different and does not detract from the direct impact of genetically determined BCAA metabolism on T2D, as shown in earlier studies.[410] It should be considered, for instance, that the *PPM1K* locus, which is most strongly associated with isoleucine, leucine, and valine, is only nominally associated with T2D, and was therefore not included as a genetic instrument in the T2D PGS.[410] Further associations are reported at *GCKR* which was excluded from the T2D PGS in this study.[410]

This study used a T2D PGS that included only lead variants at the primary association signals reported by Mahajan *et al* (2018)[411], rather than a genome-wide score, as this is most specific to T2D, with only variants that represent true associations with T2D included.[412,413] Inclusion of a wider range of variants, as would be done using a genome-wide PGS approach, risks introduction of confounding and T2D-unrelated signals into the analysis, decreasing the specificity of the associations with an negligible increment in predictive performance.[190,413]

The causal mediation framework employs natural indirect effect models and has utility in assessing causality where suitable genetic instruments for MR are missing. This framework does not need to rely on stringent and instable criteria for genetic instrument selection. Therefore, this enables the assessment of the candidate causal effect of all variants in a PGS in totality, and more readily reflects the polygenic nature of common disease risk. However, there are still significant limitations for use of a causal medication framework in that you need to have measured all the potential confounders in the study to allow full account for this. This limitation can be seen in the presented analysis, with insulin resistance needing to be approximated using the adjustment of WHRadjBMI, due to lack of more suitable measure to account for insulin resistance in this analysis. Further, such approach is hindered by the sample size of the study which has suitable measures. For instance, in the EPIC-Norfolk T2D case cohort the causal mediation framework was only employed to 25 metabolites that were significantly associated in this cohort, despite evidence of more T2D PGS associations in the larger quasi-random cohort which has approximately 7x larger sample size. Therefore, not all metabolites were considered.

Such approach is particularly of use when examining effects of metabolites or other 'omics intermediates where limited genetic instruments may exist for employment in an MR framework. However, with the increased availability of biobank-scale 'omics measures, it can be recommended that including complementary techniques, such as statistical colocalisation and MR, as illustrated through the example of glycine, to not only establish the candidate causal sequence, but also understand contributing pathways (**Figure 4.10**).

Using biobank-scale GWAS for glycine (N=188,290), it was demonstrated that changes in glycine metabolism, and likely other pathways, are more likely to be a consequence of insulin resistance, as seen by previous analyses.[383] Genetic colocalisation and MR have some advantages to assess a causal role, where powered genetic analyses and suitable genetic data are available, including being able to leverage genetic discovery efforts with a larger sample than single studies with individual-level data.

However, lack of suitable genetic instruments for other candidate causal metabolites, meant that this approach could not be applied these methods to test other metabolites using this approach, due to only few loci being identified for such metabolites, or these genetic instruments being highly pleiotropic. Further investigation when there is increased sample size of GWAS for these metabolites available will be needed to fully elucidate causal role of T2D-associated metabolites in disease onset.

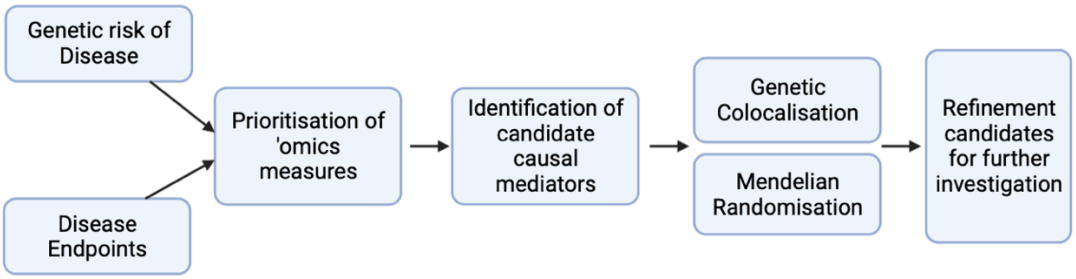


Figure 4.10: Proposed genetically driven framework to refine and assess the causal role of 'omics measures in common diseases such as type 2 diabetes

Created with BioRender.com.

Strengths and limitations

This study systematically assessed and refined the metabolomics signatures with T2D, showing that increased genetic predisposition and T2D incidence are associated with changes in plasma metabolite levels collected prior to disease onset. To our knowledge this is the first study to systematically integrate PGS, both complete and partial, disease incidence and untargeted metabolomics at scale. The causal mediation framework along with genetic colocalisation and several MR analyses demonstrated the use of comprehensive assessment of the potential causal role for metabolites in T2D, through the example of glycine. This included several sensitivity analyses to test the stability and sensitivity of the candidate causal associations identified.

However, limitations need to be considered in the context of this study. Firstly, the causal mediation analyses were restricted to metabolites significantly associated with the T2D PGS and incidence in the smaller T2D-case cohort. The overlapping number of metabolites does not subsume all metabolites for which associations were identified with for the T2D PGS in the larger quasi-random sub-cohort. Therefore, it is likely that we have missed potentially causal associations, although the overall relevance might be questionable as this would also imply much lower effect sizes. The PGS used in these analyses were restricted to lead variants at genome-wide significant loci ($P < 5x10^{-8}$). Whilst this was done to ensure specificity of the PGS to T2D and its associated genetic subtypes, as highlighted above, it is likely that these do not truly capture the complete polygenic architecture of T2D. Further the dynamic nature of metabolites over the lifetime and the fact that genetic risk is fixed at conception means that we cannot discount the fact that metabolites effects identified in this study may be under- or overestimated. Finally, it cannot be disregarded that whilst the Metabolon platform provides one of the largest coverages of untargeted metabolomics platforms available, that we possibly missed true causal mediating metabolites of T2D genetic risk to T2D incidence that are not detected using this measurement platform.

Conclusion

This large-scale and comprehensive study demonstrates, through the employment of genetic risk scores and a causal mediation framework, refinement of confounded metabolite associations with incident T2D, and further highlights genetic support of key metabolic pathways in T2D disease aetiology. Although there is little evidence of a role of these metabolites as causal mediators in T2D, these refined metabolites could be further investigated for their potential utility as clinical biomarkers over and above traditional measures in highlighting T2D endotypes which have potential utility in personalised treatment approaches in the future.

Chapter 5: Phenotypic consequences of loss of function mutations in the human Melanocortin 3 Receptor (MC3R)

Chapter 5 Overview:

The state of somatic energy stores in metazoan organisms is communicated to the brain, which regulates key aspects of behaviour, growth, nutrient partitioning, and development. The central melanocortin system is an important component of this, with the Melanocortin 4 Receptor (*MC4R*) controlling appetite, food intake and energy expenditure. Humans lacking *MC4R* are obese, hyperphagic and have reduced basal energy expenditure. Melanocortin 3 Receptor (MC3R) is the only other melanocortin receptor that is predominantly expressed in the brain. Rare functionally compromised heterozygous variants in *MC3R* have been reported in humans, however no clear consistent phenotype has been reported.

In this study I aimed to establish the role of MC3R in human physiology using a reverse genetics approach. I identified naturally occurring coding mutations in *MC3R* in 3 population-based cohorts – Avon Longitudinal Study of Parents and Children (ALSPAC), UK Biobank and Genes & Health using exome sequencing and array-based genotyping. I functionally studied these mutations *in vitro* to identify those resulting in functional impairment of the receptor and subsequently studied the relationship of these variants with human phenotypes. Collaborators both within the Institute of Metabolic Science, and in other institutions, conducted further studies which allowed us to further characterise the role of *MC3R* in human and mouse physiology.

Humans who carry characterised loss of function mutations in *MC3R*, including a single homozygote consanguineous individual, have a later onset of puberty. These individuals also had reduced linear growth, having reduced height in both childhood and adulthood. The carriers of these mutations had reduced lean mass and lower circulating IGF-1 levels, an indirect measure of growth hormone exposure. This includes a single homozygous carrier who is completely deficient in *MC3R* who has a consistent but exaggerated phenotype.

Work from collaborators, show a role of *Mc3r* in the mouse hypothalamus in population of neurons expressing Growth hormone-releasing hormone (*Ghrh*) or Kisspeptin/Tachykinin B (*Kiss1*/*Tac2*) which regulate growth and reproductive function, respectively.

Across the animal kingdom, nutritional status is a critical determinant of linear growth and the timing of reproductive maturity. *MC3R* appears to play an important role in linking signals of caloric sufficiency that act through POMC expressing neurons to the central control of growth and reproduction.

Publication related to this chapter:

The functional characterisation of MC3R variants identified in UK Biobank, Genes & Health and ALSPAC, and genetic analysis in UK Biobank formed part of joint-first author publication[414]:

Lam, B.Y.H.*, Williamson, A.*, Finer, S.*, [34 authors], Langenberg, C.#, Perry, J.R.B.#, Yeo, G.S.#, O'Rahilly, S.#, 2021. MC3R links nutritional state to childhood growth and the timing of puberty. Nature, 599(7885), pp.436-441. https://doi.org/10.1038/s41586-021-04088-9

* Joint first authors # Joint senior authors

Contributions and collaborations

The pooled sequencing of *MC3R* in ALSPAC, was completed prior to the start of my PhD by Dr Brian Lam, Professor Giles Yeo and Kara Rainbow all based at the Institute of Metabolic Science Metabolic Research Laboratories and MRC Metabolic Diseases Unit.

I performed all functional *in vitro* assays and all genetic association analyses described in this chapter unless otherwise indicated.

In Results sub-section "Supporting evidence: Additional results from collaboration" I describe results of several collaborators that built on work I performed and formed part of the publication related to this chapter[414]. These results from collaborators are highlighted in this chapter as they add further evidence for the overall conclusions drawn.

5.1 Aims and Objectives

With the role of *MC3R* to date being unclear, I aimed to explore the role of *MC3R* further using human genetics. This involved a two stranded approach combining *in vitro* functional characterisation and assessment of genetic association with variants in *MC3R* in human population studies – termed the "Functional reverse genetic approach in humans" (workflow summarised in **Figure 5.1**).

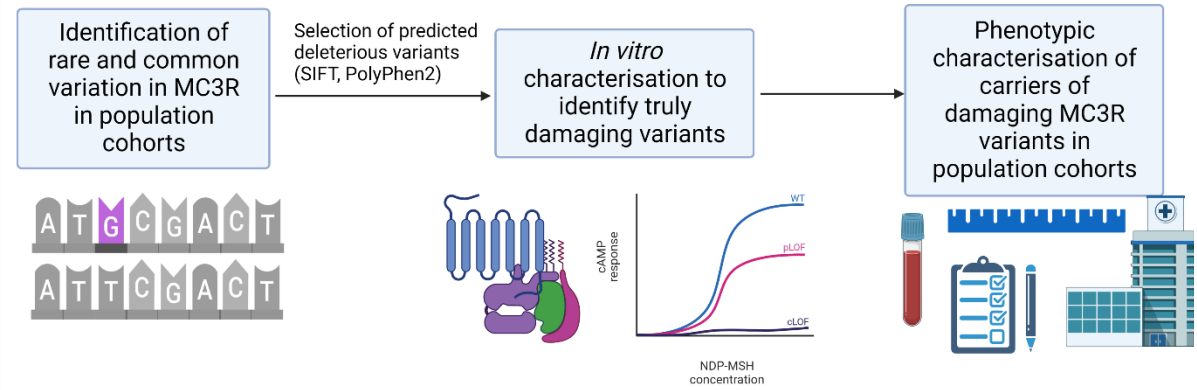


Figure 5.1: Workflow summary of functional reverse genetic approach to study role of *MC3R* in humans.

Figure created with BioRender.com

Through this systematic assessment I aimed to identify the phenotypic impact of *MC3R* loss of function. The work presented in this chapter had the following Objectives:

1. Identify coding *MC3R* variation in human population cohorts – ALSPAC, the Fenland Study, UK Biobank and Genes & Health.
    a. Prioritise variant based on their *in silico* predicted impact on function.
2. *In vitro* functional characterisation of prioritised *MC3R* variant to identify variants that truly result in *MC3R* loss of function.
3. Genetic association analyses: assess the phenotypic association of dysfunctional *MC3R* variants with human phenotype in large scale population cohorts. This analysis was done in two tiers:
    a. Systematic assessment of phenotypes previously reported in the literature in human and mouse studies.
    b. Hypothesis free unbiased assessment of a wider range of phenotypes.

5.2 Introduction

The melanocortin receptors, are a family of five G protein coupled receptors (GPCRs) – MC1R to MC5R, which are responsible for a wide range of processes. The melanocortins show tissue specific, and broad, expression patterns.[415] MC3R and MC4R are the only melanocortin receptors to be expressed centrally in the brain, while MC1R, MC2R and MC5R are predominantly expressed outside the central nervous system (CNS) in skin, adrenal cortex and skeletal muscle respectively.[415]

5.2.1. Melanocortin receptor agonists and antagonists

These melanocortin receptors are bound and activated by pro-peptide pro-opiomelanocortin (POMC) derived peptides – melanocortins – which are derived through post-translational cleavage of POMC into several active peptides (**Figure 5.2**). POMC is widely expressed throughout the body and centrally it is expressed in the arcuate nucleus of the hypothalamus and nucleus tractus solitarius of the brain stem as well as the pituitary (reviewed in [416]).

POMC is post-translationally processed by proconvertase cleavage proteins PC1 (PCSK1) and PC2 (PSCK2).[417–419] This results in the production of POMC-derived peptides including melanocyte stimulating hormones (MSH) including alpha-, beta- and gamma-MSH as well as ACTH which act as agonists of the melanocortin receptors.[420]

PC1 first cleaves the unprocessed POMC pro-peptide into an N-terminal fragment, ACTH and beta-LPH.[421] The MSH peptides are then derived though further cleavage of these products by PC2 with gamma-MSH being a cleavage product of the N-terminal fragment[422], alpha-MSH from ACTH[423] and beta-MSH, which is not found in rodents, from beta-LPH.[419] Cleavage of beta-LPH further generate gamma-LPH and beta-END.[419]

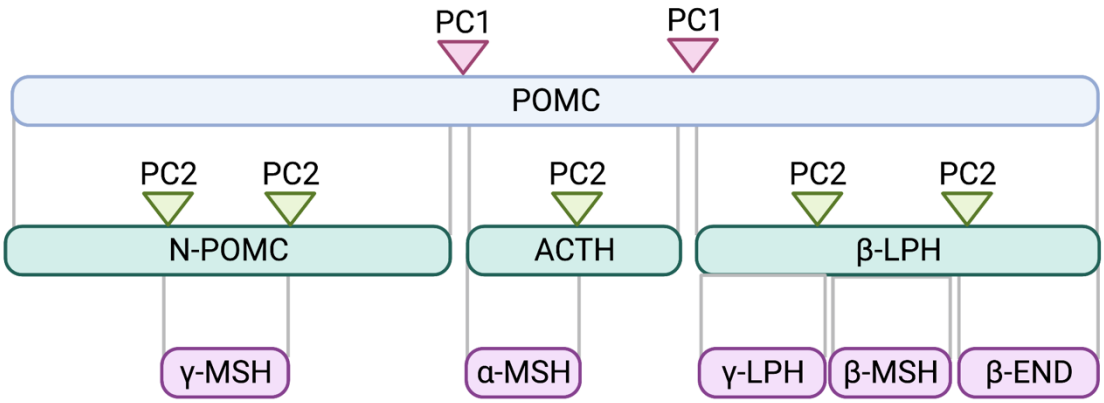


Figure 5.2: Posttranslational processing of POMC into POMC derive peptides.
Full length POMC is shown in blue. Green shows cleavage products of POMC through cleavage by PC1, whose cleavage points are indicated by pink arrows. Pink show peptides from subsequent cleavage by PC2, whose cleavage points are shown by green arrows. (Figure created using Biorender.com).

Conversely binding of antagonists Agouti related peptide (AgRP) and Agouti prevents activation of these receptors and thus downstream signalling processes.[424,425] AgRP is predominantly expressed in the arcuate nucleus of the hypothalamus and the adrenal gland, with low levels in the lung, testis and kidney[424]. Agouti is expressed more widely, being expressed in skin, adipose tissue, liver, kidney, hear, testis and ovaries.[426–428]

The 5 Melanocortin receptors show differing affinities for these POMC-derived peptides (extensively reviewed including in[429]). Of the POMC-derived agonists, ACTH is the only ligand known to activate all 5 melanocortin receptors with this being the single agonist known to activate MC2R.[430] Contrastingly, gamma-MSH selectively agonises MC3R, and with significantly lower affinity reported for other melanocortin receptors.[431] In humans, alpha-MSH and beta-MSH are able to agonise MC1R, MC3R, MC4R and MC5R with a range of potencies, with a general trend in binding affinity of: MC1R > MC3R ≥ MC4R > MC5R.[429] Antagonists agouti and AgRP also have differing affinities to the melanocortin receptors. AgRP can antagonise MC3R and MC4R at nanomolar potency, whereas it has less affinity for MC5R.[429] Agouti is able to antagonise MC1R, MC3R and MC4R at nanomolar potency whereas, similar to AgRP, this has less affinity for MC5R with micromolar concentrations required to antagonise human MC5R.[429]

Several synthetic ligands of the melanocortin receptors have been developed. These include [Nle4, DPhe7]-α-MSH (NDP-MSH) which has been characterised as a highly potent alpha-MSH analogue that has high affinity for MC1R, MC3R, MC4R and MC5R.[432,433]

5.2.2 Peripheral role of melanocortin receptors (MC1R, MC2R, MC5R).

MC1R

MC1R has been identified to be involved in the production of melanin. MC1R is therefore involved in skin pigmentation and mediates UV protection through the tanning response. MC1R, is expressed in melanocytes and leukocytes.[434] Loss of function of MC1R is associated with decreased melanin synthesis and thus fair skin as well as red hair colouration.[434–436] Roles of MC1R have also been suggested in inflammation and DNA repair mechanisms.[434]

MC2R

Also known as ACTH receptor, MC2R is the only melanocortin receptor known to be bound by a single ligand – ACTH. MC2R has been identified to play a role in functions in the adrenal cortex, where it is predominantly expressed. Here MC2R is involved adrenal development with activation of MC2R by ACTH playing a key role in the regulation of corticosterone/cortisol production.[437]

MC5R

MC5R is predominately expressed in skin and skeletal muscle, but the role of MC5R remains to be fully elucidated.[186,187] In mice MC5R was identified to play a role in sebum production with *Mc5r* knockout mice being unable to produce sebum resulting in defects in coat water repellence and thermo-regulation.[438]

5.2.3 The central melanocortin system (MC3R and MC4R)

Over the last two decades, central leptin-melanocortin system had been characterised to be responsible for the central control of food intake and energy homeostasis through the relay of signals of peripheral hormones such as leptin and insulin which are released from the periphery into the blood stream (reviews include[42,420,429,439–441]). After crossing the blood brain barrier these hormones bind to their respective receptors expressed on the surface of pro-opiomelanocortin (POMC neurons). *MC3R* and *MC4R* are the melanocortin receptors that are expressed in the brain are involved in this

pathway. While the downstream role of MC4R is well characterised, the exact role of MC3R is unclear.

The unequivocally clear role of leptin, POMC and MC4R in the central control of food intake and energy balance means much of the literature to date as focused on the role of MC3R and MC4R in these processes.[429,442]

Overview of the central melanocortin pathway

The central melanocortin pathway in the arcuate nucleus of the hypothalamus is activated through the binding of nutritional cues, including leptin and insulin, to their respective receptors on POMC neurons which result in activation of these neurons and in turn the suppression of food intake.[443] Conversely, inhibition of AgRP neurons by leptin results in the suppression of feeding.[444,445]

The activation of POMC neurons by these signals results in the increase in transcription of POMC via JAK-STAT1/3 pathway (leptin) and PI3K (insulin), therefore increasing the production of POMC-derived peptides.[446–448] These peptides subsequently act on MC3R and MC4R expressing neurons whose activation results in the regulation of downstream peripheral processes, in the case of MC4R, food intake and energy expenditure.[449]

The activation of MC3R/MC4R predominantly results in activation of downstream cyclic adenosine 3',5'-monophosphate (AMP) (cAMP) production through the induction G protein of Gαs which activates adenylyl cyclase and the conversion of ATP to cAMP. cAMP is then able to activate protein kinase A (PKA) and downstream transcriptional activation a cAMP response elements (CRE) by CREB, which is activated and phosphorylated by PKA[450–452] (**Figure 5.3a**). The EPAC/ERK pathway can also be activated downstream of cAMP, which also induces transcriptional activation at CRE.[453]

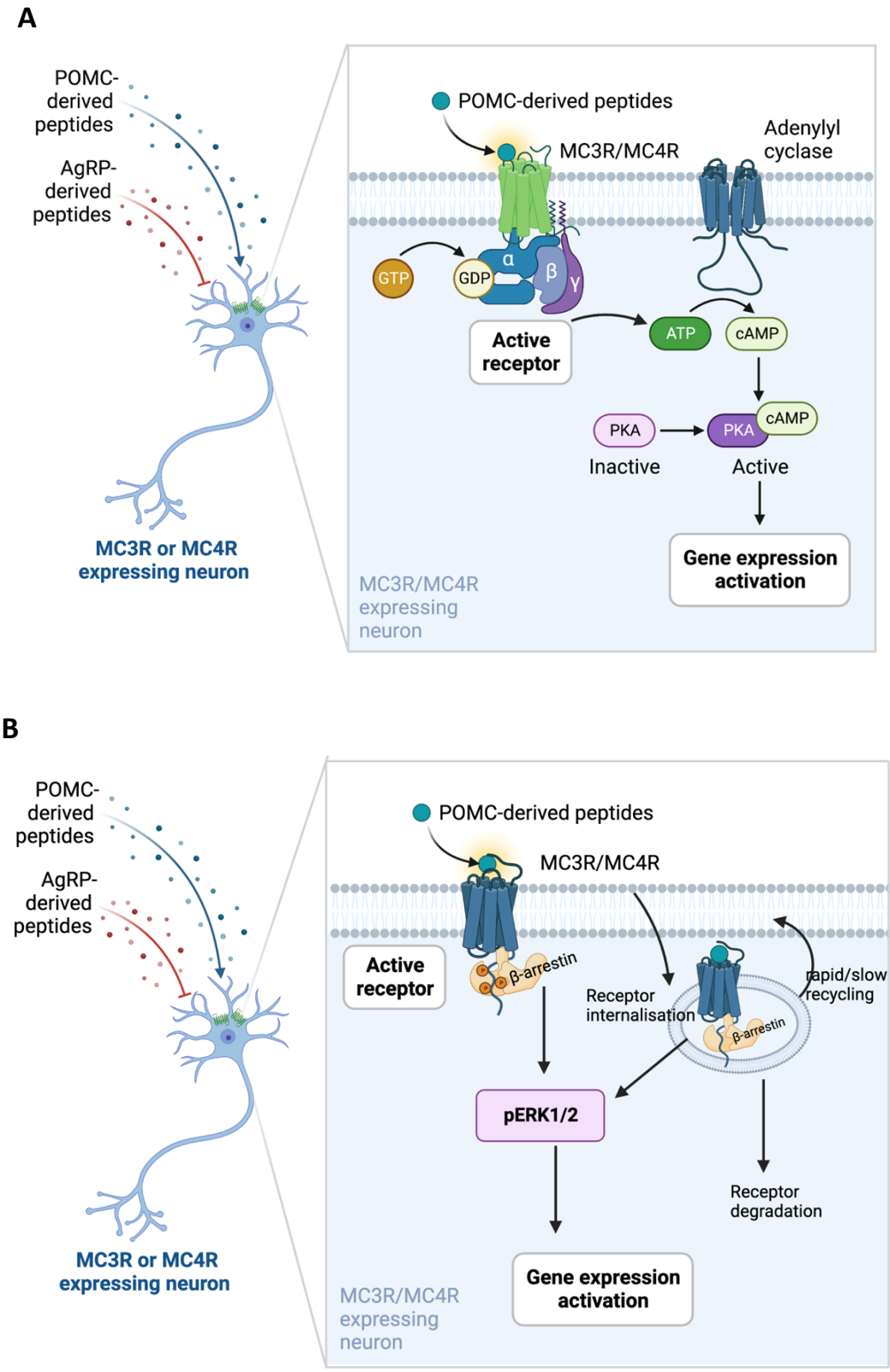


Figure 5.3: Downstream signalling of MC3R/MC4R in neurons.
(A) MC3R and MC4R are activated by POMC-derived peptides. Downstream generation of cAMP upon activation of the receptor by binding of POMC-derived peptides results in promotion of gene expression at CREB elements. (B) activation of these downstream pathways can also occur via beta-arrestin. Figure adapted from Biorender template "A Potent Melanocortin MC3 and MC4 Receptor Antagonist: SHU-9119" by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates

Beta-arrestin recruitment has also been demonstrated to be involved in the control of melanocortin receptors (**Figure 5.3b**). The interaction of beta-arrestin with MC4R has been demonstrated to result in the endocytosis of the receptor followed by either its recycling to the plasma membrane, or degradation.[454] Internalisation of these receptors can cause a second wave of activation of gene expression via the MAPK pathway which results in activation of ERK1/2.[454]

5.2.4 Genetic evidence of the role of proteins in the central melanocortin system.

Genetic studies in humans and animal models have been essential for the identification of the role of proteins within the central melanocortin pathway. In this section I will summarise the results of studies focused on key elements of the central melanocortin pathway. **Table 5.1** summarises the phenotypic changes reported with deficiency of these proteins by studies in mice and humans.

PCSK1

PCSK1, which encodes prohormone convertase 1 (PC) is responsible for the cleavage of POMC into its cleavage products (**Figure 5.1**). As well as cleaving POMC in the hypothalamus, it also plays a role in cleavage of several other hormone precursors, for example proinsulin. *PCSK1* was reported as the first monogenic obesity gene by O'Rahilly and colleagues.[455,456] The female proband was compound heterozygous for a missense (p.Gly483Arg) and splice site mutation at intron donor 5 in *PCSK1*. The proband had severe early onset obesity, as well as reactive hypoglycaemia, with impaired processing of prohormones including proinsulin and POMC identified. She further had hypogonadotropic hypogonadism and impaired adrenal function.[455,456] An additional unrelated individual also compound heterozygous for different mutations was also identified to be severely obese in childhood.[457] However the primary clinical presentation of this proband was severe diarrhoea after birth, which has later need identified in other *PCSK1* mutation carriers.[457–459] Later population-based genetic studies further identified common variation in *PCSK1* to be associated with obesity and increased BMI , providing robust replication of the role of *PCSK1* in obesity.[460]

Contrastingly, *Pcsk1* knockout mice are not obese and have slow growth, and so do not reflect the human phenotype[461]. This failure of growth is due to growth hormone deficiency due to impaired processing of pro-growth hormone releasing hormone

(proGHRH).[461] However, mouse models with coding mutations in *Pcsk1* such as p.N222D, which results in partial loss of function, were obese and glucose intolerant with evidence of defects in processing of prohormones including proinsulin and POMC, reflective of the human phenotype.[462,463]

Leptin (LEP)

Leptin is an adipokine and is well established to play a key role in the regulation of food intake and energy expenditure.[464] This evidence has come from mouse studies and also patients with severe early onset obesity in which mutations resulting in *LEP* have been identified. Levels of leptin generally increase in proportion to fat mass and act as a signal to the brain about the state of energy stores.[465]

Naturally occurring mutations in *ob* in mice – later identified to be the gene encoding leptin in mice- is naturally occurring was originally described over 70 years ago in the Obesity mouse where it was identified to have early onset obesity, reduced energy expenditure and hyperphagia.[466,467] These mice were later identified to have defects in the gene *Ob* and the human homolog mapped and subsequently identified as leptin almost 40 years later by Freidman and colleagues.[468] Following studies demonstrated that all phenotypic defects in these mice were corrected through treatment with leptin.[469–472]

Congenital leptin deficiency was first described in two severely obese children of Pakistani ancestry who had undetectable leptin levels, being severely obese from an early age. These children were identified to have a frameshift mutation resulting from deletion of guanine at codon 133 in leptin, resulting in a premature stop codon.[473] As well as early onset obesity, in line with observations in the *ob/ob* mouse, these children had marked hyperphagia.[473] Later identification of other homozygous missense mutations in two adults from a Turkish patient cohort highlighted the role of Leptin in reproductive function, with these patients never going into puberty, with diagnosis of hypogonadism.[474] This demonstrates a role of leptin in both regulation of food intake and energy expenditure as well as the hypothalamic-pituitary-gonadal axis.

Leptin Receptor (LEPR)

The Diabetes (*db/db*) mouse, identified alongside the Obese (ob/ob) mouse, has severe obesity, hyperphagia, diabetes and defects in pubertal development.[467] Later

identified to have a mutation in the gene encoding the leptin receptor, these mice additionally have leptin resistance, with extremely high levels of circulating leptin.[475,476] Leptin receptor mutations were first described in humans in 1998 with the identification of 3 siblings homozygous for a splice site mutation in *LEPR*.[477] These probands had early onset obesity and hyperphagia, as well as lack of pubertal development.[477] They additionally had shorter final height, coupled with impaired growth hormone secretion. These phenotypes were later reported and confirmed in carriers identified in a cohort of severe early onset obese patients with hyperphagia.[478] Individuals with *LEPR* deficiency have been suggested to have very high levels of circulating leptin – leptin resistance[479], however this has been contested with normal levels of leptin identified in other *LEPR* mutation carriers.[478]

Pro-opiomelanocortin (POMC)

Pro-opiomelanocortin (POMC) derived peptides are the agonists of the melanocortin system, responsible for activation of melanocortin receptors and the activation of downstream signalling pathways. Unlike *LEP* and *LEPR*, *POMC* function was first identified through the identification of *POMC*-deficient human patients.[480] In these patients it was characterised that *POMC* deficiency results in early onset obesity, red hair pigmentation and adrenal insufficiency.[480] The probands were additionally identified to be taller than non-carriers.

POMC knockout mice that were subsequently generated had severe early onset obesity, increase in body length as well as subtle yellowing of coat colour.[481] The mice additionally had lack of adrenal glands, with aldosterone and corticosterone below limit of detection[481], however different *POMC* knockout lines have been reported with adrenal glands, albeit hypofunctional.[482,483]

Melanocortin 4 Receptor (MC4R)

MC4R, is now well established to be critical for control of food intake and energy expenditure, with variants resulting in loss of *MC4R* resulting in severe early onset obesity.[449,484,485] In mice, deletion of *Mc4r* has been shown to result in severe obesity, hyperphagia, as well as insulin resistance, hyperglycaemia and increased linear growth.[486] It was further identified that even heterozygous loss of *Mc4r* in mice results in obesity, in a gene dose dependent manner.

In humans, *MC4R* deficiency was identified to result in severe early onset obesity, hyperphagia and increased linear growth.[485,487] And further unlike *LEPR, LEP* and *POMC* deficiency MC4R deficiency was not identified to impact puberty, fertility, or adrenal function. *MC4R* was the first dominantly inherited obesity gene to be reported with milder phenotype reported in heterozygous compared to homozygous carriers.[485,487]

More recently, gain of function mutations in *MC4R* , resulting in increased activation of beta-arrestin pathway, have been associated with protection of obesity[488], and heterozygous carriage of loss of function *MC4R* missense variants have been identified in 0.3% of the UK population, associated with a mean increase in body weight of 17.76kg as well as higher BMI and fat mass.[489]

Table 5.1. Summary of key phenotypic features of reported genetic deficiency in the central melanocortin system

Genetic Deficiency (gene symbol Human, Mouse)	Mouse phenotype	Human phenotype
Prohormone convertase 1 *(PCSK1, Pcsk1)*	Knockout: Obesity: no Failure to thrive, with slow growth. Loss of function mutation carrier: Obesity: yes Hyperphagia: yes Defects in prohormone processing.	Obesity: Severe early onset obesity Glycaemic control: Reactive hypoglycaemia Other endocrine: impairment of adrenal and thyroid function, Puberty onset: No, hypogonadotropic hypogonadism Digestive: frequent diarrhoea. Defects in prohormone processing including POMC and proinsulin.
Leptin *(LEP, Ob)*	Obesity: Early onset severe Hyperphagia: yes Fertile: No hypogonadism. Glycaemic control: high glucose and insulin levels Leptin levels: undetectable Linear growth: short	Obesity: Early onset severe Hyperphagia: yes Fertile: No hypogonadism. Glycaemic control: high glucose and insulin levels Leptin levels: undetectable Stature/Body length: short Puberty onset: No
Leptin Receptor *(LEPR, Db)*	Obesity: Early onset severe Hyperphagia: yes Fertile: No hypogonadism. Glycaemic control: high glucose and insulin levels, type 2 diabetes Leptin levels: high	Obesity: Early onset severe Hyperphagia: yes Fertile: No. Hypogonadotropic hypogonadism. Glycaemic control: high glucose and insulin levels Leptin levels: high/normal Stature/Body length: short Puberty onset: No
Pro-opiomelanocortin *(POMC, Pomc)*	Obesity: Early onset severe Hyperphagia: yes Adrenal gland: deficiency Leptin levels: Normal Corticosterone: undetectable	Obesity: Early onset severe Hyperphagia: yes Adrenal gland: deficiency, low ACTH Leptin levels: Normal Stature: Tall Fair Skin Red Hair
Melanocortin 4 Receptor *(MC4R, Mc4r)*	Obesity: Early onset severe Hyperphagia: yes Glycaemic control: insulin resistance Linear growth: increased	Obesity: Early onset severe Hyperphagia: yes Leptin levels: Normal Adrenal function: normal Fertile Puberty onset: yes Stature: Tall

5.2.5 Melanocortin 3 Receptor (MC3R)

The role of MC3R, the other melanocortin receptor expressed centrally, is less clear. No human homozygous phenotype has been reported prior to this project (see results below). In addition, mice lacking *Mc3r* have a complex phenotype.

Mc3r deficient mice

In mouse studies, *Mc3r* knockout mice have been reported to have reduced lean-mass and increased fat-mass compared to control animals, with no overall change in body weight.[490] Further phenotypes reported in *Mc3r* knockout mice include reduced body length, with shorter femur length, and reduced energy expenditure.[491,492] Contrastingly to *Mc4r* deficient mice, *Mc3r* deficient male mice are hypophagic under standard conditions[490], however *Mc3r* has been suggested to act as a rheostat with *Mc3r* deficient mice having defects in energy set point maintenance following nutritional challenge.[493] *Mc3r* knockout mice have been reported to have increased obesity compared to wildtype littermates on high fat diet.[490] Further, double *Mc3r* and *Mc4r* knockout mice had higher body weight on higher body weight than the single gene knockouts suggesting these genes have distinct roles.[490] However other studies have found no impact of *Mc3r* deficiency in animal models on food intake, suggesting it is solely mediated through *Mc4r.*[494]

MC3R variation in Humans

There is limited phenotypic information of *MC3R* variants in humans, with complete deficiency of *MC3R* through homozygous mutations not previously being reported. Rare, functionally compromised heterozygous variants in *MC3R* have been identified in humans but no consistent phenotype has been reported.[495] Several small-scale family-based linkage studies have reported possible association of *MC3R* with obesity.[431,495–498] However, these reports are not replicated in larger scale GWAS with no associations with traits related to body mass or weight having been reported at the MC3R locus.[499] In large scale GWAS studies variants near MC3R have been reported to be associated with height and age of menarche.[499,500]

5.3 Methods

5.3.1 Variant identification across the different cohorts

MC3R coding variants were extracted from 3 population cohorts of interest – Avon Longitudinal Study of Parents and Children (ALSPAC), UK Biobank, and Genes & Health. These variants were identified using different genotyping technologies across the different population cohorts.

Pooled high-throughput sequencing approach used to identify variants in ALSPAC

In brief, in the ALSPAC study variants were identified previously in the O'Rahilly group, prior to the start of my PhD, using a targeted pooled exome sequencing approach. In ALSPAC this targeted pooled exome sequencing approach, MC3R was sequenced in a random subset of 5993 individuals in pools of 50 individuals.

This section summarises the methods used by others in the group to identify these variants in ALSPAC, as outlined in Contributions & Collaborations for this chapter.

Pooled DNA (10 ng) was used for *MC3R* single-exon PCR with Q5 Hot Start High-Fidelity DNA Polymerase (NEB) and *MC3R* exon primers −331 bp upstream (5′-TGGAACAGCAAAGTTCTCCCT-3′) and +61 bp downstream (5′-CCTCACGTGGATGGAAAGTC-3′) of the protein-coding region, yielding a PCR product of 1,375 bp. The PCR product was purified using Agencourt Ampure XP beads (Beckman Coulter) and quantified with the QuantiFluor dsDNA system (Promega) using the Tecan Infinite M1000 Pro plate reader. Purified PCR product (1 ng) was used to construct the sequencing libraries with the Nextera XT Library Preparation Kit and Nextera XT Index V2 barcodes (Illumina) according to manufacturer's instruction. Ampure Xp beads were used to purify the libraries, which were then quantified using the Kapa Library Quantification Kit (Roche) and the Quantstudio 7 Flex Real Time PCR instrument (Thermo Fisher Scientific). All libraries were combined at 10 nM for paired-end sequencing at 150 bp (PE150) on the Illumina HiSeq4000 instrument at the CRUK Cambridge Institute Genomics Core.

BWA MEM (0.7.12) was used to align the sequence reads to the human GRCh38 (hg38) genome.[501] PCR de-duplication was performed using Picard 1.127 (https://broadinstitute.github.io/picard/). GATK 3.8 (https://gatk.broadinstitute.org/) was used to perform insertion and deletion (indel) realignment and base quality score

recalibration according to GATK Best Practices. The variants were called by mpileup2snp and mpileup2indel function from Varscan 2.4.2[502] with the following parameters: variant coverage of 100X or more, 'Strand Filter' = 'ON', variant allele frequency (VAF) of 0.6% or more and $P < 0.05$.

Genotype selection and measurement in UK Biobank

To identify directly genotyped variants covered on the UK Biobank Axiom array (Affymetrix) we extracted genotype counts in the coding region of *MC3R* available in UK Biobank using plink v1.9.[503] Genotyping quality was assessed using plink v1.9 and cluster plots of raw genotype intensity data. Variants which have a MAF > 0.05% were taken forward in analysis**.**

Imputed genotype data were used for 2 variants – rs3827103 and rs61735259, to maximise sample size (Info score > 0.96). Directly genotyped data was used for rs143321797 due to its low MAF (0.06%), and genotype cluster plots were manually inspected to ensure genotype reliability. Furthermore, genotype concordance for non-reference carriers was examined across WES (from the 200k release), genotyping array and imputation for rs143321797 and rs61735259, and genotyping calls were found to be highly concordant across technologies (**Table 5.2**).[504]

Table 5.2 Genotype concordance amongst whole-exome Sequencing (WES), genotyping arrays and imputation in UK Biobank

Variant	Comparator group	N overlapping WES non-reference carriers	Concordance with WES
rs143321797 (p.F45S)	WES non-reference carriers	234	1
	Genotyped non-reference carriers	233	0.9957
	Imputed non-reference carriers	186	0.7949
rs61735259 (p.R220S)	WES non-reference carriers	688	1
	Genotyped non-reference carriers	686	0.9971
	Imputed non-reference carriers	687	0.9985

Identification of MC3R variants in Genes & Health

Non-synonymous variants for *MC3R* were identified from public exome data available on the Genes & Health (G&H) website (https://genesandhealth.org), summary data downloaded in September 2019. The exome sequencing of G&H is described in [454].

Variant in silico annotation

Variant Effect Predictor (VEP v99)[505] and CADD (v1.6)[278] were used to annotate the extracted variants and assess their predicted deleteriousness.

5.3.2 In-vitro cyclic-AMP (cAMP) accumulation assay

Human Embryonic Kidney (HEK293) cells were obtained from lab stock and maintained with Dulbecco's Modified Eagle Medium High Glucose (Invitrogen, Carlsbad, CA, USA), supplemented with 10% foetal bovine serum (Invitrogen), 1% Glutamax (Invitrogen), 100 U/ml penicillin and 100 mg/ml streptomycin (Sigma-Aldrich, IL, USA). HEK293 cells were kept in 37˚C humidified air with 5% CO_2. The cell line was tested negative for mycoplasma contamination, it was not a commonly misidentified cell line and not authenticated.

Site-directed mutagenesis on wild-type Human N-FLAG-MC3R pcDNA3.1(+) was performed using Agilent QuikChange Lightning kit (Santa Clara, CA, USA) to generate all *MC3R* variants for cAMP activity measurement.

10ng of plasmid carrying *MC3R* wild-type (WT) and variants were transfected into HEK293 cells using Lipofectamine 3000 (Invitrogen) 48 hours prior to starting the cAMP assay. An increasing dose of [Nle^4, D-Phe^7]-α-melanocyte-stimulating hormone (NDP-αMSH, Bachem, Bubendorf, Switzerland) from 10^{-13} to 10^{-5}M was administered the following day for 2 hours in phosphate buffered saline (PBS, Sigma-Aldrich) before intracellular cAMP concentration measurement using a luminescence based HitHunter cAMP Assay for Small Molecules (Cat# DiscoverX 90-0075SM25, Eurofins DiscoverX, Fremont, CA, USA) and Tecan Infinite M1000 Pro microplate reader. cAMP standard curve was measured for each experiment following the standard manufacturer protocol and used to transform luminescence values to cAMP concentrations for downstream analyses.

The baseline and maximal cAMP concentrations were normalised to MC3R WT from the same experiment and a 3-point sigmoidal dose-response curve was fitted to each individual replicate to determine the E_{max} and the logEC_{50}. The average relative maximal efficacy (E_{max}) and log half maximal effective concentration (logEC_{50}) values were used for LoF determination.

The $logEC_{50}$ was not used for mutants which exhibited no response. One-way ANOVA was used to compare the E_{max} and $logEC_{50}$ of each MC3R mutation to the MC3R WT response. All calculations were performed with GraphPad Prism 7.

The LoF classifications were defined as follows:

Complete LoF (cLoF): $E_{max} \leq 25\%$ WT or $EC_{50} \geq 50x$ WT

Partial LoF (pLoF): 25% WT $< E_{max} \leq 75\%$ WT or $5x$ WT $\leq EC_{50} < 50x$ WT

WT-like: 125% WT $> E_{max} > 75\%$ WT or $0.2x$ WT $< EC_{50} < 5x$ WT

5.2.3 Phenotypic Association of MC3R in UK Biobank

UKBB Cohort Information

UK Biobank is a large and prospective study of ~500,000 participants aged 40-69 years, recruited between 2006 and 2010.[506] All analyses conducted using the UK Biobank Resource were done under application numbers 32974 and 44448.

UKBB Phenotype Measurements

We considered candidate anthropometric, puberty timing, and metabolic traits. The following specific filters were used: age at menarche was filtered to correspond to the ReproGen consortium definition[500] for analyses conducted with genotyping array and imputed data, and the full cohort (~100,000 female participants) was used for the whole-exome sequencing (WES) data. Type 2 diabetes was identified on the basis of probable diabetes[507] plus any mention of E11 in HES (main or secondary) or death (underlying or contributory cause); body composition variables (total lean mass and appendicular lean mass) were derived from prediction equations based on demographic, anthropometric and bioelectrical impedance values[508]. Waist-to-hip ratio was adjusted for BMI and the residuals from this were rank-based inverse normally transformed.

Statistical analyses

Individual variant associations with outcomes were assessed under additive genetic models. For the individual variants, associations were tested using mixed linear models implemented in BOLT-LMM[381], which allow the inclusion of related individuals. Phenome-wide analyses were performed in up to 451,301 individuals. The variant-based models were performed adjusted for age, sex (where appropriate) and the first

10 genetic principal components as provided by UKBB[504], with two outcomes additionally controlling for height where this is stated.

MC3R Genetic Risk Scores in UKBB

Genetic risk scores (GRS) which allow the assessment of the combined effect of the three variants of interest identified in UKBB were constructed using several variant weightings. GRS for each individual were constructed by taking the sum of allele scores for each variant which had been multiplied by the weight given for each variant.

Four weighting strategies were assessed in the construction of these genetic risk scores (see table below). Firstly, variants were weighted by the CADD (V1.6) Phred score[278] which indicates the predicted functional impact with higher values reflecting that these are predicted to be more damaging. There scores are on the Phred scale meaning that they reflect the rank of magnitude of predicted effect, for example a CADD score of 20 reflects a score within the top 1% and CADD score of 30 reflects a score within the top 0.1% of all scores.

The CADD weighted score was used as the primary GRS and sensitivity analysis was run with 3 additional weightings (**Table 5.3**). First variants were weighted by their independent effect on age of menarche, the phenotypic trait on which *MC3R* variants were observed to have the largest effect. These independent effects were calculated using joint conditional models, adjusting for all other variants.

Table 5.3 MC3R Genetic Risk Score Weighting

SNP	CADD	Menarche	Diff_E_{max}	Diff_LogEC50
rs3827103	19	0.04580293	7.1	0.03914
rs143321797	26.2	0.51303297	22.95	2.797
rs61735259	23	0.18607887	32.49	0.3772

Further 2 weights were used in sensitivity analyses which were derived from the *in vitro* functional data. These are Diff_E_{max} which is the difference in E_{max} compared to WT (100%), and Diff_LoFEC50 which is the difference in logEC50 compared to WT *MC3R*.

GRS were tested for association with relevant traits of interest in European unrelated individuals in UKBB. Linear regression models were run adjusting for age, sex, and the

1st 10 genetic principal components. Sex was not adjusted for in analysis of sex specific traits – age of menarche and age of voice breaking.

5.3.4 Phenome-wide association study

A phenome-wide association study (pheWAS) was conducted using publicly available genome-wide analysis (GWAS) summary statistics from five different repositories: GWAS of 633 ICD10-coded disease phenotypes from the UKBB provided by the Neale lab (http://www.nealelab.is/uk-biobank) where data was systematically coded using an algorithm-based approach to determine the most appropriate analysis[509], Open Targets Genetics[509], Open GWAS IEU[509], Global Biobank Engine[509], and Phenoscanner[510]. Summary statistics were extracted for the three coding *MC3R* variants rs3827103, rs143321797 and rs61735259.

We considered studies with > 5,000 individuals and excluded binary traits where there were less than 0.1% of cases in the cohort. We manually pruned the list of phenotypes to retain only non-redundant traits by choosing the largest available study covering all variants in cases where a phenotype available included in multiple data sets.

We used *grs.summary()* function from R package gtx (v0.0.8; http://cran.r-project.org/web/packages/gtx), which enables multi-SNP genetic risk score analysis using single SNP summary statistics, across 478 traits for which summary statistics for all three variants were available. Weights for each variant's CADD Phred-score (v1.6) were used in the analysis: rs143321797-C=26.2; rs6173525-A=23; rs3827103-A=19. We used a Bonferroni significance threshold to control for multiple testing ($P< 1.046x10^{-4}$).

Summary statistics were extracted for the three coding *MC3R* variants rs3827103, rs143321797 and rs61735259 for glycaemic traits from the MAGIC consortia – fasting insulin, fasting glucose, 2 h glucose, HbA1c.[87]

5.3.5 Genes & Health

Cohort Information

Genes & Health is an ongoing community-based population study comprising (on 31st August 2021) 48,960 British Bangladeshi individuals and British Pakistani

individuals.[511] Genes & Health operates under approval from the National Research Ethics Committee (London and South-East), and Health Research Authority (reference 14/LO/124), and Queen Mary University of London is the Sponsor. Genes & Health incorporates Stage 1 (health record access, saliva DNA collection) on all volunteers and Stage 2 (focused recall studies) procedures on selected volunteers, including recall-by-genotype. Exome sequencing has been performed on all volunteers reporting parental relatedness (n=5,236) and genotyping (Illumina GSAv3EA+MD chip) on all. Informed consent is taken at both Stage 1 and Stage 2 and allows analysis of health and genetic data and publication of results.

G&H Proband comparison to UKBB

The G&H proband was compared to males who have DEXA imaging data available in UKBB. This cohort was further stratified by self-reported ethnicity (field 21000) into European males, the majority of the cohort, (N=2,367; 2,356 with both BMI and DEXA measures) and South Asian males (N=36) to allow matched assessment of the proband with individuals of the same ethnic background. South Asian ethnicity was defined as individuals who reported to be of Indian, Bangladeshi, and Pakistani ethnicity. Total lean and total fat percentage were compared with age matched males of European ethnicity to account for age effect. These included males within a 10-year span closest to that of the proband – aged 44-54 at the second study visit when DEXA images were obtained (N=417).

Individuals with missing data were removed from the comparison. Percentage of lean and fat mass were calculated using DEXA total lean and fat mass variables and total mass as defined by the DEXA measurements. BMI at the second health check was used to allow comparison across different BMI ranges, to match the study visit when DEXA images were obtained. Appendicular lean mass was calculated using the sum of lean mass from legs and arms in kilograms, divided by BMI. Z-scores of these measures were calculated to aid cross-trait comparison within these subgroups of interest.

5.3.6 The Fenland Study

Fenland proteomic association results were a look up from in-house summary statistics and are now published as part of a larger effort.[169] The methods below outline what was done.

Cohort Information
The Fenland study is a population-based cohort of 12,435 participants born between 1950 and 1975 who underwent detailed phenotyping at the baseline visit between 2005 and 2015, which has previously described in detail.[240] The study was approved by the Cambridge Local Research Ethics Committee (ref. 04/Q0108/19) and all participants provided written informed consent. Briefly, the participants were recruited from general practice surgeries in the Cambridgeshire region in the UK. Individuals were not enrolled in the cohort if they were clinically diagnosed diabetes mellitus, a terminal illness or psychotic disorder, unable to walk unaided, or were pregnant or lactating.

Measurements
Proteomic profiling has previously been described.[512] Proteomics profiling was performed on fasted EDTA samples collected at baseline by SomaLogic Inc. (Boulder, CO, USA) using DNA aptamer-based technology. Relative protein abundances of 4,775 human protein targets were evaluated by 4,979 aptamers (SomaLogic V4).

Statistical analyses
10,708 Fenland participants had both phenotypes and genetic data after excluding ancestry outliers and related individuals. Association analyses for variants of interest was performed as described previously.[512] Briefly, within the 3 genotyping subsets, aptamer abundances were transformed to follow a normal distribution using rank-based inverse normal transformation, and were then adjusted for age, sex, sample collection site and first 10 genetic principal components. The residuals were then used as input for the genetic association analyses using an additive model with BGENIE (v1.3)[504]. The results for the three genotyping arrays were combined in a fixed-effects meta-analysis in METAL (v 2011-03-25)[247].

We first prioritised a total of 14 proteins from the insulin-like growth factor family of proteins targeted by 15 aptamers at a rigorous Bonferroni significance threshold ($p< 0.0033$). We further considered all proteins targeted by the platform at a lenient multiple testing threshold of $P < 1x10^{-4}$.

Other traits of interest were analysed in models then adjusted for age, sex, and first 10 genetic principal components, with and without additional adjustment for BMI. The glycaemic traits and other blood biochemistry measures were natural log transformed and then scaled prior to analysis. As described above, analysis was restricted to individuals who had phenotype and genotype data after excluding ancestry outliers and related individuals for each trait.

5.2.7 The European Prospective Investigation of Cancer (EPIC)-Norfolk study

Cohort Information

The European Prospective Investigation of Cancer (EPIC)-Norfolk study is a prospective cohort of 25,639 individuals aged between 40 and 79 and living in the county of Norfolk in the United Kingdom at recruitment. The study was approved by the Norfolk Research Ethics Committee (REC 500 ref. 98CN01) and all participants gave their written consent before entering the study.[375]

Measurements

Genotyping, imputation, and untargeted metabolite profiling of baseline non-fasted serum samples from 9,712 unrelated European individuals in the EPIC-Norfolk cohort was performed using the Discovery HD4 platform (Metabolon, Inc., Durham, USA), as previously described.[377,382]

Statistical Analysis

Linear regression models adjusted for age, sex, time of blood sample, time fasting and the first 10 genetic PCs were run for each *MC3R* variant and metabolite pair in R (v3.6.0). A total of 656 metabolites with a known chemical identity were included in the analysis. Statistical significance was considered at a Bonferroni significance threshold of $P < 7.6x10^{-5}$.

5.2.8 Assessment of frequency of MC3R variants in individuals with delayed puberty

Variant extraction and annotation

Variants were extracted from the coding region of MC3R (Hg38 chr20:56248732-56249815) from the UK biobank 450,000 WES data via DNAnexus SwissArmyKnife (V4.5.1). Prior variant quality control had been done to remove poor quality genotype calls using the MRC Epidemiology Unit in-house QC pipeline.[513]

Any variants which were flagged as poor quality were removed prior to genotype call preparation.

Qctool (v2.0.8; https://www.well.ox.ac.uk/~gav/qctool/) was used to extract variants form QC variant files and plink 1.9 was subsequently used to convert bgen file into raw genotype calls (0 = reference, 1 = heterozygous, 2 = homozygous).[503]

Variants were annotated using Ensembl variant effect predictor (VEP) and CADD (v1.6)[278,505]. Annotations for the MANE select transcript (ENST00000243911) were taken forward for variant clustering[505].

Variants were groups in tiers of predicted severity prior to the enrichment analysis:

- HIGH – all deletion and frameshift variants annotated as HIGH impact by VEP.
- Moderate_exclbenign- missense variants, excluding those that are annotated as benign by SIFT or Polyphen2
- Moderate_high_exclbenign – HIGH and Moderate_exclbenign combined
- CADD > 25 – variants with a CADD (phred scaled) score greater than 25.

Variant groupings were selected with individuals carrying a variant meeting the variant grouping above coded as 1 and non-carriers as 0. These were used to calculate the relative frequency of carriers in different puberty groups (see below).

Puberty definition

Puberty timing in females was defined as the reported age of menarche in years. We clustered individuals into three age of menarche groups based on clinical guidelines – delayed, early and normal age of puberty (N total =246,262). Individuals who reported age of menarche at age 15 or later were classed as having delayed age of menarche, with this being an age at which individuals would have been referred for further clinical assessment. Individuals who reported age of menarche at age 10 or earlier were classed as having early puberty onset. Individuals who reported age of menarche between ages of 10 and 15 were classified as having a normal age of menarche.

For male puberty timing we utilised information on reported age of voice breaking in males in UK biobank (N total = 190,485). This is a categorical variable where participants recall if they had earlier, average, or later age of voice breaking compared to their peers. Individuals who reported later age of voice breaking were classed as delayed and those that reported earlier than average were classed as having early puberty onset.

Once clusters were defined, we calculated the relative frequency of MC3R variant carriers based on the variant groups described above in each puberty timing group. Chi-square test was used to identify significant enrichment of MC3R variants in the delayed/early groups compared to controls for each variant class using R package epitools (version 0.5-10.1). Odds ratios were used to assess the relative enrichment of *MC3R* variants in the early/delayed groups compared to controls. Significant enrichment was defined at Bonferroni significance threshold for each sex ($P < 0.00625$, N tests = 8).

5.3 Results

5.3.1 Prioritisation of *MC3R* variants in several population cohorts:

Previous work in the O'Rahilly group aimed to identify coding variation in *MC3R* using pooled targeted exome sequencing of the coding region of *MC3R* in ALSPAC. This characterisation identified a total of 20 *MC3R* nonsynonymous coding variants in the ALSPAC cohort (**Table 5.4**).

To prioritise variants taken forward for functional characterisation, I further filtered these variants based on *in silico* prediction of functional impact. Using a strategy used for previous work in the O'Rahilly group characterising *MC4R*[489], I took forward variants for functional characterisation which were predicted to be damaging by SIFT[276] and PolyPhen2[277].

Table 5.4: Number of variants prioritised for *in vitro* functional follow up across 3 population studies.

Population	Genotype technology	N MC3R coding variants*	N variants prioritised for follow up*
ALSPAC	Pooled targeted sequencing	20	7
Genes & Health (Rare homozygous variants only)	Whole exome sequencing	2	2
UK Biobank	Genotyping array	5	3

*Note: some of the variants prioritised for follow up are overlapping between the different populations.

All variants studied in ALSPAC and UK Biobank are found in heterozygous carriers. To enable exploration of the effect of homozygous carriage of *MC3R* variants, we additionally prioritised variants detected by whole exome sequencing Genes & Health Study for which there are homozygous carriers.

We further identified variants that were directly genotyped in the UK Biobank Study, with whole exome sequencing data not being available in all individuals at the time of variant characterisation.

5.3.2 Functional characterisation of *MC3R* coding variants identifies variants that result in loss of function.

I measured the ability of the 11 prioritised predicted damaging variants to generate cAMP in human embryonic kidney (HEK293) cells *in vitro*, upon stimulation at several concentrations of [Nle_4, D-Phe_7]-α-MSH (NDP-MSH), ranging from 10^{-13} to 10^{-5}M. All variants taken forward for characterisation are rare, except *MC3R* p.V44I (**Table 5.5**). Predicted functional impact is summarised in **Table 5.6**.

Variants were subsequently classified based on their functional impact on E_{max} and EC50 (see methods). I additionally characterised 2 variants which were predicted to be benign (p.V44I and p.I50T) as well as *MC3R* Q11* which has been previously shown to be complete loss of function. P.Q11* and p.I50T have been previously characterised as benign and therefore acted as positive control[514]. P.M97I, which was found in a single homozygous carrier in Genes & Health, was selected despite being predicted benign due to this being found in homozygosity in this study.

I identified a total of the 5 of 11 variants that showed a significant difference in cAMP response ($P < 0.002$, Bonferroni; ANOVA) from wildtype *MC3R*, meeting the criteria of complete or partial loss of function (see methods). Five were classified as complete loss of function (cLoF) having < 25% of maximal wildtype (WT) *MC3R* activity and/or 50x EC50 compared to WT.

Table 5.5 Frequencies of variants of interest in population cohorts

Study	Amino Acid Change	Gnomad (overall) MAF	Alt Het Count	Alt Hom Count	Cohort MAF
UKBB	p.V44I	1.361E-01	79024/484019	5318/484019	1.009E-01
UKBB	p.F45S	6.774E-05	611/487743	0/487743	6.267E-04
UKBB	p.R220S	1.718E-03	1826/487988	4/487988	1.883E-03
ALSPAC	p.F45S	6.774E-05	4/5724	0/5724	3.494E-04
ALSPAC	p.L53R	-	1/5724	0/5724	8.735E-05
ALSPAC	p.I57T	2.389E-05	1/5724	0/5724	8.735E-05
ALSPAC	p.S90F	-	2/5724	0/5724	1.747E-04
ALSPAC	p.A214P	-	1/5724	0/5724	8.735E-05
ALSPAC	p.L260V	9.558E-05	3/5724	0/5724	2.621E-04
ALSPAC	p.S292T	1.591E-05	1/5724	0/5724	8.735E-05
G&H	p.M97I	6.011E-04	67/8921*	1/8921*	3.867E-03
G&H	p.G240W	1.990E-05	0/8921*	1/8921*	1.121E-04

Table 5.6: Variants identified in ALSPAC, Genes & Health (G&H) and UK Biobank (UKBB) studies.

Study	Technology	Chromosome Position (GRCh38)	Ref/Alt	Accession Number(s)	Amino Acid Change	Sift	Polyphen2	CADD Phred Score (v1.6)
UKBB	Genotyping Array	20:56248973	G/A	rs3827103,CM053305,COSV54763631	P.V44I	tolerated(0.29)	benign(0.05)	19.5
UKBB	Genotyping Array	20:56248977	T/C	rs143321797,CM0910908	p.F45S	deleterious(0)	probably_damaging(0.994)	26.2
UKBB	Genotyping Array	20:56249501	C/A	rs61735259,CM0911382,COSV54763198	p.R220S	tolerated(0.2)	possibly_damaging(0.902)	23
ALSPAC	Amplicon NGS	20:56248977	T/C	rs143321797,CM0910908	p.F45S	deleterious(0)	probably_damaging(0.994)	26.2
ALSPAC	Amplicon NGS	20:56249001	T/G	rs61736060	p.L53R	deleterious(0)	probably_damaging(0.998)	25.6
ALSPAC	Amplicon NGS	20:56249013	T/C	rs372944947	p.I57T	deleterious(0.04)	possibly_damaging(0.575)	24.6
ALSPAC	Amplicon NGS	20:56249112	C/T	-	p.S90F	deleterious(0)	probably_damaging(0.999)	26.7
ALSPAC	Amplicon NGS	20:56249483	G/C	-	p.A214P	deleterious(0)	probably_damaging(0.999)	27.5
ALSPAC	Amplicon NGS	20:56249621	C/G	rs146000025,CM127381	p.L260V	deleterious(0.03)	probably_damaging(0.998)	25.3
ALSPAC	Amplicon NGS	20:56249717	T/A	rs750492023	p.S292T	deleterious(0.04)	probably_damaging(0.995)	26.9
G&H	WES	20:56249134	G/A	rs199684791	p.M97I	tolerated(0.52)	benign(0.131)	21.3
G&H	WES	20:56249561	G/T	rs776955610	p.G240W	deleterious(0)	probably_damaging(1)	32

These include *MC3R* p.F45S which was identified in heterozygous carriers in ALSPAC and UKBB, and *MC3R* G240W which was identified in a single homozygous carrier in Genes & Health, meaning this individual has no functional MC3R (**Figure 5.5, Table 5.7**).

A further variant, p.R220S, was identified as partial loss of function having between 25 and 75% of maximal WT receptor activity and/or 5-50x EC50 compared to WT. No variants were identified to result in a gain of *MC3R* function compared to WT (**Table 5.7**). These cLoF and pLoF variants are predominantly located the transmembrane domains of MC3R (**Figure 5.4**).

The remaining 6 variants were identified to not be significantly different from WT, with between 75% and 125% of WT receptor activity and <5x WT EC50 (**Figure 5.5, Figure 5.5, Table 5.7**). This highlights the utility of *in vitro* characterisation to classify variants prior to any genetic association analyses with 53 % of the variants assayed, all of which were predicted *in silico* to be damaging, not having any true functional impact on downstream cAMP response of *MC3R*.

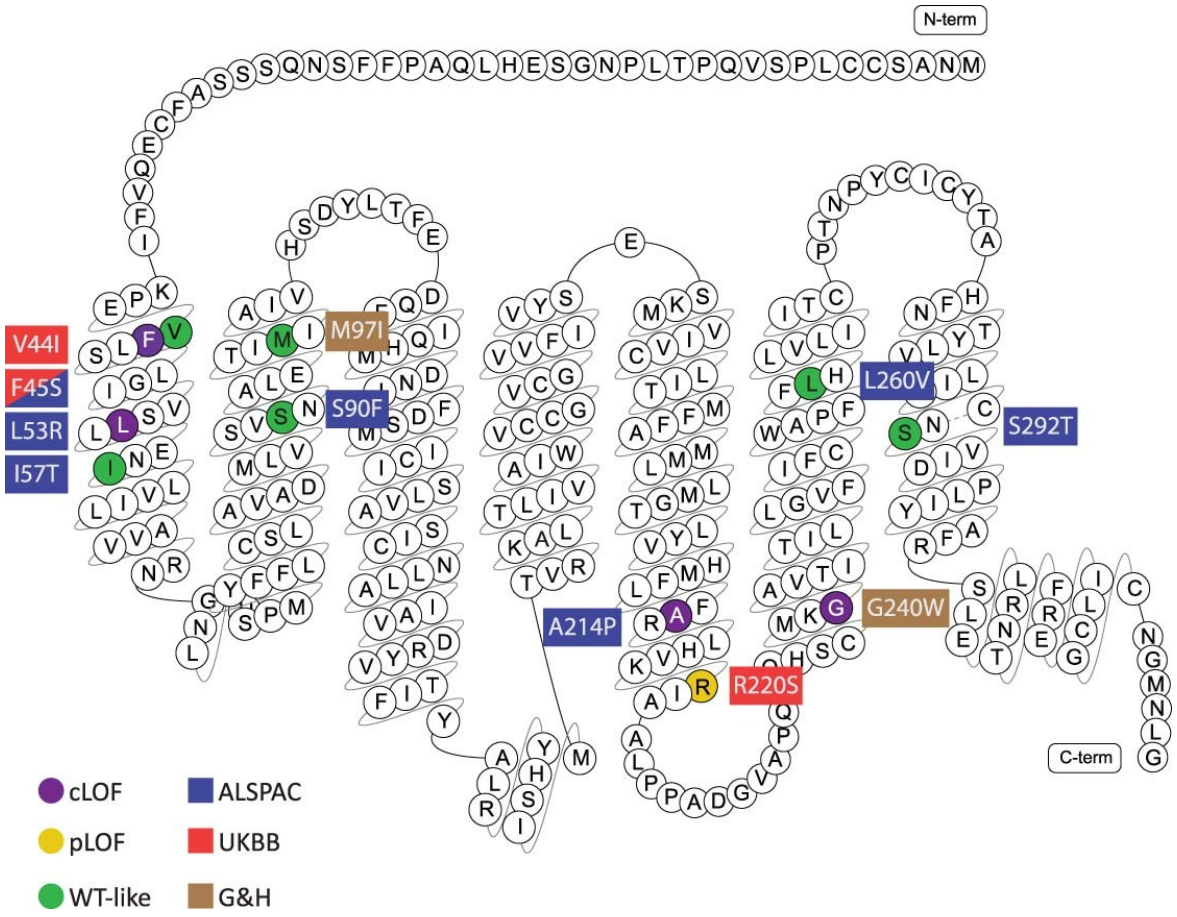


Figure 5.4: Schematic overview of the location of variants of interest in MC3R.

Colours represent the functional consequence and population in which these variants were identified, as indicated by the figure legend.

Table 5.7: MC3R cAMP assay results for variants of interest.

Study	Amino Acid Change	No. of biological replicates (N)	Mean logEC50 (logM;SEM)	Diff. LogEC50 compared to WT (-8.079)	Relative EC50 (linear fold-WT)	LogEC50 p-value	Bonferroni Significant?	Mean cAMP Relative E_{max} (%WT; SEM)	E_{max} p-value	Bonferroni Significant?	cAMP LoF Class
UKBB	P.V44I	7	-8.32 (0.20)	-0.241	0.57	0.379	N	93.61 (4.03)	0.494	N	wt-like
UKBB	p.F45S	13	-6.55 (0.06)	1.529	33.81	<0.0001	Y	53.12 (8.70)	<0.0001	Y	cLoF
UKBB	p.R220S	10	-8.08 (0.08)	-0.001	1.00	0.989	N	67.04 (7.49)	<0.0001	Y	pLoF
ALSPAC	p.F45S	13	-6.55 (0.06)	1.529	33.81	<0.0001	Y	53.12 (8.70)	<0.0001	Y	cLoF
ALSPAC	p.L53R	4	ND	ND	ND	ND	-	15.32 (2.045)	<0.0001	Y	cLoF
ALSPAC	p.I57T	3	-8.48 (0.36)	-0.401	0.40	0.365	N	101.00 (17.35)	0.932	N	wt-like
ALSPAC	p.S90F	6	-7.75 (0.30)	0.329	2.13	0.319	N	120.50 (12.98)	0.020	N	wt-like
ALSPAC	p.A214P	6	ND	nd	ND	ND	-	24.98 (22.04)	<0.0001	Y	cLoF
ALSPAC	p.L260V	6	-8.33 (0.14)	-0.251	0.56	0.482	N	117.50 (13.98)	0.065	N	wt-like
ALSPAC	p.S292T	15	-7.57 (0.05)	0.509	3.23	0.044	N	77.63 (6.70)	0.002	Y	wt-like
G&H	p.M97I	3	7.79 (0.26)	0.289	1.95	0.512	N	79.38 (7.66)	0.081	N	wt-like
G&H	p.G240W	3	ND	ND	ND	ND	-	-4.491 (7.14)	<0.0001	Y	cLoF

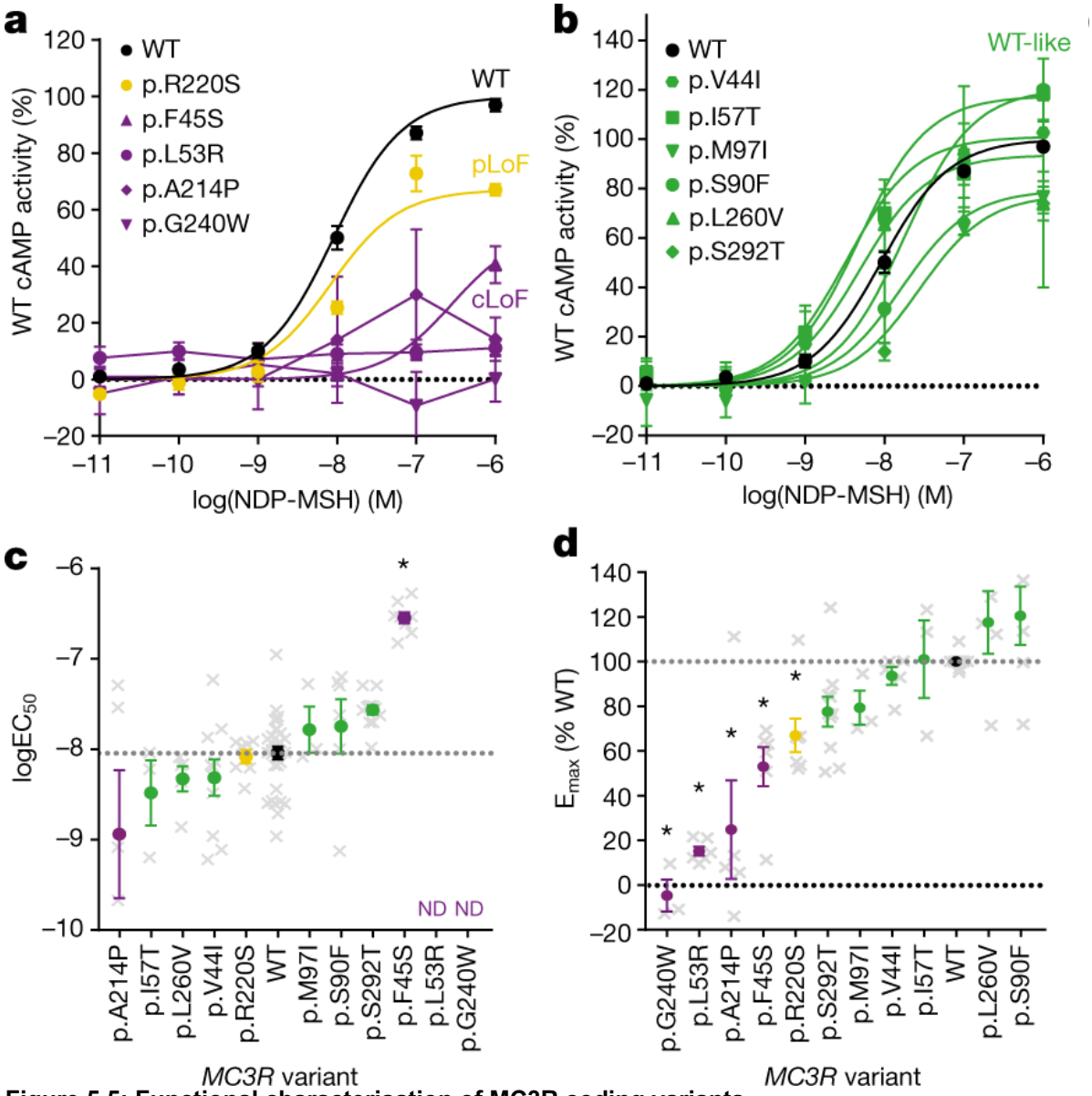


Figure 5.5: Functional characterisation of MC3R coding variants

a, b) Dose-dependent cAMP accumulation activity of MC3R mutants stimulated by NDP-MSH. MC3R mutants are grouped by their functional classification: LoF (**a**) and WT-like (**b**). Data are normalized to the percentage of WT (black). Mean ± s.e.m. is shown. Dotted line shows 0% WT cAMP activity; N and P values are listed in **Table 5.7**. cLoF, complete LoF; pLoF, partial LoF. **C, d**, logEC50 (**c**) and E_{max} (**d**) values of MC3R mutants. Mean ± s.e.m. is shown. Grey crosses show individual data points; grey dotted line shows WT response; black dotted line shows 0% WT activity. Bonferroni adjusted *P < 0.05 using one-way ANOVA; N and P values are listed in **Table 5.7**.

5.3.3 Heterozygous carriage of dysfunctional *MC3R* variants

UK Biobank

Previous studies in mouse and humans have shown possible association of *MC3R* with metabolic, developmental, and anthropometric traits. We therefore assessed the association of the three variants identified in UKBB with an array of traits within these categories – p.F45S, p.R220S and p.V44I. All three *MC3R* variants were associated with a delay in puberty in male and female participants. Individuals heterozygous for the rarer p.F45S and p.R220S variants, which result in a more substantial disruption of cAMP signalling, had a greater delay in pubertal onset, with female carriers of p.F45S mutation having a 5.16-month delay. This suggests a dose-dependent relationship of degree of *MC3R* loss of function and effect on phenotype (**Figure 5.6**). These variants were also associated with reduced growth, as indicated by shorter total and sitting height in adults (**Fig 5.6a&c, Table 5.8**), and shorter relative childhood height at age 10 years (**Fig 5.6a, Table 5.8**), a similar dose dependent effect was seen for these traits.

The much more common p.V44I variant was also significantly associated with age at puberty and height, albeit with a substantially smaller effect size (**Figure 5.5 and 5.6**). Although this variant exhibits no significant difference from WT in the cAMP assay, we hypothesise that the large numbers of carriers (~50,000) allowed us to discern a phenotypic impact of a reduction in signalling resulting from this mutation not discernible in a heterologous over-expression system (~94% of WT, **Figure 5.5 and 5.6**). Alternatively, the effect may be explained by its LD with the 5'UTR variant (or other non-coding variants) that could affect the expression of *MC3R.* Carriers of these three variants also had lower total body lean mass and a reduced ALM/BMI ratio (**Figure 5.5 and 5.6, Table 5.8**).

There was some heterogeneity between individual variant associations – notably, associations with childhood height were more consistent than with adult height (**Figure 5.6a, Table 5.8**) and p.R220S was also associated with lower circulating IGF1 levels (**Figure 5.6a&e, Table 5.8**).

No variant showed any association with BMI, waist-to-hip ratio, fat mass, type 2 diabetes, HbA1c or random glucose (**Figure 5.6e, Table 5.8**). This is an important observation as it discounts the associations reported in previous small scale family studies of dysfunctional, heterozygous MC3R variants, and clearly highlights the lack of impact of heterozygous carriage of deleterious MC3R variants on energy expenditure and food intake, when looking at these variants at population-scale.

Gene burden tests conducted by Dr John Perry at the MRC Epidemiology Unit, confirmed the association of phenotypes with burden of predicted loss of function MC3R variants seen for the individual genotype variants (**Table 5.8**).

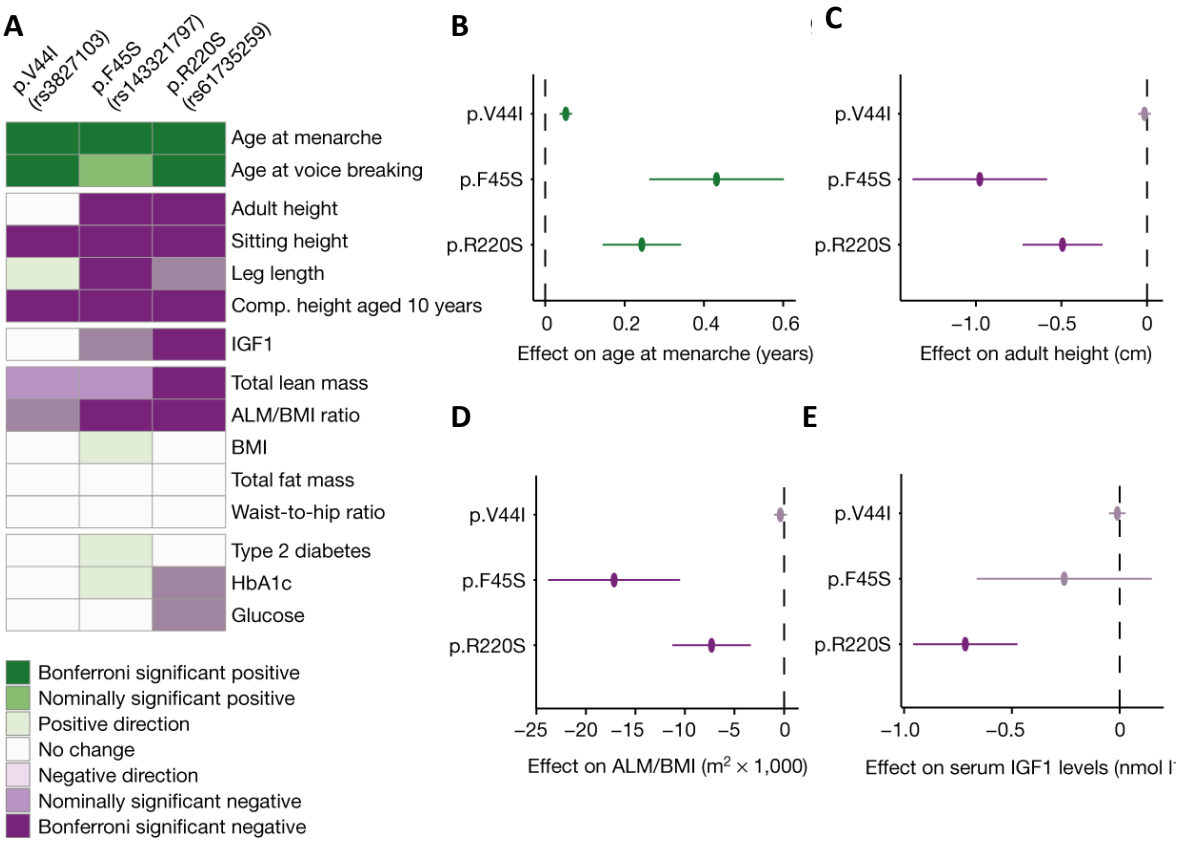


Figure 5.6: Association of MC3R variants with relevant traits in UK Biobank
A) Direction and significance of association of MC3R variants of interest with relevant traits in UK Biobank. **B-E**) The effect of UKBB *MC3R* variants on the age at menarche (B), adult height (C), the ALM/BMI ratio (D), and plasma levels of IGF1 (in nmol l^{-1}) (E). Beta ± 95% CI is shown. Dashed line indicates an effect size of 0; colour indicates level of significance (key in A).

Table 5.8: Phenotype association of array genotyped MC3R variants identified in UKBB

		Gene Burden Test of *MC3R* WES variants*	p.V44I, rs3827103 G>A WT-like, MAF 10.09%	p.F45S, rs143321797 T>C cLoF, MAF 0.06%	p.R220S, rs61735259 C>A pLoF, MAF 0.19%
Puberty Timing	**Age at menarche in females (years)**	**0.39 (0.22; 0.49)** **P = 6.4×10^{-12}**	**0.05 (0.04; 0.07)** **P = 7.60×10^{-11}**	**0.43 (0.26; 0.60)** **P = 7.10×10^{-7}**	**0.24 (0.14; 0.34)** **P = 1.40×10^{-6}**
	Age at voice breaking in males (3 categories)	**0.051 (0.027; 0.075)** **P = 2.6×10^{-5}**	**0.007 (0.004; 0.011)** **P = 7.20×10^{-5}**	**0.047 (0.008; 0.086)** **P = 0.019**	**0.038 (0.015; 0.061)** **P = 8.60×10^{-4}**
Adult Height	**Adult height (cm)**	**-0.43 (-0.9; -0.17)** **P = 0.004**	-0.015 (-0.05; 0.02) P = 0.480	**-0.98 (-1.37; -0.59)** **P = 2.10×10^{-6}**	**-0.49 (-0.73; -0.26)** **P = 6.10×10^{-5}**
	Sitting height (cm)	**-0.37 (-0.54; -0.21)** **P = 2.6×10^{-5}**	**-0.04 (-0.07; -0.018)** **P = 5.60×10^{-4}**	**-0.55 (-0.80; -0.29)** **P = 2.10×10^{-5}**	**-0.34 (-0.49; -0.19)** **P = 7.50×10^{-6}**
	Leg length (cm)	-0.05 (-0.24; 0.14) P = 0.79	0.02 (-0.005; 0.05) P = 0.14	**-0.48 (-0.77; -0.19)** **P = 0.002**	-0.17 (-0.34; 0.004) P = 0.063
	Sitting height / Height ratio (*1000)	**-0.95 (-1.64; -0.26)** **P = 0.008**	**-0.17 (-0.27; -0.07)** **P = 6.80×10^{-4}**	-0.18 (-1.26; 0.90) P = 0.59	-0.42 (-1.06; 0.21) P = 0.23
Childhood Height	**Comparative Height at age 10 (3 categories)**	**-0.067 (-0.1007; -0.034)** **P = 2.3×10^{-4}**	**-0.014 (-0.019; -0.010)** **P = 3.00×10^{-4}**	**-0.093 (-0.143; -0.043)** **P = 1.80×10^{-4}**	**-0.087 (-0.116; -0.057)** **P = 1.90×10^{-8}**
	Comparative Height at age 10 (adj. Adult Height)	**-0.043 (-0.070; -0.016)** **P = 0.002**	**-0.013 (-0.017; -0.009)** **P = 3.70×10^{-10}**	-0.037 (-0.080; 0.007) P = 0.095	**-0.052 (-0.078; -0.026)** **P = 9.00×10^{-5}**
Blood biochemistry – growth	**IGF1 (nmol/l)**	**-0.37 (-0.63; -0.11)** **P = 0.007**	-0.01 (-0.05; 0.03) P = 0.61	-0.26 (-0.66; 0.15) P = 0.12	**-0.72 (-0.96; -0.47)** **P = 9.40×10^{-10}**
Obesity-Related Traits	**BMI (kg/m2)**	-0.043 (-0.26; 0.18) P = 0.65	-0.015 (-0.048; 0.018) P = 0.44	0.22 (-0.13; 0.58) P = 0.26	-0.05 (-0.26; 0.16) P = 0.63
	Total fat mass (grams)	47.2 (-480.2; 385.8) P = 0.76	-25 (-89; 39) P = 0.48	345 (-335; 1026) P = 0.32	-155 (-559; 250) P = 0.42

	Body Fat percentage (inverse normal transformed)	0.001 (-0.047; 0.048) P = 0.98	-0.002 (-0.009; 0.006) P = 0.73	0.057 (-0.018; 0.133) P = 0.15	0.003 (-0.041; 0.048) P = 0.92
	Total lean mass (grams)	**-332.79 (-599.64; - 65.94)** **P = 0.01**	**-59 (-98; -19)** **P = 0.004**	**-638 (-1054; -222)** **P = 0.0038**	**-497 (-744; -250)** **P = 5.70×10^{-5}**
	Total lean mass (grams) adjusted for height	-108 (-335; 119) P = 0.32	**-44 (-78; -10)** **P = 0.011**	**16 (-378; 410)** **P = 0.86**	**-221 (-434; -8)** **P = 0.04**
	Lean mass Index	-0.03 (-0.11; 0.05) P = 0.43	**-0.013 (-0.024; -0.001)** **P = 0.036**	-0.009 (-0.133; 0.114) P = 0.90	**-0.064 (-0.137; 0.009)** **P = 0.093**
	ALM/BMI (m^2) (*1000)	**-6.75 (-11.10; -2.42)** **P = 0.004**	-0.38 (-1.01; 0.25) P = 0.22	**-17 (-24; -10)** **P = 1.30×10^{-6}**	**-7.3 (-11.3; -3.4)** **P = 3.90×10^{-4}**
	Waist-hip ratio (normalised, adj. BMI)	0.00073 (-0.0021; 0.0035) P = 0.64	-0.00012 (-0.00060; 0.00036) P = 0.60	-0.0013 (-0.0064; 0.0037) P = 0.53	0.00001 (-0.00301; 0.00302) P = 1
Glycaemic Control	**Type 2 diabetes (odds ratio)**	0.99 (-0.98; 1.001) P = 0.20	1.00 (0.97; 1.02) P = 0.78	1.24 (0.88; 1.74) P = 0.21	0.92 (0.72; 1.18) P = 0.51
	HbA1c (mmol/mol)	-0.015 (-0.058; 0.029) P = 0.44	0.0004 (-0.0063; 0.0070) P = 0.85	0.064 (-0.006; 0.135) P = 0.16	-0.035 (-0.076; 0.007) P = 0.25
	Random glucose (mmol/L)	-0.044 (-0.09; 0.006) P = 0.08	-0.00062 (-0.00839; 0.00714) P = 0.86	0.029 (-0.052; 0.11) P = 0.54	-0.031 (-0.080; 0.018) P = 0.21

The effect estimate (*beta*) and its 95% confidence intervals are shown in the table. ALM = Appendicular Lean Mass
For single variant analyses P values are determined by the non-infinitesimal mixed model association test from BOLT-LMM.
*WES – Whole Exome Sequencing in ~200K UKBB participants. Gene burden analysis was conducted by Dr John Perry, MRC Epidemiology Unit, University of Cambridge including variants annotated as “high” or “moderate” impact by VEP and excluded those annotated as benign by Polyphen2, with a MAF < 0.2% within MC3R[414]

5.3.4 Phenotypic follow up

Genetic risk scores were constructed which allow the assessment of the combined effect of these 3 variants of interest in UK Biobank (p.V44I, p.F45S and p.R220S). Allele scores were weighted using several strategies including their effect on age of menarche, the most strongly associated trait, CADD score and percentage reduction E_{max} (%) compared to wildtype *MC3R in vitro* (see previous section).

I determined the interdependent effect of each variant in the score by running joint conditional model adjusting for each variant to test for association with age of menarche in UKBB (**Table 5.9**). This showed that each variant had an independent effect on puberty timing.

The higher GRS (indicative of lower MC3R activity) was associated with later puberty timing in both sexes, shorter adult and childhood stature, lower sitting height, lower total body lean mass, and lower appendicular lean mass to BMI ratio, with better overall sensitivity than any of the individual variants (**Table 5.10**).

Table 5.9 Independent association of array genotyped MC3R variants with age of menarche in UK Biobank

rsid	Beta	SE	Pvalue
rs3827103	0.045803	0.009826	3.14E-06
rs14332179	0.513033	0.103371	6.94E-07
rs61735259	0.186079	0.061020	2.29E-03

I also constructed a menarche-weighted GRS for sensitivity analyses as well as weighting by functional parameters from the cAMP assays. These showed similar associations to those with the CADD-weighted GRS. The results further broadly reflect that of the individual component SNPs (**Table 5.10, Table 5.8**). These include a strong association with delay in puberty timing and reduced height in childhood and adulthood. No association was identified for obesity related traits or those related to glycaemic control, consistent with our observations for the individual variants.

Table 5.10 MC3R GRS associations with traits of interest in UK Biobank

MC3R GRS	Menarche-weighted GRS		CADD-weighted GRS		difference in E_{max} – weighted GRS		difference in EC50 – weighted GRS	
Trait (Units)	**Beta**	**P**	**Beta**	**P**	**Beta**	**P**	**Beta**	**P**
Age at menarche in females (years)	1.00000	8.28E-15	0.00292	7.50E-10	0.00697	4.46E-12	0.22397	2.88E-10
Age at voice breaking in males (3 categories)	0.12058	2.91E-05	0.00043	4.67E-05	0.00099558	1.01E-05	0.01998275	1.19E-02
Height at age 10 (3 categories)	-0.29320	2.23E-13	-0.00096	6.47E-11	-0.00229	2.19E-13	-0.05505	5.65E-07
Height at age 10 (adj Adult Height)	-0.17269	7.59E-08	-0.00071	1.59E-09	-0.00156	5.07E-10	-0.02188	1.33E-02
Adult height (cm)	-1.94638	1.47E-07	-0.00391	4.11E-03	-0.01163609	6.05E-05	-0.54303	9.34E-08
Sitting height (cm)	-1.38429	8.42E-11	-0.00339	1.62E-05	-0.00916	4.24E-08	-0.34277	4.79E-09
Sitting height / Height ratio	-0.00200	2.33E-02	-0.00001	2.11E-02	-0.00002	1.24E-02	-0.00030	2.13E-01
Leg length (cm)	-0.55931	2.91E-02	-0.00054	5.66E-01	-0.00245	2.23E-01	-0.19913	4.66E-03
Appendicular lean mass (Kg)	-0.76513	2.65E-05	-0.00224	8.28E-04	-0.00588544	3.71E-05	-0.15470	1.98E-03
Appendicular lean mass/Height2 (kg/m^2)	-0.08633	1.04E-01	-0.00039	4.36E-02	-0.00094	2.36E-02	-0.00519	7.22E-01
Appendicular lean mass/BMI (m^2)	-0.02992	1.28E-07	-0.00006	8.17E-03	-0.00017	1.45E-04	-0.00879	1.60E-08
Total lean mass (grams)	-1476.48620	1.29E-05	-4.23494	6.80E-04	-11.19587	2.44E-05	-305.87097	1.00E-03
Total lean mass (grams, adjusted for height)	-497.10843	7.58E-02	-2.24874	2.91E-02	-5.28714	1.59E-02	-34.71843	6.52E-01
Lean mass index (total lean mass Kg/height^2)	-0.11872	2.21E-01	-0.00064	0.0730215	-0.00149	4.99E-02	0.00135	9.60E-01
Total fat mass (grams)	212.72849	6.91E-01	-0.82167	6.77E-01	-1.57625	7.07E-01	178.57280	2.25E-01
Body Fat Percentage (inverse normal transformed)	0.07462	2.02E-01	0.00003	9.04E-01	0.00017	7.08E-01	0.03242	4.34E-02
IGF1 (nmol/l)	-0.92870	3.86E-03	-0.00157025	1.85E-01	-0.00805	1.42E-03	-0.19621	2.59E-02
Random glucose (mmol/L)	-0.00383	9.59E-01	-0.00004	8.98E-01	-0.00009	8.74E-01	0.00173	9.32E-01
HbA1c (mmol/mol)	0.37797	3.15E-01	0.00020	8.84E-01	-0.00002	9.95E-01	0.18773	6.92E-02
Type 2 Diabetes, Feb 2021 Definition (odds ratio)	1.17106	4.89E-01	0.99993	9.36E-01	0.99998	9.90E-01	1.08130	1.84E-01
Waist-hip ratio (normalised, adj. BMI)	-0.01952	7.39E-01	-0.00004	8.67E-01	0.00005	9.08E-01	-0.01040	5.17E-01
BMI (kg/m2)	0.13231	6.29E-01	-0.00047	6.38E-01	-0.00100033	6.41E-01	0.11116	1.41E-01

The Fenland Study: association of MC3R LoF with additional metabolic traits

I further assessed the association of the identified *MC3R* LoF variants assessed in UK Biobank in the Fenland Study. I assessed the association of MC3R variants with glycaemic measures during an oral glucose tolerance test (OGTT) in individuals without diabetes (diagnosed or biochemical; see methods) and additional blood biochemistry measures which are not available in UK biobank.

I identified nominal association of the *MC3R* GRS, F45S and V44I with higher fasting proinsulin and fasting insulin levels. However, these associations did not replicate larger meta-analyses for these traits by the MAGIC consortium[87] with a much larger sample size, further suggesting there is no association of *MC3R* loss of function with diabetes and glycaemic traits, unlike *MC4R* and other proteins in the central melanocortin system.

Phenome-wide association study (pheWAS) of MC3R

To explore other phenotypic associations of MC3R in a hypothesis free manner, I ran a phenome wide association study (pheWAS) for each of the 3 variants identified in UK biobank as well as the CADD weighted GRS using publicly available GWAS data from several databases (see methods). These variants were selected as these are genotyped on genotyping arrays and therefore are many in several GWAS studies, unlike the other variant identified using exome sequencing approaches. Fewer phenotypes were available for MC3R p.F45S and therefore the GRS compared to p.R220S and p.V44I due to minor allele frequency filters in many studies resulting in p.F45S not being included. This confirmed significant associations with age of menarche and sitting height surviving multiple testing correction ($p < 1.046x10^{-4}$) (**Figure 5.7**).

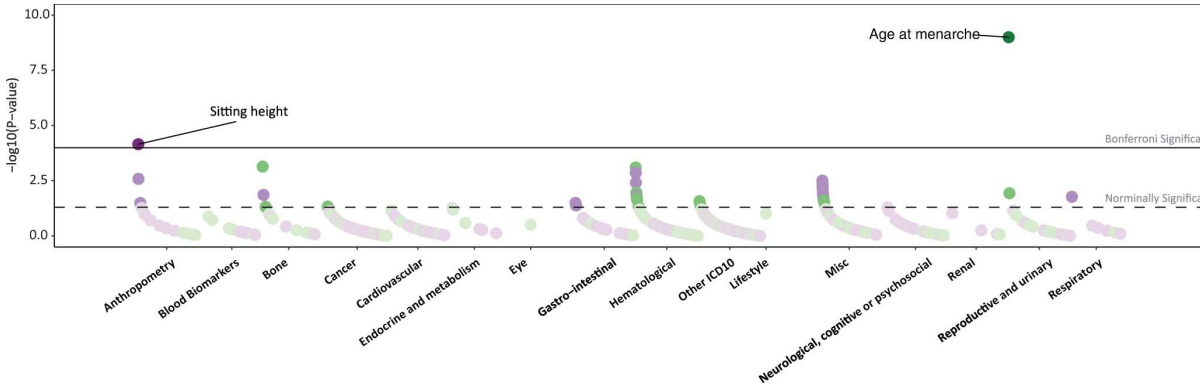


Figure 5.7: PheWAS analysis of MC3R genetic risk score.

A CADD-weighted MC3R genetic risk score was constructed (see Methods) and used to conduct a phenome-wide analysis (pheWAS) with publicly available summary statistics. Solid black line indicates Bonferroni multiple-testing threshold of $p < 1.046 \times 10^{-4}$, dashed line indicates nominal significance threshold $p < 0.05$.

'Omic associations with MC3R LoF

To explore the effect of *MC3R* variants on the plasma proteome and metabolome, I used data from the Fenland Study[382] and EPIC Norfolk[377,515], respectively. I identified IGFBP1, a liver-derived protein which is known to be suppressed by growth hormone (GH)[515] (**Supplementary Table S5.1**). The two most strongly associated metabolites with *MC3R* p.F45S, pipecolate (beta = 1.1, SE =0.33, P =9.6×10^{-4}) and 4-hydroxyphenylpyruvate (beta =0.96, SE =0.31, P =0.0025) are metabolites of lysine and tyrosine respectively and therefore likely reflect increased proteolysis (**Supplementary Table S5.2)**. These associations, while potentially illuminating, did not reach stringent, multiple test corrected thresholds (see methods). Further studies with a larger sample size will be needed to further confirm these observations.

5.3.5 A human knockout: homozygous carriage of *MC3R* LoF variant MC3R p. G240W

In the exome data from participants in the Genes & Health (G&H) study, of whom 18.8% report parental relatedness[511], I identified two rare, homozygous non-synonymous mutations p.M97I and p.G240W, each in one participant. While p.M97I signalled normally, the p.G240W mutant receptor was completely unresponsive in cAMP assays (**Figure 5.4; Table 5.7)**.

The participant carrying p.G240W was Invited for phenotypic assessment under ethically approved recall protocols by Dr Sarah Finer and colleagues at Queen Mary University of London and gave informed consent for publication of results. He is a male of Bangladeshi origin, in his early 40s whose parents are second cousins. The mutation is in an 8.3Mb genomic region of homozygosity, consistent with consanguineous inheritance. He reported a history of significantly delayed puberty, starting in his early 20s after which he subsequently fathered children. He was of markedly short stature, -2.95 SDs of the mean by WHO reference[516]. His sitting height ratio was below the normal range for South Asians and had reduced circulating levels of IGF1 (**Table 5.11**).

In contrast to the finding in heterozygotes, he has been overweight/obese since early childhood and currently has a BMI of 40.4 kg/m^2 (**Table 5.11**), accompanied by type 2 diabetes and hypertension, both well-controlled. I screened the proband's exome sequence for all known monogenic obesity genes, no pathogenic mutations in obesity-genes were identified. It therefore is unclear whether *MC3R* deficiency in the homozygous state results in obesity and defects in glycaemic control, in contrast to heterozygous carriers.

Table 5.11: Clinical characteristics of homozygous MC3R G240W proband

Phenotypic characteristic of MC3R G240W proband		**Normal range**
Age (years)	Early 40s	
Height (standing)	**155.0cm**	WHO (2007) Z-score of proband -2.95*
Weight	96.96kg	
Body mass index	**40.4kg/m^2**	
Sitting height (DEXA-derived measurement)	**75.2cm**	
Sitting height ratio (DEXA-derived measurements)	**48.5**	50.9 – 52.8#
Total Body fat (DEXA)	**48.50%**	UKBB average: 29.2%^
Total lean mass (DEXA)	**49.20%**	UKBB average: 67.1%^
Appendicular lean mass:body mass index (DEXA)	**0.54**	< 0.9 defined as sarcopenia
Blood pressure	120/75mmHg	
IGF1 (nmol/L)	**6.3 (L)**	11.6 – 31.3
IGFBP1 (ng/mL)	7.1	May-34
IGFBP3 (mg/L)	2.9	2.5 – 6.0
LH (U/L)	2.8	1.5 – 6.3
FSH (U/L)	3.1	1.0 – 10.1
Testosterone (nmol/L)	14.2	7.2 – 31.3 (Adult male)
Sex Hormone Binding Globulin (nmol/L)	47.1	10.0 – 57.0
TSH (mU/L)	1.01	0.35 – 5.50
Free T4 (pmol/L)	14.4	10.5 – 21.0
Fasting Plasma Glucose (mmol/L)	5.9 (in the context of treated diabetes)	3.5 – 6.0
HbA1C (IFCC) mmol/mol	41 (in the context of treated diabetes)	< 42
Insulin (pmol/L)	**102**	0 – 80
C-Peptide (pmol/L	**1,110**	174 – 960
9am Cortisol (nmol/L)	**395**	> 420
ACTH ng/L	45	≤ 50
NEFA (umol/L)	**361**	≤ 64
Adiponectin (ug/mL)	5.2	2.9 – 14.9+ Consistent with obesity and insulin resistance
Leptin (ug/L)	**40.2**	7.8 – 31.7 for BMI >35

All clinical characteristics of the proband were obtained by Dr Sarah Finer and colleagues, Queen Mary University of London.
Blood sampling was undertaken by venepuncture after a 10 hour overnight fast, and using standard clinical laboratory protocols and assays performed at the University of Cambridge Core Biochemistry Assay Laboratory.
*Using WHO 2007 growth reference (de Onis et al (2007)[494], plotted using the maximum age, 19 years.
#Sitting height ratio (SHR) for South Asian males in UKBB (N=3982), 1st and 3rd quartiles shown.
^European males with DEXA data in UKBB (N=2,367).
'+Based on Semple et al. (2006)[494]

Whole body dual-energy X-ray absorptiometry (DEXA) scanning (**Figure 5.8**) revealed a high percentage of body fat at 48.5% (**Figure 5.8a-c**), but a low total lean mass for his level of BMI (**Figure 5.8d**).

I compared the total body lean mass of the proband to matched male control groups in UK biobank (ethnicity, age, and BMI) which showed that the proband had significantly lower lean mass than expected irrespective of BMI (**Figure 5.8b-d**). We further show that the proband has significantly lower ALM/BMI ratio, an index of sarcopenia, was below normal (**Figure 5.8e, Table 5.11**).

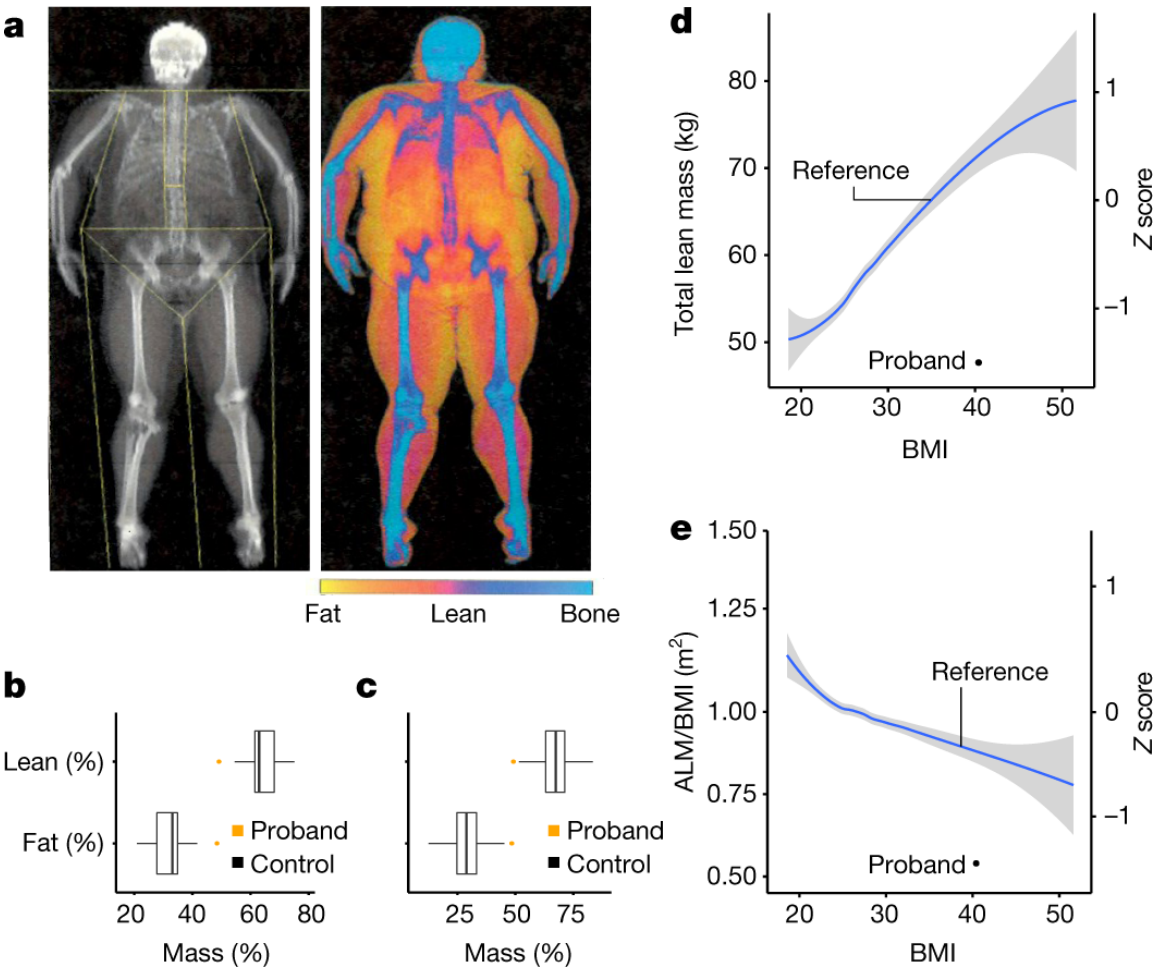


Figure 5.8: Characteristics of an individual who is homozygous for the MC3R p.G240W mutation.
A, Whole-body dual-energy X-ray absorptiometry image of the *MC3R* p.G240W proband. **B,C**, Boxplots of the percentage of fat and lean mass in proband (orange) compared with South Asian men ($N = 36$) (**b**) and age-matched European men aged 44–54 years ($N = 417$) (**c**) in the UKBB. The centre line indicates the median, the box is the interquartile range, and the whiskers indicate 1.5 times the interquartile range. **D**, **E**, The total lean mass (black circle) (**d**) and the ALM/BMI ratio (**e**) of the proband compared with men from the UKBB with a BMI from 18 to 52 (the mean is indicated in blue and the grey shading indicates the 95% CI of the mean; $N = 2,356$).

5.3.6 Investigation of the enrichment of MC3R variants in individuals who reported delayed puberty in UK Biobank

Following the identification that carriage of LoF *MC3R* variants is associated with delay in puberty I utilised the UK Biobank 450k WES dataset to assess whether there is an enrichment in MC3R coding variants that are predicted to be damaging in individuals who reported a significant delay in puberty onset.

There was a significant enrichment in frequency of carriers of *MC3R* predicted damaging variants in the delayed puberty groups for both males and females compared to the normal and early onset groups. Frameshift and premature stop variants (variant class = HIGH) which have high confidence of resulting in cLoF of *MC3R* are rare in UKBB 450k, found in 0.16 % of females and 0.059% males in UKBB 450k with average age of puberty onset. These predicted cLoF variants are found in 0.056% of females who reported age of menarche at 15 years or older, compared to 0.024% in the normal age range (**Table 5.12**), a 2.37x enrichment (P chi square = $4.68x10^{-3}$). This enrichment is also seen for missense variants which are predicted to be damaging by several prediction algorithms with a 1.49- and 1.67-fold enrichment in MC3R predicted LoF carriers in the delayed menarche group compared to normal age controls (**Table 5.12**), for “Moderate excl benign” and CADD > 25 variant grouping respectively (**Table 5.12**). There was nominal of enrichment of MC3R predicted LoF variants in individual who reported early age of menarche (**Table 5.12**), with a trend towards lower percentage of MC3R *predicted* LoF variants compared to normal range groups. This suggests that variants that result in gain of function of MC3R, which would be predicted to result in earlier onset of puberty, are rare. A similar pattern of enrichment was seen of reported age of voice breaking in males (**Table 5.13**).

The identification of an enrichment in *MC3R* predicted LoF variant in groups who reported delayed puberty, adds evidence of the role of *MC3R* in development and suggests that *MC3R* variants should be screened for in patient cohorts presenting with delayed puberty.

Table 5.12: Enrichment of MC3R predicted damaging variants in UKBB participants reporting delayed age of menarche (N = 246,262)

Variant class	Menarche timing	N Noncarriers	N Carriers	Carrier Frequency (%)	OR vs normal timing	95% CI	p-value
HIGH	Delayed (≥16Yrs)	23216	13	0.0560	2.3466	1.17-4.38	*0.0098*
HIGH	Early (< 10Yrs)	9321	2	0.0215	0.8992	0.11-3.42	*1*
HIGH	Normal	213725	51	0.0239	-	-	-
MODERATE excl. benign	Delayed (≥16Yrs)	23088	141	0.6070	1.5155	1.26-1.81	*1.45E-05*
MODERATE excl. benign	Early (< 10Yrs)	9298	25	0.2682	0.6672	0.43-0.99	*0.0430*
MODERATE excl. benign	Normal	212918	858	0.4014	-		-
MODERATE/HIGH excl. benign	Delayed (≥16Yrs)	23076	153	0.6587	1.5527	1.30-1.85	*1.77E-06*
MODERATE/HIGH excl. benign	Early (< 10Yrs)	9296	27	0.2896	0.6802	0.45-1.00	*0.0489*
MODERATE/HIGH excl. benign	Normal	212867	909	0.4252	-	-	-
CADD ≥25	Delayed (≥16Yrs)	23097	132	0.5683	1.6633	1.37-2.01	*3.90E-07*
CADD ≥25	Early (< 10Yrs)	9300	23	0.2467	0.7198	0.45-1.09	*0.1437*
CADD ≥25	Normal	213044	732	0.3424	-	-	-

5.13: Enrichment of MC3R predicted damaging variants in UKBB participants reporting delayed age of voice breaking

Variant class	Age at voice breaking	N Noncarriers	N Carriers	Carrier frequency (%)	OR	95% CI	p-value
HIGH	Older than average	11211	2	0.0178	0.6223	0.07-2.37	*0.7698*
HIGH	Younger than average	8347	1	0.0120	0.4179	0.01-2.44	*0.7307*
HIGH	About average age	170927	49	0.0287	-	-	-
MODERATE excl. benign	Older than average	11141	72	0.6421	1.5520	1.20-1.98	*7.46E-04*
MODERATE excl. benign	Younger than average	8319	29	0.3474	0.8372	0.56-1.21	*0.3819*
MODERATE excl. benign	About average age	170267	709	0.4147	-	-	-
MODERATE/HIGH excl. benign	Older than average	11139	74	0.6599	1.4918	1.16-1.90	*1.80E-03*
MODERATE/HIGH excl. benign	Younger than average	8318	30	0.3594	0.8099	0.54-1.17	*0.3082*
MODERATE/HIGH excl. benign	About average age	170218	758	0.4433	-	-	-
CADD ≥25	Older than average	11154	59	0.5262	1.4558	1.09-1.90	*8.05E-03*
CADD ≥25	Younger than average	8327	21	0.2516	0.6941	0.43-1.07	*0.1093*
CADD ≥25	About average age	170357	619	0.3620	-	-	-

5.4 Supporting evidence: results from collaboration.

The following section describes in brief some of the additional results contributed by collaborators as part of publication Lam, Williamson, Finer et al (2021)[414]. These summaries are included as the results provide basis of some of the concluding discussions in this chapter.

MC3R LoF impacts height throughout development

To study the impact of *MC3R* LoF throughout development, we studied 5,993 unrelated participants from the Avon Longitudinal Study of Parents and Children (ALSPAC). This birth cohort allows us to assess the impact of *MC3R* variation on traits of interest throughout childhood and early adulthood with longitudinal data collected at several time points throughout this period.

An aggregated burden test of the three LoF variants (p.F45S, p.L53R and p.A214P) identified in a total of 6 heterozygous carriers on was performed on anthropometric trajectories and pubertal timing. We found that despite the small sample size ($N = 6$), *MC3R* LoF mutations were associated with lower height throughout childhood, adolescence, and early adulthood, with a trend towards lower lean mass and lower weight. No effect on pubertal onset was discernible in this small group.

MC3R expression in the hypothalamus

Dr Brian Lam generated a harmonised single cell RNA sequencing (scRNA-seq) map of the mouse hypothalamus, incorporating 18,427 neurons from 4 published studies. Within this there were identified to be 1,166 Mc3r expressing neurons, which formed 11 distinct clusters. This analysis identified a cluster of Mc3r expressing neurons co-expressing *Kiss1* and *Ghrh*, indicating neuronal populations that control reproduction and growth respectively. The expression of *Mc3r* in these neuronal populations was further confirmed using single-molecular *in situ* hybridisation in mouse and human hypothalamus.

MC3R deficiency in mice results in delayed puberty

A delay in puberty was observed in *Mc3r* knockout (KO) mice compared to wildtype control littermates. In male *Mc3r* KO mice a delay of 2 day was identified Kruskal–

Wallis test, $P = 0.015$; N values: 4 WT, 14 Mc$3r^{+/-}$ and 9 $Mc3r^{-/-}$. A similar trend was observed for female mice, with *Mc3r* KO having a delay in puberty compared to wildtype littermate controls however this did not meet significance (Kruskal–Wallis test, $P = 0.280$); likely due to low numbers of female mice included: N values: 4 (WT), 15 ($Mc3r^{+/-}$) and 7 ($Mc3r^{-/-}$). In adult female MC3R KO mice, the length of oestrous cycle was prolonged in both the fed ($P < 0.05$) and fasted condition ($P < 0.0001$; two-way ANOVA with Bonferroni post-hoc), and in WT and MC3R KO mice subject to an overnight fast, the lengthening effect of fasting on oestrous cycle was abolished.

5.5 Discussion

Summary of findings

Through the systematic assessment of *MC3R* loss of function in humans, along with studies of human and mouse physiology from collaborators we have identified the role of MC3R in human growth and development. This shows, across several population cohorts the robust association of *MC3R* LoF with delayed puberty in males and females, shorter height in childhood and adulthood and reduced lean mass. This along with evidence from mouse studies shows the role of MC3R in growth through likely GHRH neurons and puberty timing likely through KISS1 neurons.

It is well documented that caloric deprivation, such as in periods of famine or insufficient access to nutrition is associated with linear growth and delayed puberty.[517] On the other hand, over-nutrition has been identified to result in rapid childhood growth and earlier puberty onset.[517] There is a trend in developed countries towards taller human height and earlier onset of puberty, which has occurred in parallel to higher caloric availability.[517] This is a phenomenon that is conserved across the animal kingdom.[517]

MC3R in the reproductive axis

The robust association identified between MC3R LoF mutations and delayed puberty indicates a key role of *MC3R* in the control of the human reproductive axis and development. This is supported by work from our collaborators that showed a subtle delayed puberty effect in *Mc3r* knockout mice as well as an insensitivity of mice deficient in *Mc3r* to the reproductive impact of fasting, compared to wild-type mice. This indicates that this effect is conserved from mouse to human. We further identified co-expression of a population of KISS1 neurons with *Mc3r* in the hypothalamus mouse using scRNA-seq data, and further that this co-expression is found in human through RNA-scope of a human donor hypothalamus.

This link between *MC3R* and puberty is further highlighted in females who are homozygous for mutations in POMC, who therefore lack sufficient signalling to both MC3R and MC4R. These individuals do not enter puberty and are severely obese from

early childhood.[480] When treated with setmelanotide, a melanocortin receptor agonist with more than 10 times the selectivity for MC4R than MC3R, these individuals lose weight, indicative on activation of MC4R, but remain hypogonadal suggesting in combination with evidence presented in this chapter that this is due to insufficient activation of MC3R.[518] This further highlights the potential utility of MC3R agonists in individuals with proteins upstream of MC3R in central melanocortin system to facilitate puberty onset and development. This work highlights the utility of investigation and investment in research to develop such pharmacological interventions. These effects of LoF correlate with the severity of receptor dysfunction as well as genotype – with the homozygous carrier of complete loss of function variant *MC3R* p.G240W having the most severe impact on pubertal delay.

MC3R plays a role in growth.

I further identified a robust association of *MC3R* LoF with linear growth and height. This observation, which was reported in the original papers for *Mc3r* deficient mice[490,492], with these animals having shorter body length as well as femur length. The identification of Mc3r LoF variants being associated with shorter height in childhood and adulthood, as evidenced by the association analyses in UKBB and ALSPAC, demonstrates that this effect impacts growth during development and final height reached in adulthood. This association further showed a dose-effect relationship with the degree of *MC3R* LoF being correlated by the degree of impact on height seen with carriage of a given *MC3R* variant.

The effect of *MC3R* deficiency in humans on height appears to have a greater impact on trunk length than leg length. We hypothesise that this is due to delay in puberty resulting in delay in epiphyseal fusion which allows an extended period of long growth. This is on a background of growth hormone deficiency which impedes overall growth and therefore height. Further evidence of MC3R involved in the growth hormone- IGF1 axis, as evidenced by an association of *MC3R* LoF with lower circulating IGF1 levels is consistent with previous observations of *Mc3r* deficient mice.

MC3R is involved in the accrual of lean mass

The association of MC3R LoF with lower total body lean mass suggests that MC3R is involved in the accrual of lean mass. This is consistent with previous studies in *Mc3r* deficient mice which had a low lean-to fat mass ratio[490,492]. This effect on lean mass is most marked in the homozygous *MC3R* LoF proband whose total body lean mass and lean mass percentage was markedly lower than average compared to males matched by age, ethnicity, or BMI. The probands ALM/BMI ratio is also below to threshold for sarcopenia, an involuntary loss of muscle mass. This observation suggests that MC3R agonists may further have potential utility in treatment of sarcopenia, which is characterised by low lean mass, including muscle. This suggests that MC3R agonists may have utility beyond puberty and height in a wide range of chronic disorders in which sarcopenia is found.[518] However these results need to be taken with caution that there may be confounding of height in these observations.

Phenotypic associations are identified for WT-like MC3R p.V44I

In UK Biobank we see consistent associations of MC3R p.V44I with which have a significant but small effect on phenotypes of interest, this is even though p.V44I is classed as a WT-like variant displaying 94% of WT *MC3R* cAMP response. This variant is the most common of the variant studied, being found in 10.09% of Europeans in UKBB. This disconnect between WT-like cAMP response and association with relevant phenotypes can be hypothesised to be due to several outcomes. Firstly, our studies of cAMP response of *MC3R* WT vs mutant constructs may not be sensitive enough to distinguish such subtle changes in MC3R receptor activity compared to WT MC3R. While not significant, p.V44I resulted in a cAMP response that was consistently reduced compared to the WT receptor.

Secondly, the commonality of this variant compared to others identified may allow us to identify subtle effects that we do not have enough power to identify for other WT-like variants which are found at lower allele frequencies. Thirdly, p.V44I is in strong linkage disequilibrium (r2 = 0.97) with a previously identified GWAS lead variant for age at menarche (rs3746619 located in the 5′ untranslated region of *MC3R*). This suggests that V44I may simply be tagging the potential regulatory effect of rs3746619. Finally, while cAMP is the dominant downstream signalling mechanism of MC3R, it is possible that V44I has a severe impact on an alternative mechanism that is not captured using

the cAMP assay for example an impact on beta-arrestin signalling[518]. Further studies which can assess these alternative pathways will be needed to confirm whether this is the case.

Is MC3R deficiency associated with obesity?

Small scale family studies reported co-segregation of heterozygous carriage of dysfunctional MC3R variants with severe obesity.[431,495–498] Further Mc3r knockout mice gain more weight than WT mice when on a high fat diet, although this is to less of a degree than that of Mc4r deficient animals. Further, double Mc3R and Mc4r knockout mice are more obese than single knockouts of each gene.[490]

In this study no significant association was identified for heterozygous *MC3R* loss of function with BMI and other obesity related traits in childhood and adulthood UKBB and ALSPAC, despite assessing this association at population scale. Further large scale GWAS meta-analyses have not identified any significant association of variants near or in MC3R with BMI.[499]

However, the homozygous MC3R LoF proband was severely obese, despite having no mutations in known monogenic obesity genes. Subsequent work in the O'Rahilly group in collaboration with Professor Danish Saleheen (University of Columbia) has identified 4 Pakistani individuals who are homozygous carriers of complete loss of function variants who are not obese (personal communication).

Novelty

This is the first study to systematically assess the impact on *MC3R* function of genetic variants and subsequently assess the impact on phenotype at population scale. We further identified a single homozygous carrier of a *MC3R* cLoF mutation for the first time, and therefore describe for the first time a genetic syndrome of *MC3R* deficiency in humans.

This work along with that other work outlined from collaborators proposes that MC3R plays a key role in this mechanism with MC3R relaying signals of nutritional status,

through signals of sufficient nutrition – leptin and insulin, to control the timing of puberty and rate of linear growth.

Strengths and Limitations

There are some limitations to this study that need to be considered. Firstly, while cAMP is the dominant downstream signalling pathway of MC3R, it is possible that *MC3R* variants may specifically impact some other alternative pathways downstream of MC3R activation such as beta-arrestin which were not assessed in this project. Further this was assessed in an overexpression system where MC3R variants were transient transfected into HEK293 cells, a cell line that does not express endogenous MC3R. While this means that any cAMP response identified is from the transfected MC3R it would be insightful to investigate the impact of *MC3R* variants in alternative cell lines or systems where *MC3R* is expressed at endogenous levels. It would further be interesting to assess the role of the *MC3R* mutations identified in humanised mouse models where these specific mutations are introduced, to obtain insight into the phenotypic consequences beyond what can be done in human population studies.

Our genetic analyses of heterozygous carriage of *MC3R* LoF variants was restricted to individuals of European ancestry. It is important that future efforts look to assess these variants and those common in other populations of other genetic ancestries.

We identified a single homozygous human completely deficient in *MC3R*. While most of the phenotypic findings were consistent with our observations in heterozygous carriers identified in population studies, this individual is also obese. Further homozygous carriers will need to be identified in future studies to identify if homozygous carriage of MC3R LoF variants is associated with obesity or if this is an incidental finding in this individual.

Despite some limitations described above, this project has several strengths. Firstly, we demonstrate robust association of MC3R LoF with human phenotype at population scale showing robust association of MC3R LoF with human phenotype in UK Biobank,

one of the largest population studies including almost 500,000 individuals. The ALSPAC study further allowed us to assess and confirm associations observed in UKBB and their impact throughout childhood and development. Further, as ALSPAC is a birth cohort the recruitment is unbiased and therefore likely to reflect the true effect of MC3R LoF in the UK population. The assessment of MC3R LoF variant association with 'Omics measures further reflects the largest possible sample size available for these measures and genetic data at the time of data analysis.

Conclusions

These data have allowed us to build evidence for a bifurcating model of central melanocortin signalling with nutritional cues being relayed through MC4R to the control of appetite, food intake and energy expenditure, while through MC3R these nutritional cues, indicative of sufficient energy stores and nutation, result in the promotion of linear growth and accrual of lean mass, as well as the timing of sexual maturation (**Figure 5.9)**. This work has highlighted the role of MC3R in the central melanocortin system and provides evidence for the role of MC3R in several key pathways, highlighting the potential of MC3R agonism in therapeutic innervations for conditions including delayed puberty and sarcopenia. The identification of enrichment of predicted deleterious MC3R variants in individuals who reported delayed puberty further highlights the value of screening for *MC3R* mutations in individuals who have delayed puberty.

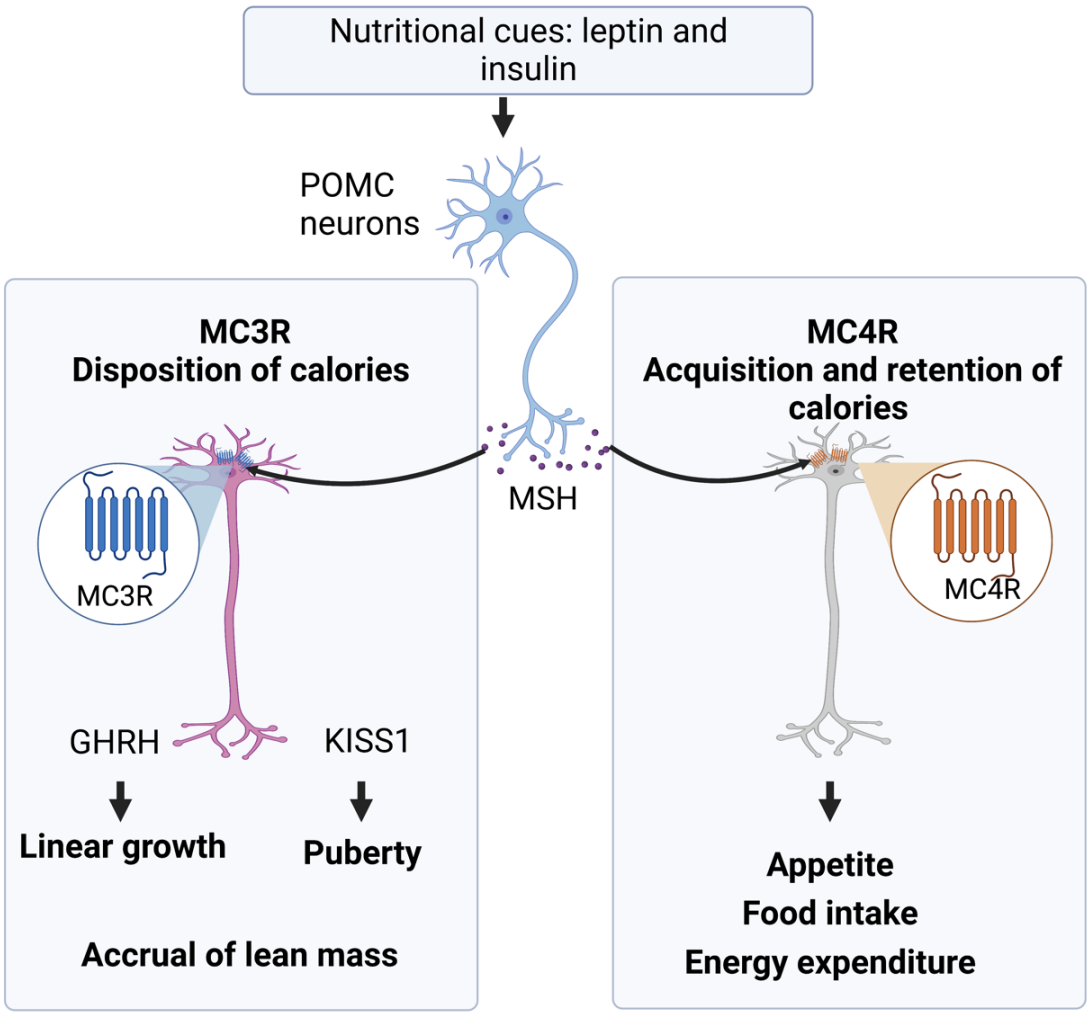


Figure 5.9 A bifurcating model of central melanocortin signalling

Figure generated using Biorender.com

Chapter 6: Concluding Discussion

6.1 Summary of findings

This thesis aimed to leverage the power of large-scale population genetic analyses coupled with tailored *in vitro* follow up to further the understanding of T2D aetiology through characterisation of metabolic sub-phenotypes.

In this concluding discussion I will summarise the overall findings of this thesis, the strengths and limitations of this work more broadly across the projects presented, possible implications of these findings, and potential future research directions.

Refined glycaemic traits intermediate to T2D have great utility in furthering the understanding of T2D aetiology

Genomics is a powerful tool to understand the aetiology of T2D, as demonstrated by the exponential increase in discovery of genetic loci associated with T2D over the last ~15 years[78,80,146,411,519], with the most recent effort identifying 338 distinct genetic signals in multi-ancestry analyses.[80] However, the molecular mechanisms by which these loci are associated to increased T2D risk is unclear in the vast majority of cases, with causal genes and mechanisms unknown. Examining the genetic architecture of glycaemic traits, and how this relates to T2D, has been shown to be useful in beginning to get insight to the molecular impact of these loci in T2D, revealing candidate genes and processes involved in T2D aetiology.[87,89,93,95,200,239,248,520] Mid-to-high throughput *in vitro* functional follow up of loci and genes of interest in T2D have predominately taken a pancreas-centric view of T2D, focusing on the role of genes identified in GWAS studies in beta-cell dysfunction[204–208,521], and fasting insulin resistance, implicating the liver.[209–211]

Studying the genetic architecture of post-prandial insulin resistance reveals loci contributing to T2D risk via this pathway

Post-prandial insulin resistance has been demonstrated to be a key pathophysiological pathway contributing to T2D onset[296], however despite evidence of a genetic contribution to this in T2D from isolated and patient populations[109,111,522], this had not been studied at scale within the general population. In Chapter 2 I demonstrated how OGTT-derived indices can be leveraged at population scale in combination with genetics to study the genetic architecture of post-prandial insulin resistance, a key pathophysiological pathway in T2D, to further understand aetiological mechanisms at T2D associated loci. A study of the genetic architecture of two OGTT measures of post-challenge insulin resistance, insulin fold change and Modified Stumvoll ISI, in

more than 55,000 individuals revealed 10 novel genetic loci, 9 of which were identified in individuals of European ancestry and 1 in individuals of non-European ancestry. This is the largest and most comprehensive study of post-challenge insulin resistance to date. Eight of these novel loci colocalised with type 2 diabetes, suggesting that these loci may be associated with T2D risk through their impact on post-challenge insulin action. This is important particularly in the context of dominant focus of the genetic architecture of insulin resistance in the fasting, and not postprandial states. Focused analyses of one of the loci, *SLC2A4,* revealed that this locus specifically associated with post-challenge specific insulin resistance, indicated by this locus being associated with elevated 2hr glucose and post-challenge insulin resistance (higher IFC, lower ISI), without associations with fasting measures. This is likely due to an impact of this locus on transcriptional regulation of *SLC2A4*, with the effect allele (rs117643180-A) being associated with reduced transcriptional activity in muscle cell lines, and the fact that this is an eQTL for reduced *SLC2A4* mRNA expression in skeletal muscle. Whilst this is a biologically obvious candidate, encoding the critical insulin-stimulated glucose transporter GLUT4 that is responsible for insulin-stimulated glucose uptake in the postprandial state in muscle and adipose tissues, it is the first time such locus has been described for post-challenge insulin traits. Further this study revealed other loci where currently the exact mechanism of action is not clear and warrant future functional investigation. This includes loci prioritised using an integrated GWAS approach, such as *ERAP2*, which colocalises for an eQTL for *LNPEP* (also known as *IRAP*), another key player in insulin-stimulated glucose uptake.

Genetically informed functional follow up reveals novel regulators of GLUT4 trafficking

In the post-prandial state GLUT4 is the key player in insulin stimulated glucose uptake in muscle and adipose tissues, responsible for the majority of postprandial glucose disposal[273]. Dysregulation of this pathway has further been demonstrated to result in postprandial insulin resistance and T2D[274,300,303].

In Chapter 3 I went from genome to function, characterising 36 genes located at post-challenge insulin resistance loci (Chapter 2) for their function in insulin-stimulated GLUT4 trafficking in a cellular system (**Figure 6.1**). This approach highlighted the power of genetically informed approaches to prioritise genes for functional follow up. This assessment in 3T3-L1 adipocytes allowed the identification of 14 out of the 36

genes as being implicated in GLUT4 trafficking, 9 of which have not been described before. This demonstrates the utility of combining genetic analyses of traits intermediate to disease, in this case post-challenge insulin resistance in T2D, and how the results of such analyses can be used to prioritise candidates with a high prior for functional investigation into their potential role in biologically relevant pathways. Literature review revealed diverse functional roles of these genes, and future detailed mechanistic follow up will be needed to understand the exact mechanisms by which there are implicated in GLUT4 trafficking.

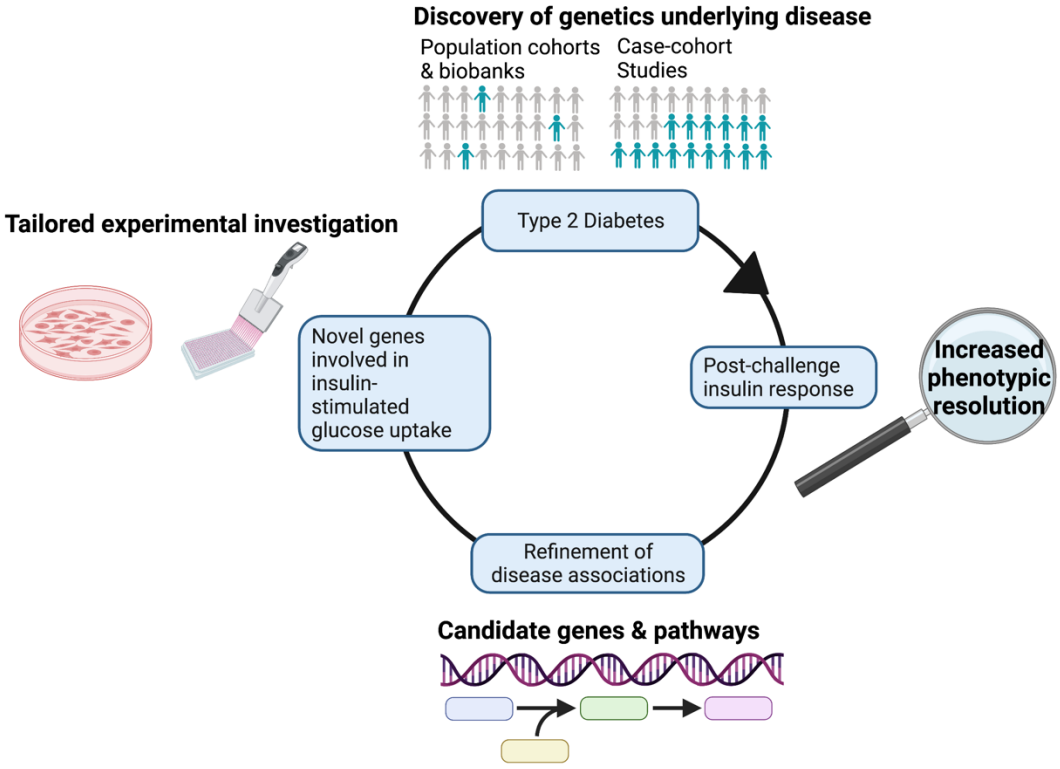


Figure 6.1. Integration of population genetics and *in vitro* follow up to identify novel candidates involved in post-prandial insulin action and T2D
Figure generated using biorender.com

Genetic subtypes of T2D have distinct metabolomic signatures, implicating specific aetiological pathways.
Genetic subtypes for T2D have been proposed to cluster patients with T2D and identify those with differences in disease progression and contributions of different aetiologies.[134] Whilst the majority of patients with T2D do not fit one single subtype when considering such genetic or phenotypic subtypes of disease, these are useful to gain insight into the relative contribution of different pathways to disease onset.[134,137,372]

Several metabolomics measures have been proposed to be useful biomarkers to identify individuals who are at risk of developing T2D (Chapter 4 literature review, and reviewed in[523]). However, such studies have, generally, been small scale and relied on targeted profiling of metabolites. This has resulted in the levels of hundreds of metabolites to be associated with T2D incidence and proposed as candidate biomarkers, but the reproducibility of many of these observational associations, outside of the most robust associations, is poor across studies.

In Chapter 4 I examined the association of plasma metabolomics measures at baseline with incidence of T2D and demonstrated how confounded observational associations can be refined through the integration of genetics, using PGS for T2D and T2D endotypes. This study represents one of the largest studies into the metabolomic signatures of T2D genetic risk, both in terms of the sample size used and the number of metabolites assessed. I further demonstrated a systematic framework to use genetics to refine confounded observational associations of metabolites with T2D incidence. This revealed that the refined list of metabolomic associations with T2D risk mirror in direction the association of these metabolites with T2D incidence. In addition, the study of genetic T2D endotypes revealed distinct metabolomic signatures, suggesting biomarkers for T2D contributed to by specific aetiological pathways. Using a causal mediation framework, as well as MR and genetic colocalisation for a single example, I demonstrated that the suggestive causal associations identified were likely because of insulin resistance, rather than these metabolites being direct causal mediators in T2D.

A functional reverse genetic approach allowed identification of the role of *MC3R* in humans.

MC3R is a key player in the central leptin-melanocortin pathway, however the exact downstream effects of this gene in humans were not well understood. Small scale family studies suggested that heterozygous loss of function of *MC3R*, like closely related *MC4R*, resulted in obesity and therefore it was thought *MC3R* may play a role in the maintenance of energy balance.[431,495–498] However, such effects were less clear in knockout mouse studies[490], and common GWAS studies of BMI and related adiposity traits had not identified genetic loci near *MC3R* to be associated with these traits. Therefore, it was hypothesised that systematic interrogation of *MC3R* function

using coding variants found in human population cohorts may allow greater insight into gene function.

In Chapter 5, through the *in vitro* identification and prioritisation of truly damaging variants in *MC3R*, I identified an association between *MC3R* loss of function with timing of puberty, accrual of lean mass and linear growth in UK biobank, which is supported by detailed *in vitro* and *in vivo* follow up from collaborators. This work further disproved, at population scale, the association of heterozygous *MC3R* loss of function with obesity and adiposity related traits. The *in vitro* identification of the true functional impact of *MC3R* variants that were predicted to be damaging demonstrated the benefit of *in vitro* characterisation to refine variants considered in targeted genetic analyses. It further demonstrated how focused genetic characterisation of targets of interest has the potential to yield important and novel insights into gene function in combination with detailed experimental work.

6.2 Strengths and Limitations

Strengths

Large-scale investigation of the genetic contribution to subphenotypes intermediate to T2D reveals insights into disease aetiology

In chapters 2 and 4, I demonstrated that the integration traits intermediate to T2D with genetics can aid further understanding of disease aetiology. Intermediate traits act as layers between genetic variant and disease that can allow further understanding of possible mechanisms by which a particular locus is associated with disease – as was done in chapter 2 with post-challenge insulin resistance in T2D, a related physiological process to the disease of interest. This is particularly useful when diseases are heterogeneous, as seen with T2D, with pleiotropic associations across traits at a given locus able to reveal phenotypic connections.[524] Similarly, integration of molecular trait data such as mRNA expression (via eQTLS, Chapter 2) and metabolites (Chapter 4), further allows unpicking of candidate molecular mechanisms at loci of interest.

Further, the integrative GWAS approach presented in Chapter 2 demonstrates how the integration of intermediate traits can aid refinement of genetic loci associated with phenotypes best represented by multi-trait signatures, even at suggestive significance thresholds. Such methods have been demonstrated previously to successfully prioritise loci associated with insulin resistance more broadly, which is not solely explained by a single trait, such as fasting insulin.[257] These approaches allow the prioritisation of genetic loci that have a cumulative genetic signature across multiple traits that represent the multifactorial phenotype of interest. In Chapter 2 this successfully allowed prioritisation of genetic loci below genome-wide significance that have a post-prandial specific glycaemic signature, including overlap with skeletal muscle eQTLs, indicative of genomic regulatory role of these loci in a key tissue in postprandial insulin action.

Expanding the understanding of disease associated loci, not just sample size

The characterisation the genetic architecture of metabolic subphenotypes intermediate to disease has highlighted utility in the further detailed characterisation of disease-associated loci, as opposed to further expansion of complex disease GWAS. The latter increases the number of loci discovered, but without an attempt to close the knowledge gap in the molecular mechanisms behind these associations, there will be little gain in aetiological understanding. In this thesis I have demonstrated in Chapter 2 how such

detailed characterisation using post-challenge insulin traits can be used to begin to unpick the biological mechanisms behind these disease associations.

Integration of in vitro characterisation with genetic analyses

Interdisciplinary approaches integrating large-scale genetic analyses and *in vitro* follow up have been shown to be useful in furthering the interpretability and utility of GWAS results (reviewed in [525–528]).

In this thesis I show the utility of this interdisciplinary approach form two angles. Firstly, in Chapter 3 I demonstrate how genetic association studies can act as a powerful source of gene prioritisation for functional assessment in a pathway of interest – in this case genes at loci associated with post-challenge insulin resistance and their role in GLUT4 trafficking. In Chapter 5 I show how *in vitro* functional characterisation and refinement of genetic analyses to include only truly damaging mutations allows identification of phenotypic associations without "dilution" of association tests with variants that are predicted to be damaging *in silico* despite this not truly being the case. This, along with subsequent *in vitro* and *in vivo* functional characterisation of *MC3R* by collaborators was essential in determining the fine-scale mechanisms of *MC3R* in humans.

A framework to infer causality of 'omics measures in disease

In Chapter 4 I demonstrated how integration of genetics, metabolomics, and information on T2D incidence has utility in furthering the understanding of T2D aetiology and inferring causality of circulating metabolites directly in T2D. The framework built in this work can be readily applied to assess the molecular signatures of other diseases and/or the associations of other 'omics measures.

Causal mediation analyses using natural indirect effect models allowed the use of individual data to determine what metabolites mediate T2D genetic risk to T2D incidence. The advantage of using individual level data means that PGS can be considered across the effects of all SNPs in totality, unlike MR methods where SNPs must meet stringent inclusion criteria. Further the use of individual level data allows sensitivity analyses adjusting for possible confounding factors e.g., measures of insulin resistance as demonstrated in Chapter 4.

However, MR also has several advantages. Two sample MR allows the use of GWAS summary statistics meaning that the largest possible sample size can be utilised for traits of interest and is not constrained by traits available in a single study. In Chapter 4 this allowed use of GWAS for fasting insulin to assess the causality of glycine on insulin resistance, a trait that is not available in the EPIC Norfolk study, and further use of summary statistics from a sample size several orders of magnitude above that of the EPIC-Norfolk T2D case-cohort.

Insight into GLUT4 total and surface levels in the same cell

In Chapter 3 the GLUT4 trafficking assay developed by the Fazakerley lab has a number of strengths to studying this process. Most importantly this approach allows the measurement of both surface levels of GLUT4 and total levels of GLUT4 within the same cell thus allowing disentanglement of effects on GLUT4 trafficking vs regulation. Secondly the use of secondary antibodies conjugated to fluorophores of different wavelengths allows parallel assessment of multiple proteins of interest. Further the 96 well format of this assay means that it is relatively high throughput in terms of the number of genes assessed in parallel.

Integration of true impact of variant function into genetic association analyses

As mentioned above, one of the strengths of the results presented in Chapter 5 for *MC3R* includes the integration of information of the true functional impact of coding *MC3R* variants on function. This allowed the refinement of variants that were predicted to be damaging with ~55% of variants assayed, all of which were predicted to be damaging by SIFT[276] or Polyphen[277], having no significant impact on *MC3R* function. The refinement of genetic association analyses to only include truly damaging variants allows more specific and refined analyses of *MC3R* loss of function on traits of interest.

Limitations and considerations

General considerations of a molecular view in studying metabolic diseases.

This thesis and other work more broadly in the field of genetics views metabolic diseases, in this case T2D, from a purely biomedical perspective. These investigations therefore focus on the molecular mechanisms most proximal to disease. Such approaches are incredibly useful in gaining insight into disease aetiology and mechanisms that contribute to disease risk, as demonstrated in this thesis, and thus inform avenues for treatment. However, one also needs to consider such studies in a

wider societal context taking into consideration the other determinants of health. It is known that socioeconomic and behavioural factors play a key role in disease risk including diet, physical activity, accessibility of healthcare, and the built environment[529–531] and these likely have a greater overall impact on common metabolic disease risk, at a population level, downstream of genetic influences.

Considerations for large-scale GWAS studies

Possibility of false positive findings

Whilst the stringent genome-wide significance threshold ($P < 5x10^{-8}$) is widely accepted for in GWAS to define significant findings, it is possible that some of the associations that have been reported in Chapter 2 in this study are false positive.

The integrated GWAS examination of post-prandial insulin resistance, and the suggestive loci identified (outlined in Chapter 2) also needs to be considered in terms of potential for false positive results. Whilst several of the loci highlighted are biologically plausible when considering candidate genes implicated, future work will be needed to see if these loci meet more stringent genome-wide significance with increased sample size in future efforts. Further replication in subsequent, efforts of larger scale including independent cohorts will be needed to prove the stability of these associations. There have been several suggestions for the research community to consider using more stringent significance thresholds considering for example increased density of imputation panels and adoption of sequencing methods meaning more and increasingly rare variants are assessed.[253,532] This is therefore something that may be reconsidered in the future.

This also needs to be considered for the PGS used in Chapter 4 for T2D and T2D endotypes. Whilst these PGS were restricted to loci that met stringent genome-wide significance threshold ($P<5x10^{-8}$) in effort to ensure specificity for T2D, it is possible that some of the loci included are false positive findings, which could independently influence the association of these PGS with metabolites of interest. Whilst not considered in the current analysis, leave-one-out analyses for the association of the PGS with metabolites, excluding individual variants from the PGS could be used to test the stability of the metabolite signatures identified and ensure that associations are not driven by a single variant.

GWAS does not allow the identification of all genetic determinants of complex traits

The genetic variants examined across this thesis were predominately identified by imputation of genome-wide genotype arrays. Whilst increased imputation density now facilitates the reliable detection of low-frequency variants, as demonstrated in Chapter 5 with MC3R, such technologies do not allow the detection of all variation across the genome – namely rare and ultra-rare variants present in < 0.5% of the population.[102] Further genotyping arrays do not allow complete coverage of the coding genome – exons – which have been shown in recent biobank-scale WES efforts such as in UKBB to capture coding variants which have a significant impact on gene function and phenotype.[253,532] Secondly, structural variants and copy number variants (CNVs) are often not considered, or poorly captured by SNP array and current NGS sequencing methods used in WES and WGS studies.[533]

Such variants are likely to make up this missing heritability not explained by GWAS and WES studies, with increased sample size and ability to detect variants of increasing rarity in future population genomics studies.[533]

GWAS studies remain dominated by samples of European ancestry and do not reflect global diversity

Population genetic work presented in this thesis has predominantly relied on cohorts of European ancestry, meaning that it is not yet known whether the findings are generalisable to the global population overall. Whilst three studies of non-European ancestry – 1 of East Asian and 2 of Hispanic American ancestry – were included in the genetic discovery of post-challenge insulin resistance in Chapter 2, these studies made up only 3% of the overall sample size and therefore were underpowered for ancestry-specific analyses. Future efforts to characterise dynamic glycaemic traits outside the European population is of great importance and will allow detection of possible ancestry specific effects as well as increase overall power for genetic discovery across populations. From a broader ethical viewpoint, it is further paramount to ensure the generalisability of these associations across ancestries, particularly in the context of identifying candidates which may be investigated as drug targets with the potential of differential adverse effects across ancestries.

Increased diversity of genetic discovery will be beneficial for example, in understanding the relationship of the *SLC2A4* locus with T2D and post-challenge insulin resistance. As discussed in Chapter 2 signals do not colocalise for these traits in cohorts of European ancestry but there is some evidence that investigating this signal further in studies of East Asian ancestry populations, with differing allele frequencies apparent in these ancestry groups likely due to differences in LD. These differences in LD are also beneficial when conducting multi-ancestry analyses allowing improved fine-mapping and identification of candidate causal variants.

Further in Chapter 4, the metabolomic signatures of T2D were assessed in the EPIC-Norfolk study, a cohort of individuals of European ancestry, with MR analyses utilising instruments derived from GWAS of European ancestry. When further mQTL studies are available from other ancestries, it would be useful to apply this framework to infer the effects are generalisable over wider populations. This is particularly important in the context of using this framework to gain insight into the effects of drug targets and whether this differs across populations.[533]

Causal inference from genetic studies

Use of the proposed causal mediation framework to infer causality of exposure on outcome

Causal mediation analyses, as presented in Chapter 4, has a number of strengths including the use of individual level data. However, the power of such analyses therefore relies on the sample size for the study used for such analyses. The EPIC-Norfolk T2D case-cohort has a relatively modest sample size meaning that not all PGS-metabolite associations detected in the larger quasi-random subcohort were detected at significance in this subcohort. This restricted the number of metabolites for which causality could be assessed. Further due to the individual-level data required by such method, analyses are restricted to the traits available in a given study. This is in contract with MR approaches which can utilise publicly available GWAS summary statistics with the largest sample size and likely a wider range of traits than will be available in any given study.

Inferences of causality by MR are reliant of suitable genetic instruments

Causal interference using MR approaches has a number of assumptions, including most critically that the genetic effects are unconfounded from exposure to outcome.

Whilst efforts were taken to ensure that bias was minimised, for example, using BMI unadjusted summary statistics for T2D and insulin resistance related phenotypes examined in MR analyses presented in Chapter 4, it is possible that other confounders and sources of bias were not accounted for.

Further, the scope of MR analyses presented in Chapter 4 was hindered by insufficient number or lack of suitable genetic instruments for other metabolites. Therefore, it is possible that several causal associations from metabolite to T2D risk were missed in these analyses, and future work should be done to examine the potential causality of these associations when larger sample size are available in GWAS for these metabolites.

Challenges in assigning "cis acting" genetic instruments for metabolites

In MR analyses of expression or protein QTLs, genetic instruments can be refined to those that are only in the *cis* region of the target gene of interest to ensure specificity to the gene of interest.[533] I considered doing similar in MR analyses for glycine to T2D presented in Chapter 4, but this has several challenges for metabolite QTLs (mQTLs). Whilst one can identify loci that are near genes encoding enzymes known to be involved in metabolite metabolism, in this case glycine, for example, it is not known for certain whether these are the truly causal genes at a locus, even with higher biological prior of being involved in a related pathway. Further the effects of metabolite catabolism vs synthesis may be different across candidate pathway-specific "*cis* acting" mQTLs.

Colocalisation analyses assume a single causal variant for trait pairs at a given locus

In Chapter 2 I demonstrated how pairwise[177] and multi-trait colocalisation[176] can use used to identify shared genetic architecture between sets of traits at loci of interest. However, both methods need to be considered considering the core assumption that there is a single causal variant at a locus of interest effecting both traits. It is possible that shared effects of a locus on two traits with 2 independent variants, that effect a single gene, may be missed with such approach. Using an alternative method that allows consideration of multiple causal variants at a given locus, such as recently developed method SuSiE, would have allowed testing of this assumption.[533]

Phenotype considerations

In Chapter 2, post-prandial insulin resistance was estimated using dynamic measures of insulin and glucose during an OGTT. As discussed previously, it is not practical to obtain hyperinsulinemic-euglycemic clamps coupled with stable isotope labelled glucose at sufficient scale for GWAS studies – the gold standard method to assess post-prandial insulin sensitivity. These OGTT-derived measures allow approximation of such measures in genetic studies; however, it is possible these may capture other effects such as insulin secretion which needs to be taken into consideration when interpreting results of genetic analyses of these traits.

In addition, in Chapter 4, untargeted plasma metabolomics measures were assessed to identify molecular signatures of T2D genetic risk and that of T2D endotypes. Plasma is easy to sample and convenient when considering potential applications for metabolites as biomarkers for disease. However, it is possible that the levels and relative changes in metabolite levels compared between healthy and disease states, or those at high or low genetic risk, may not be reflective of changes in disease relevant tissues – for example in T2D muscle, adipose, liver and pancreas. For example, elevated levels of a metabolite in plasma may be explained by increased flux from a tissue into blood, and thus reduced levels in disease-relevant tissue, or reduced downstream reaction rates, or increased upstream pathway rates.[534] It is not possible to determine such changes from metabolite levels alone and therefore it is important that this is considered in interpretation of association results.

6.3 Implications

The work presented in this thesis has allowed further aetiological understanding for T2D through the interrogation of molecular subphenotypes and intermediate traits. Assessment of the genetic architecture of post-challenge insulin resistance in Chapter 2 has highlighted an importance of post-prandial insulin resistance in the genetic contribution to T2D risk in the general population. This further highlights the utility of intermediate traits, and pathways in T2D to further understanding of T2D genetic aetiology and importance of such pathways in pathogenesis. This study emphasises that there is benefit in examining the "layers" below disease association using refined, less commonly measured traits, rather than simply expanding the number of loci identified for disease traits with little gain in understanding of biological mechanism. In Chapter 4 I further demonstrated that genetic scores for T2D endotypes are associated with distinct metabolomic signatures, highlighting that such subphenotypes are detectable and distinguishable using 'omics measures that are candidate circulating biomarkers, which also overlap with those associated with T2D incidence.

The *in vitro* work presented in this thesis highlights the utility of functional characterisation of variants and/or genes to both inform and follow up genetic association analyses. Such characterisation allowed tailored and refined analyses of truly damaging MC3R coding variants in Chapter 5, going beyond standard *in silico* prediction when interpreting the impact of such coding variant. In Chapter 3 the characterisation of candidate genes from post-challenge insulin resistance associated loci allowed the identification of novel candidate players in GLUT4 trafficking. This highlights the advantage of combining genetic discovery of refined traits, systematic gene selection workflows, with pathway specific follow up to gain increased understanding about disease mechanism and related biological pathways.

The methodological approaches used in this thesis provide analytical frameworks that can be applied to other diseases and traits of interest when studying complex diseases. For instance, in Chapter 5 the characterisation of *MC3R* function using functional characterisation of variants of interest followed by population-scale genetic analyses highlights how such interdisciplinary approaches can allow great progress in understanding gene function in humans. In Chapter 2 the integrated GWAS approach that was previously used to gain insight into insulin resistance[535], emphasises that, with caution, that biologically relevant loci can be identified below genome-wide significance

allowing increased insight into the genetic architecture of post-challenge insulin resistance even with modest sample sizes compared to large scale population biobanks. In Chapter 4, the systematic approach, utilising several analytical approaches, to identify candidate causal metabolites mediating T2D genetic risk to T2D incidence, can readily be applied in other cohorts to assess the causality of other 'omics measures in other diseases.

There are potential future clinical implications stemming from this work with implications for directions of research into drug development and pharmacological interventions. Chapters 2 and 3 further highlighted the importance of post-challenge insulin resistance in contributing to T2D onset. This highlighted several new candidates playing a role in insulin-stimulated glucose uptake in the post-prandial state, which could be explored as a potential new class of drug targets for T2D, with this pathway not targeted by current drugs available for T2D[4]. For MC3R (chapter 5), the identification of a role in puberty timing suggests that activation of MC3R could be tractable as a treatment for delayed puberty. In addition, the role of MC3R in accrual of lean mass suggests that promotion of MC3R receptor activity could be used in conditions where there is loss of muscle mass such as sarcopenia.[536]

Finally, the identification of robustly associated metabolites suggests that panels of biomarkers for T2D subtypes could be generated, which in the future, in combination with future work, could potentially inform different treatment avenues.

6.4 Future directions

The work presented in this thesis has opened several avenues for future work. Firstly, the genetic discovery of post-challenge insulin resistance should be expanded to include more diverse ancestries within the study, and future discovery efforts should also consider assessing the potential for sex-specific genetic contribution to post-challenge insulin resistance in T2D. Detailed functional characterisation of the loci identified could also be useful in pinpointing the causal gene and mechanism of action at most of these loci which is currently not known.

The identification of 9 novel genes implicated in GLUT4 trafficking provides a starting point for detailed follow up work to fully understand how these are implicated in this pathway. This could include assessment of the effects of these genes in muscle cell lines as well as assessment of the interactome of these proteins to begin to elucidate mechanistic links to GLUT4 and its function.

Larger sample sizes are needed to detect further genetic loci associated with T2D subphenotypes and intermediate traits

Whilst the genetic discovery presented in Chapter 2 represents the largest sample size for post-challenge insulin traits to date, including more than 55,000 individuals, it is likely that these were insufficiently powered to detect all genetic variation in these traits. Further in Chapter 4, the updated meta-analysis of glycine represents the largest sample size for this trait, but similar challenges are likely as described for post-challenge insulin traits. Both efforts are still swamped by large scale biobank efforts such as in UK Biobank for other traits which includes more than 500,000 individuals with genetic and phenotypic information. Such large-scale discovery efforts include measures that are more readily available and easier, cheaper, and/or more convenient to characterise – such as BMI[499] and height[105], and clinical blood biochemistry traits measured in the unfasted state.[537] Further meta-analyses for such traits mean that genetic discovery efforts for such traits now include more than 1 million individuals, with the most recent and largest genetic discovery effort for height including approximately 5.4 million individuals[105].

Future work including larger sample sizes will allow the identification of additional genetic loci as well as assessment of the stability of the loci currently identified.

Updated efforts including larger sample sizes also mean that updated analyses including new PGS, derived from studies with larger sample size for T2D and T2D sub phenotypes could be run to assess the metabolomic signatures of T2D risk and see how stable these are with updated PGS.[80,538]

Exploring the influence of sex and ancestry on the genetic architecture of T2D subphenotypes

Previous efforts have demonstrated that there are some sex specific effects on T2D and insulin resistance[146,214], however such differences were not explored in this study. Future work that examines the sex specific genetic architecture of post-challenge insulin resistance would potentially allow the identification of such effects.

As discussed above, future work should also assess possible ancestry specific effects of loci on T2D subphenotypes, to allow identification of the generalisability of the genetic associations presented in this thesis to the wider global population. Including additional more diverse samples will further allow increased sample size and thus increased power to detect multi-ancestry as well as ancestry-specific associations. Recent multi-ancestry analyses for T2D identified and glycaemic traits identified several novel multi-ancestry loci as well as those specific to individual ancestry groups[80,87]. Further increased specificity and resolution of fine-mapping in multi-ancestry analyses will allow more refinement of candidate causal variants at loci.[533]

Detailed molecular characterisation of novel candidate genes and their role in GLUT4 trafficking

In Chapter 3, 9 novel candidate genes influencing GLUT4 trafficking were described. Future work is needed to test the effect of depletion of these genes ideally in human-derived skeletal muscle cell line in which GLUT4 trafficking can be reliably measured. Human primary myotubes may be worthwhile exploring, being more reflective of the endogenous state compared to immortalised cell lines.[539–541] Further mechanistic characterisation of these genes is needed to understand the exact mechanisms by which these implicate GLUT4 trafficking, and how directly these genes are involved in the GLUT4 trafficking pathway. Such analyses could examine for example the intracellular localisation of candidate genes and GLUT4 and how these change upon insulin stimulation, for example.

Integrated network view of intermediate traits in T2D,

In Chapter 2 the focus on post-challenge insulin resistance represents just one pathway influencing T2D pathogenesis and focused on mechanisms in skeletal muscle and adipose tissues. Integration across analyses of intermediate traits and tissues representing other pathways e.g., insulin secretion from pancreatic beta cells and liver insulin action, could allow a broader network view of the genetic contribution of T2D, potentially allowing increased aetiological insight through joint analyses across pathways. Such future work could include the integration of additional tissue specific 'omics measures, down to a single cell level, which would allow potential identification of common cell types implicated across tissues.[371,542–544]

Similarly, integration of a "whole organism" view of metabolite flux could gain insight into how associations of differential metabolite levels identified with T2D in plasma, relate to effects in disease relevant tissues.

Conclusion

Overall, this thesis has demonstrated the power of an interdisciplinary approach to further the understanding of the aetiology of T2D and its metabolic sub-types. The work presented in this thesis contributes to advanced understanding of the genetic contribution to post-prandial insulin resistance in T2D, molecular insight into T2D subtypes, and the role of *MC3R* in humans. The investigations presented in this thesis highlight several candidate genes that warrant further mechanistic investigation for their utility in treatment of T2D and related disorders.

References:

1. Sun, H. *et al.* IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res. Clin. Pract.* **183**, (2022).
2. World Obesity. World Obesity Atlas 2022 | World Obesity Federation. https://www.worldobesity.org/resources/resource-library/world-obesity-atlas-2022 (2022).
3. National Insitute of Clinical and Health Excellence. Overview | Type 2 diabetes in adults: management | Guidance | NICE. *Nice* 1–44 (2015).
4. DeFronzo, R. A. *et al.* Type 2 diabetes mellitus. *Nat. Rev. Dis. Prim. 2015 11* **1**, 1–22 (2015).
5. National Institute of Health and Care Excellence. Obesity: Indentification, Assessment and Management, Clinical Guideline [CG189]. 63 (2014).
6. Samuel, V. T. & Shulman, G. I. The pathogenesis of insulin resistance: integrating signaling pathways and substrate flux. *J. Clin. Invest.* **126**, 12–22 (2016).
7. Batista, T. M., Haider, N. & Kahn, C. R. Defining the underlying defect in insulin action in type 2 diabetes. *Diabetol. 2021 645* **64**, 994–1006 (2021).
8. Brüning, J. C., Winnay, J., Cheatham, B. & Kahn, C. R. Differential signaling by insulin receptor substrate 1 (IRS-1) and IRS-2 in IRS-1-deficient cells. *Mol. Cell. Biol.* **17**, 1513–1521 (1997).
9. Sun, X. J. *et al.* Structure of the insulin receptor substrate IRS-1 defines a unique signal transduction protein. *Nature* **352**, (1991).
10. Backer, J. M. *et al.* Phosphatidylinositol 3′-kinase is activated by association with IRS-1 during insulin stimulation. *EMBO J.* **11**, 3469–3479 (1992).
11. Sakamoto, K. & Holman, G. D. Emerging role for AS160/TBC1D4 and TBC1D1 in the regulation of GLUT4 traffic. *Am. J. Physiol. Endocrinol. Metab.* **295**, (2008).
12. Wan, M. *et al.* A noncanonical, GSK3-independent pathway controls postprandial hepatic glycogen deposition. *Cell Metab.* **18**, 99–105 (2013).
13. Ros, S., García-Rocha, M., Domínguez, J., Ferrer, J. C. & Guinovart, J. J. Control of Liver Glycogen Synthase Activity and Intracellular Distribution by Phosphorylation. *J. Biol. Chem.* **284**, 6370–6378 (2009).
14. Yoon, M. S. & Choi, C. S. The role of amino acid-induced mammalian target of

rapamycin complex 1(mTORC1) signaling in insulin resistance. *Exp. Mol. Med. 2016 481* **48**, e201–e201 (2016).

15. Lee, S. & Dong, H. H. FoxO integration of insulin signaling with glucose and lipid metabolism. *J. Endocrinol.* **233**, R67 (2017).
16. Ricoult, S. J. H. & Manning, B. D. The multifaceted role of mTORC1 in the control of lipid metabolism. *EMBO Rep.* **14**, 242 (2013).
17. Franke, T. F., Hornik, C. P., Segev, L., Shostak, G. A. & Sugimoto, C. PI3K/Akt and apoptosis: size matters. *Oncogene 2003 2256* **22**, 8983–8998 (2003).
18. Rui, L. Energy Metabolism in the Liver. *Compr. Physiol.* **4**, 177 (2014).
19. Han, H. S., Kang, G., Kim, J. S., Choi, B. H. & Koo, S. H. Regulation of glucose metabolism from a liver-centric perspective. *Exp. Mol. Med. 2016 483* **48**, e218–e218 (2016).
20. Puigserver, P. *et al.* Insulin-regulated hepatic gluconeogenesis through FOXO1–PGC-1α interaction. *Nat. 2003 4236939* **423**, 550–555 (2003).
21. Nakae, J. *et al.* Regulation of insulin action and pancreatic β-cell function by mutated alleles of the gene encoding forkhead transcription factor Foxo1. *Nat. Genet. 2002 322* **32**, 245–253 (2002).
22. Brunet, A. *et al.* Akt Promotes Cell Survival by Phosphorylating and Inhibiting a Forkhead Transcription Factor. *Cell* **96**, 857–868 (1999).
23. Titchenell, P. M. *et al.* Direct Hepatocyte Insulin Signaling Is Required for Lipogenesis but Is Dispensable for the Suppression of Glucose Production. *Cell Metab.* **23**, 1154–1166 (2016).
24. Zhang, X., Yang, S., Chen, J. & Su, Z. Unraveling the regulation of hepatic gluconeogenesis. *Frontiers in Endocrinology* vol. 10 802 (2019).
25. Duckworth, W. C., Bennett, R. G. & Hamel, F. G. Insulin Degradation: Progress and Potential. *Endocr. Rev.* **19**, 608–624 (1998).
26. Najjar, S. M. & Perdomo, G. Hepatic Insulin Clearance: Mechanism and Physiology. *Physiology (Bethesda).* **34**, 198–215 (2019).
27. Sylow, L., Tokarz, V. L., Richter, E. A. & Klip, A. The many actions of insulin in skeletal muscle, the paramount tissue determining glycemia. *Cell Metab.* **33**, 758–780 (2021).
28. DeFronzo, R. A. & Tripathy, D. Skeletal Muscle Insulin Resistance Is the Primary Defect in Type 2 Diabetes. *Diabetes Care* **32**, S157–S163 (2009).
29. Richter, E. A. & Hargreaves, M. Exercise, GLUT4, and skeletal muscle glucose uptake. *Physiol. Rev.* **93**, 993–1017 (2013).

30. Lewis, G. F., Carpentier, A. C., Pereira, S., Hahn, M. & Giacca, A. Direct and indirect control of hepatic glucose production by insulin. *Cell Metab.* **33**, 709–720 (2021).
31. Jensen, J., Rustad, P. I., Kolnes, A. J. & Lai, Y. C. The Role of Skeletal Muscle Glycogen Breakdown for Regulation of Insulin Sensitivity by Exercise. *Front. Physiol.* **2**, (2011).
32. Szablewski, L. Glucose transporters in healthy heart and in cardiac disease. *Int. J. Cardiol.* **230**, 70–75 (2017).
33. Tran, D. H. & Wang, Z. V. Glucose Metabolism in Cardiac Hypertrophy and Heart Failure. *J. Am. Heart Assoc.* **8**, (2019).
34. Fischer, Y. *et al.* Insulin-induced Recruitment of Glucose Transporter 4 (GLUT4) and GLUT1 in Isolated Rat Cardiac Myocytes: EVIDENCE OF THE EXISTENCE OF DIFFERENT INTRACELLULAR GLUT4 VESICLE POPULATIONS. *J. Biol. Chem.* **272**, 7085–7092 (1997).
35. Shulman, G. I. *et al.* Quantitation of muscle glycogen synthesis in normal subjects and subjects with non-insulin-dependent diabetes by 13C nuclear magnetic resonance spectroscopy. *N. Engl. J. Med.* **322**, 223–228 (1990).
36. Smith, U. & Kahn, B. B. Adipose tissue regulates insulin sensitivity: role of adipogenesis, de novo lipogenesis and novel lipids. *J. Intern. Med.* **280**, 465–475 (2016).
37. Flatt, J. P. & Ball, E. G. Studies on the Metabolism of Adipose Tissue. *J. Biol. Chem.* **239**, 675–685 (1964).
38. Herman, M. A. *et al.* A novel ChREBP isoform in adipose tissue regulates systemic glucose metabolism. *Nature* **484**, 333–338 (2012).
39. Ruud, J., Steculorum, S. M. & Bruning, J. C. Neuronal control of peripheral insulin sensitivity and glucose metabolism. *Nat. Commun.* **8**, (2017).
40. Woods, S. C., Lotter, E. C., McKay, L. D. & Porte, D. Chronic intracerebroventricular infusion of insulin reduces food intake and body weight of baboons. *Nat. 1979 2825738* **282**, 503–505 (1979).
41. Milstein, J. L. & Ferris, H. A. The brain as an insulin-sensitive metabolic organ. *Mol. Metab.* **52**, 101234 (2021).
42. Garfield, A. S., Lam, D. D., Marston, O. J., Przydzial, M. J. & Heisler, L. K. Role of central melanocortin pathways in energy homeostasis. *Trends in Endocrinology and Metabolism* vol. 20 203–215 (2009).
43. Ashcroft, F. M. & Rorsman, P. Diabetes mellitus and the β-cell: the Last Ten

Years. *Cell* **148**, 1160 (2012).
44. Baron, A. D. Hemodynamic actions of insulin. *American Journal of Physiology - Endocrinology and Metabolism* vol. 267 (1994).
45. Artunc, F. *et al.* The impact of insulin resistance on the kidney and vasculature. *Nat. Rev. Nephrol. 2016 1212* **12**, 721–737 (2016).
46. Pina, A. F. *et al.* Insulin: Trigger and Target of Renal Functions. *Front. Cell Dev. Biol.* **8**, 519 (2020).
47. DeFronzo, R. A. & Tripathy, D. Skeletal Muscle Insulin Resistance Is the Primary Defect in Type 2 Diabetes. *Diabetes Care* **32**, S157 (2009).
48. Kleemann, R. *et al.* Time-Resolved and Tissue-Specific Systems Analysis of the Pathogenesis of Insulin Resistance. *PLoS One* **5**, e8817 (2010).
49. Turner, N. *et al.* Distinct patterns of tissue-specific lipid accumulation during the induction of insulin resistance in mice by high-fat feeding. *Diabetologia* **56**, 1638–1648 (2013).
50. Burchfield, J. G. *et al.* High dietary fat and sucrose results in an extensive and time-dependent deterioration in health of multiple physiological systems in mice. *J. Biol. Chem.* **293**, 5731–5745 (2018).
51. Fazakerley, D. J., Krycer, J. R., Kearney, A. L., Hocking, S. L. & James, D. E. Muscle and adipose tissue insulin resistance: malady without mechanism? *J. Lipid Res.* **60**, 1720–1732 (2019).
52. Gastaldelli, A., Gaggini, M. & DeFronzo, R. A. Role of Adipose Tissue Insulin Resistance in the Natural History of Type 2 Diabetes: Results From the San Antonio Metabolism Study. *Diabetes* **66**, 815–822 (2017).
53. Kahn, S. E., Hull, R. L. & Utzschneider, K. M. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature* **444**, 840–846 (2006).
54. Bevilacqua, S. *et al.* Acute elevation of free fatty acid levels leads to hepatic insulin resistance in obese subjects. *Metabolism* **36**, 502–506 (1987).
55. Benoit, S. C. *et al.* The Catabolic Action of Insulin in the Brain Is Mediated by Melanocortins. *J. Neurosci.* **22**, 9048–9052 (2002).
56. Obici, S. *et al.* Central melanocortin receptors regulate insulin action. *J. Clin. Invest.* **108**, 1079 (2001).
57. Fan, W. *et al.* The Central Melanocortin System Can Directly Regulate Serum Insulin Levels. *Endocrinology* **141**, 3072–3079 (2000).
58. Cusi, K. The role of adipose tissue and lipotoxicity in the pathogenesis of type 2 diabetes. *Curr. Diab. Rep.* **10**, 306–315 (2010).

59. Krssak, M. *et al.* Intramyocellular lipid concentrations are correlated with insulin sensitivity in humans: A 1H NMR spectroscopy study. *Diabetologia* **42**, 113–116 (1999).
60. Samuel, V. T. & Shulman, G. I. Mechanisms for Insulin Resistance: Common Threads and Missing Links. *Cell* **148**, 852–871 (2012).
61. Samuel, V. T., Petersen, K. F. & Shulman, G. I. Lipid-induced insulin resistance: unravelling the mechanism. *Lancet* **375**, 2267 (2010).
62. Maude, H., Sanchez-Cabanillas, C. & Cebola, I. Epigenetics of Hepatic Insulin Resistance. *Front. Endocrinol. (Lausanne).* **12**, 504 (2021).
63. Kahn, S. E. The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of Type 2 diabetes. *Diabetologia* vol. 46 3–19 (2003).
64. Esser, N., Utzschneider, K. M. & Kahn, S. E. Early beta cell dysfunction vs insulin hypersecretion as the primary event in the pathogenesis of dysglycaemia. *Diabetologia* vol. 63 2007–2021 (2020).
65. Kitabchi, A. E. Role of insulin secretion and sensitivity in the evolution of type 2 diabetes in the diabetes prevention program: Effects of lifestyle intervention and metformin. *Diabetes* **54**, 2404–2414 (2005).
66. DeFronzo, R. A. & Tripathy, D. Skeletal muscle insulin resistance is the primary defect in type 2 diabetes. *Diabetes care* vol. 32 Suppl 2 (2009).
67. Martin, B. C. C. *et al.* Role of glucose and insulin resistance in development of type 2 diabetes mellitus: results of a 25-year follow-up study. *Lancet* **340**, 925–929 (1992).
68. Gulli, G., Ferrannini, E., Stern, M., Haffner, S. & Defronzo, R. A. The metabolic profile of NIDDM is fully established in glucose-tolerant offspring of two Mexican-American NIDDM parents. *Diabetes* **41**, 1575–1586 (1992).
69. Kahn, B. B. Type 2 diabetes: When insulin secretion fails to compensate for insulin resistance. *Cell* vol. 92 593–596 (1998).
70. O'Rahilly, S., Turner, R. C. & Matthews, D. R. Impaired Pulsatile Secretion of Insulin in Relatives of Patients with Non-Insulin-Dependent Diabetes. *N. Engl. J. Med.* **318**, 1225–1230 (1988).
71. O'Rahilly, S. P. *et al.* BETA-CELL DYSFUNCTION, RATHER THAN INSULIN INSENSITIVITY, IS THE PRIMARY DEFECT IN FAMILIAL TYPE 2 DIABETES. *Lancet* **328**, 360–364 (1986).
72. Mahajan, A. *et al.* Refining the accuracy of validated target identification

through coding variant fine-mapping in type 2 diabetes. *Nat. Genet.* (2018) doi:10.1038/s41588-018-0084-1.

73. Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143 (2014).
74. Magliano, D. J. *et al.* Young-onset type 2 diabetes mellitus — implications for morbidity and mortality. *Nature Reviews Endocrinology* vol. 16 321–331 (2020).
75. Skrivankova, V. W. *et al.* Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. *JAMA* **326**, 1614–1621 (2021).
76. Sanderson, E. *et al.* Mendelian randomization. *Nat. Rev. Methods Prim.* **2**, 1–21 (2022).
77. Muniyappa, R., Lee, S., Chen, H. & Quon, M. J. Current approaches for assessing insulin sensitivity and resistance in vivo: Advantages, limitations, and appropriate usage. *Am. J. Physiol. - Endocrinol. Metab.* **294**, 15–26 (2008).
78. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. **445**, 881–885 (2007).
79. Grant, S. F. A. The TCF7L2 Locus: A Genetic Window Into the Pathogenesis of Type 1 and Type 2 Diabetes. *Diabetes Care* **42**, 1624 (2019).
80. Mahajan, A. *et al.* Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat. Genet. 2022 545* **54**, 560–572 (2022).
81. Spracklen, C. N. *et al.* Identification of type 2 diabetes loci in 433,540 East Asianindividuals. *Nature* **582**, 240 (2020).
82. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–75, S1-3 (2012).
83. Barroso, I. & McCarthy, M. I. The Genetic Basis of Metabolic Disease. *Cell* **177**, 146–161 (2019).
84. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nat. 2019 5707762* **570**, 514–518 (2019).
85. Peterson, R. E. *et al.* Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* **179**, 589–603 (2019).
86. Wheeler, E., Ile Marenne, G. & Barroso, I. Genetic aetiology of glycaemic traits:

approaches and insights. doi:10.1093/hmg/ddx293.

87. Chen, J. *et al.* The trans-ancestral genomic architecture of glycemic traits. *Nat. Genet. 2021 536* **53**, 840–860 (2021).
88. World Health Organization. *Use of glycated haemoglobin (HbA1c) in diagnosis of diabetes mellitus: abbreviated report of a WHO consultation*. http://www.who.int/iris/handle/10665/70523 (2011).
89. Wheeler, E. *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.* **14**, e1002383 (2017).
90. Leong, A. & Wheeler, E. Genetics of HbA1c: a case study in clinical translation. *Curr. Opin. Genet. Dev.* **50**, 79–85 (2018).
91. Leong, A. & Meigs, J. B. Type 2 Diabetes Prevention: Implications of Hemoglobin A1c Genetics. *Rev. Diabet. Stud.* **12**, 351–362 (2015).
92. American Diabetes Association. 2. Classification and diagnosis of diabetes: Standards of medical care in diabetes-2021. *Diabetes Care* **44**, S15–S33 (2021).
93. Lagou, V. *et al.* Random glucose GWAS in 493,036 individuals provides insights into diabetes pathophysiology, complications and treatment stratification. *medRxiv* (2021) doi:10.1101/2021.04.17.21255471.
94. Prokopenko, I. *et al.* A Central Role for GRB10 in Regulation of Islet Function in Man. *PLOS Genet.* **10**, e1004235 (2014).
95. Strawbridge, R. J. *et al.* Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* **60**, 2624–2634 (2011).
96. Jonsson, A. *et al.* Genome-wide association study of circulating levels of glucagon during an oral glucose tolerance test. *BMC Med. Genomics* **14**, (2021).
97. Almgren, P. *et al.* Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia* **54**, 2811–2819 (2011).
98. Barnett, A. H., Eff, C., Leslie, R. D. G. & Pyke, D. A. Diabetes in identical twins - A study of 200 pairs. *Diabetologia* **20**, 87–93 (1981).
99. Willemsen, G. *et al.* The concordance and heritability of type 2 diabetes in 34,166 twin pairs from international twin registers: the discordant twin (DISCOTWIN) consortium. *cambridge.org* **18**, 2015 (2015).
100. Avery, A. R. & Duncan, G. E. Heritability of type 2 diabetes in the Washington

state twin registry. *Twin Res. Hum. Genet.* **22**, 95–98 (2019).

101. Li, J. H. & Florez, J. C. On the Verge of Precision Medicine in Diabetes. *Drugs* **82**, 1389–1401 (2022).
102. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nat. 2009 4617265* **461**, 747–753 (2009).
103. Young, A. I. Discovering missing heritability in whole-genome sequencing data. *Nat. Genet. 2022 543* **54**, 224–226 (2022).
104. Dwivedi, O. P. *et al.* Loss of ZnT8 function protects against diabetes by enhanced insulin secretion. *Nat. Genet.* **51**, 1596–1606 (2019).
105. Bonnefond, A. *et al.* Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nat. Genet. 2012 443* **44**, 297–301 (2012).
106. George, S. *et al.* A family with severe insulin resistance and diabetes due to a mutation in AKT2. *Science (80-.).* **304**, 1325–1328 (2004).
107. Højlund, K. Metabolism and insulin signaling in common metabolic disorders and inherited insulin resistance. *Dan. Med. J.* **61**, B4890 (2014).
108. Dash, S. *et al.* A truncation mutation in TBC1D4 in a family with acanthosis nigricans and postprandial hyperinsulinemia. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9350–9355 (2009).
109. Moltke, I. *et al.* A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature* **512**, 190–193 (2014).
110. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
111. Grarup, N. *et al.* Identification of novel high-impact recessively inherited type 2 diabetes risk variants in the Greenlandic population. *Diabetologia* **61**, 2005–2015 (2018).
112. Andersen, M. K. & Hansen, T. Genetics of metabolic traits in Greenlanders: lessons from an isolated population. *J. Intern. Med.* **284**, 464–477 (2018).
113. Dash, S. *et al.* Analysis of TBC1D4 in patients with severe insulin resistance. *Diabetologia* vol. 53 1239–1242 (2010).
114. Manning, A. *et al.* A low-frequency inactivating AKT2 variant enriched in the finnish population is associated with fasting insulin levels and type 2 diabetes risk. *Diabetes* **66**, 2019–2032 (2017).
115. Kurki, M. I. *et al.* FinnGen: Unique genetic insights from combining isolated population and national health register data. *medRxiv* 2022.03.03.22271360

(2022) doi:10.1101/2022.03.03.22271360.
116. L, P., P, P. & J, A. Messages from an isolate: lessons from the Finnish gene pool. *Biol. Chem. Hoppe. Seyler.* **376**, 697–714 (1995).
117. Kerminen, S. *et al.* Changes in the fine-scale genetic structure of Finland through the 20th century. *PLOS Genet.* **17**, e1009347 (2021).
118. Kääriäinen, H., Muilu, J., Perola, M. & Kristiansson, K. Genetics in an isolated population like Finland: a different basis for genomic medicine? *J. Community Genet.* **8**, 319 (2017).
119. Latva-Rasku, A. *et al.* A Partial Loss-of-Function Variant in AKT2 Is Associated With Reduced Insulin-Mediated Glucose Uptake in Multiple Insulin-Sensitive Tissues: A Genotype-Based Callback Positron Emission Tomography Study. *Diabetes* **67**, 334–342 (2018).
120. Mäkinen, S. *et al.* Finnish-specific AKT2 gene variant leads to impaired insulin signalling in myotubes. *J. Mol. Endocrinol.* **1**, (2022).
121. Semple, R. K., Savage, D. B., Cochran, E. K., Gorden, P. & O'Rahilly, S. Genetic Syndromes of Severe Insulin Resistance. *Endocr. Rev.* **32**, 498–514 (2011).
122. Challis, B. G. & Semple, R. K. Genetic Disorders of Insulin Action: Far More than Diabetes. *Curr. Obes. Rep.* **2**, 293–300 (2013).
123. Flannick, J. *et al.* Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nat. 2019 5707759* **570**, 71–76 (2019).
124. Nag, A. Human genetic evidence supports MAP3K15 inhibition as a therapeutic strategy for diabetes. *medRxiv* (2021) doi:10.1016/S2213-8587(19)30087-7.
125. Gardner, E. J. *et al.* Damaging missense variants in IGF1R implicate a role for IGF-1 resistance in the etiology of type 2 diabetes. *Cell Genomics* 100208 (2022) doi:10.1016/J.XGEN.2022.100208.
126. Jurgens, S. J. *et al.* Analysis of rare genetic variation underlying cardiometabolic diseases and traits among 200,000 individuals in the UK Biobank. *Nat. Genet.* **54**, 240–250 (2022).
127. Zhao, Y. *et al.* GIGYF1 loss of function is associated with clonal mosaicism and adverse metabolic health. *Nat. Commun.* **12**, (2021).
128. Deaton, A., Parker, M., Ward, L., reports, A. F.-C.-S. & 2021, undefined. Gene-level analysis of rare variants in 379,066 whole exome sequences identifies an association of GIGYF1 loss of function with type 2 diabetes.

nature.com.

129. Wang, Q. *et al.* Rare variant contribution to human disease in 281,104 UK Biobank exomes. *nature.com*.
130. Curtis, D. Analysis of rare coding variants in 200,000 exome-sequenced subjects reveals novel genetic risk factors for type 2 diabetes. *Diabetes. Metab. Res. Rev.* **38**, (2022).
131. Gazal, S. *et al.* Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. *Nat. Genet.* **54**, 827–836 (2022).
132. Pearson, E. R. Type 2 diabetes: a multifaceted disease. *Diabetologia* **62**, 1107–1112 (2019).
133. Nair, A. T. N. *et al.* Heterogeneity in phenotype, disease progression and drug response in type 2 diabetes. *Nat. Med.* **28**, 982–988 (2022).
134. Udler, M. S. *et al.* Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. *PLOS Med.* **15**, e1002654 (2018).
135. Ahlqvist, E. *et al.* Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol.* **6**, 361–369 (2018).
136. Ahlqvist, E., Tuomi, T. & Groop, L. Clusters provide a better holistic view of type 2 diabetes than simple clinical features. *Lancet Diabetes Endocrinol.* **7**, 668–669 (2019).
137. McCarthy, M. I. Painting a new picture of personalised medicine for diabetes. *Diabetologia* vol. 60 793–799 (2017).
138. Carrillo-Larco, R. M., Castillo-Cara, M., Anza-Ramirez, C. & Bernabé-Ortiz, A. Clusters of people with type 2 diabetes in the general population: unsupervised machine learning approach using national surveys in Latin America and the Caribbean. *BMJ Open Diabetes Res. Care* **9**, e001889 (2021).
139. Bello-Chavolla, O. Y. *et al.* Clinical characterization of data-driven diabetes subgroups in Mexicans using a reproducible machine learning approach. *BMJ Open Diabetes Res. Care* **8**, e001550 (2020).
140. Dennis, J. M., Shields, B. M., Henley, W. E., Jones, A. G. & Hattersley, A. T. Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. *Lancet Diabetes Endocrinol.* **7**, 442–451 (2019).

141. Zaharia, O. P. *et al.* Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. *Lancet Diabetes Endocrinol.* **7**, 684–694 (2019).
142. Slieker, R. C. *et al.* Replication and cross-validation of type 2 diabetes subtypes based on clinical variables: an IMI-RHAPSODY study. *Diabetologia* **64**, 1982–1989 (2021).
143. Christensen, D. H. *et al.* Type 2 diabetes classification: a data-driven cluster study of the Danish Centre for Strategic Research in Type 2 Diabetes (DD2) cohort. **10**, e002731 (2022).
144. Wesolowska-Andersen, A. *et al.* Deep learning models predict regulatory variants in pancreatic islets and refine type 2 diabetes association signals. *Elife* **9**, (2020).
145. Scott, R. A. *et al.* An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Diabetes* **66**, 2888–2902 (2017).
146. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
147. Ragvin, A. *et al.* Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to HHEX, SOX4, and IRX3. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 775–780 (2010).
148. Kim, D. S., Gloyn, A. L. & Knowles, J. W. Genetics of Type 2 Diabetes: Opportunities for Precision Medicine. *J. Am. Coll. Cardiol.* **78**, 496–512 (2021).
149. Dimas, A. S. *et al.* Impact of Type 2 Diabetes Susceptibility Variants on Quantitative Glycemic Traits Reveals Mechanistic Heterogeneity. *Diabetes* **63**, 2158–2171 (2014).
150. Scott, R. *et al.* An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. *Am Diabetes Assoc* **66**, 2888–2902.
151. … *et al.* An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* 2022.07.11.22277436 (2017) doi:10.1101/2022.07.11.22277436.
152. Lotta, L. A. *et al.* Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat. Genet.* **49**, 17–26 (2017).
153. Edwards, S. L., Beesley, J., French, J. D. & Dunning, M. Beyond GWASs: Illuminating the dark road from association to function. *Am. J. Hum. Genet.* **93**,

779–797 (2013).
154. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (80-.).* **369**, 1318–1330 (2020).
155. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nature Genetics* vol. 45 580–585 (2013).
156. Zhu, Y., Wang, L., Yin, Y. & Yang, E. Systematic analysis of gene expression patterns associated with postmortem interval in human tissues. *Sci. Rep.* **7**, 5435 (2017).
157. Jiang, L. *et al.* A Quantitative Proteome Map of the Human Body. *Cell* **183**, 269-283.e19 (2020).
158. Chiou, J. *et al.* Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature* **594**, 398–402 (2021).
159. Chiou, J. *et al.* Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature* **594**, 398–402 (2021).
160. Jagadeesh, K. A. *et al.* Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat. Genet. 2022 5410* **54**, 1479–1492 (2022).
161. Butler, L. M. *et al.* Analysis of Body-wide Unfractionated Tissue Data to Identify a Core Human Endothelial Transcriptome. *Cell Syst.* **3**, 287-301.e3 (2016).
162. Norreen-Thorsen, M. *et al.* A human adipose tissue cell-type transcriptome atlas. *Cell Rep.* **40**, (2022).
163. Shin, S. Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**, 543–550 (2014).
164. Long, T. *et al.* Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat. Genet.* **49**, 568–578 (2017).
165. Surendran, P. *et al.* Rare and common genetic determinants of metabolic individuality and their effects on human health. *Nat. Med. 2022 2811* **28**, 2321–2332 (2022).
166. Yin, X. *et al.* Integrating transcriptomics, metabolomics, and GWAS helps reveal molecular mechanisms for metabolite levels and disease risk. *Am. J. Hum. Genet.* **109**, 1727–1741 (2022).
167. Yin, X. *et al.* Genome-wide association studies of metabolites in Finnish men identify disease-relevant loci. *Nat. Commun. 2022 131* **13**, 1–14 (2022).
168. Suhre, K., McCarthy, M. I. & Schwenk, J. M. Genetics meets proteomics: perspectives for large population-based studies. *Nat. Rev. Genet. 2020 221* **22**,

19–37 (2020).
169. Pietzner, M. *et al.* Mapping the proteo-genomic convergence of human diseases. *Science (80-.).* **374**, (2021).
170. Folkersen, L. *et al.* Genomic evaluation of circulating proteins for drug target characterisation and precision medicine. Preprint at. *bioRxiv* (2020) doi:10.1101/2020.04.03.023804.
171. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **53**, 1712–1721 (2021).
172. Gudjonsson, A. *et al.* A genome-wide association study of serum proteins reveals shared loci with common diseases. *Nat. Commun.* **13**, (2022).
173. Pietzner, M. *et al.* Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat. Commun.* **12**, (2021).
174. Sun BB *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
175. Lourdusamy, A. *et al.* Identification of cis-regulatory variation influencing protein abundance levels in human plasma. *Hum. Mol. Genet.* **21**, 3719–3726 (2012).
176. Foley, C. N. *et al.* A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat. Commun. 2021 121* **12**, 1–18 (2021).
177. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLOS Genet.* **10**, e1004383 (2014).
178. Zuber, V. *et al.* Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches. *Am. J. Hum. Genet.* **109**, 767 (2022).
179. Hemani, G. *et al.* The MR-base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).
180. Smith, G. D. & Hemani, G. Mendelian randomization: Geneticanchorsfor causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, (2014).
181. Davey Smith, G. & Ebrahim, S. Mendelian randomization: Genetic variants as instruments for strengthening causal inference in observational studies. in *Biosocial Surveys* 336–366 (National Academies Press (US), 2008). doi:10.17226/11939.
182. Lawlor, D. A. Commentary: Two-sample Mendelian randomization:

opportunities and challenges. *Int. J. Epidemiol.* **45**, 908 (2016).
183. Grarup, N., Sparsø, T. & Hansen, T. Physiologic characterization of type 2 diabetes-related loci. *Curr. Diab. Rep.* **10**, 485–497 (2010).
184. DeForest, N. & Majithia, A. R. Genetics of Type 2 Diabetes: Implications from Large-Scale Studies. *Curr. Diab. Rep.* **22**, 227 (2022).
185. Grotz, A. K., Gloyn, A. L. & Thomsen, S. K. Prioritising Causal Genes at Type 2 Diabetes Risk Loci. *Curr. Diab. Rep.* **17**, (2017).
186. Miosge, L. A. *et al.* Comparison of predicted and actual consequences of missense mutations. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5189–E5198 (2015).
187. Starita, L. M. *et al.* Variant Interpretation: Functional Assays to the Rescue. *American Journal of Human Genetics* vol. 101 315–325 (2017).
188. Nica, A. C. & Dermitzakis, E. T. Expression quantitative trait loci: Present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 368 20120362 (2013).
189. Wang, J. *et al.* Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* **22**, 1798–812 (2012).
190. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
191. Pang, D. X., Smith, A. J. P. & Humphries, S. E. Functional analysis of TCF7L2 genetic variants associated with type 2 diabetes. *Nutr. Metab. Cardiovasc. Dis.* **23**, 550–556 (2013).
192. Lyssenko, V. *et al.* Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. *J. Clin. Invest.* **117**, 2155–2163 (2007).
193. Schäfer, S. A. *et al.* Impaired glucagon-like peptide-1-induced insulin secretion in carriers of transcription factor 7-like 2 (TCF7L2) gene polymorphisms. *Diabetologia* **50**, 2443–2450 (2007).
194. Mathiesen, D. S. *et al.* No detectable effect of a type 2 diabetes-associated TCF7L2 genotype on the incretin effect. *Endocr. Connect.* **9**, 1221–1232 (2020).
195. Bazwinsky-Wutschke, I., Bieseke, L., Mühlbauer, E. & Peschke, E. Influence of melatonin receptor signalling on parameters involved in blood glucose regulation. *J. Pineal Res.* **56**, 82–96 (2014).
196. Lyssenko, V. *et al.* Common variant in MTNR1B associated with increased risk

of type 2 diabetes and impaired early insulin secretion. *Nat. Genet.* **41**, 82–88 (2009).

197. Mulder, H., Nagorny, C. L. F., Lyssenko, V. & Groop, L. Melatonin receptors in pancreatic islets: Good morning to a novel type 2 diabetes gene. *Diabetologia* **52**, 1240–1249 (2009).
198. van de Bunt, M. *et al.* Transcript Expression Data from Human Islets Links Regulatory Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream Effectors. *PLOS Genet.* **11**, e1005694 (2015).
199. Tuomi, T. *et al.* Increased Melatonin Signaling Is a Risk Factor for Type 2 Diabetes. *Cell Metab.* **23**, 1067–1077 (2016).
200. Prokopenko, I. *et al.* Variants in MTNR1B influence fasting glucose levels. *Nat. Genet. 2008 411* **41**, 77–81 (2008).
201. Rubio-Sastre, P., Scheer, F. A. J. L., Gómez-Abellán, P., Madrid, J. A. & Garaulet, M. Acute melatonin administration in humans impairs glucose tolerance in both the morning and evening. *Sleep* **37**, 1715-1719B (2014).
202. Fernández-Tajes, J. *et al.* Developing a network view of type 2 diabetes risk pathways through integration of genetic, genomic and functional data. *Genome Med.* **11**, 1–14 (2019).
203. Rao, S., Yao, Y. & Bauer, D. E. Editing GWAS: experimental approaches to dissect and exploit disease-associated genetic variation. *Genome Med. 2021 131* **13**, 1–20 (2021).
204. Khamis, A. *et al.* Laser capture microdissection of human pancreatic islets reveals novel eQTLs associated with type 2 diabetes. *Mol. Metab.* **24**, 98–107 (2019).
205. Ndiaye, F. K. *et al.* Expression and functional assessment of candidate type 2 diabetes susceptibility genes identify four new genes contributing to human insulin secretion. *Mol. Metab.* **6**, 459–470 (2017).
206. Khetan, S. *et al.* Functional characterization of T2D-associated SNP effects on baseline and ER stress-responsive β cell transcriptional activation. *Nat. Commun. 2021 121* **12**, 1–13 (2021).
207. Grotz, A. K. *et al.* A genome-wide CRISPR screen identifies regulators of beta cell function involved in type 2 diabetes risk. *bioRxiv* 2021.05.28.445984 (2021) doi:10.1101/2021.05.28.445984.
208. Thomsen, S. K. *et al.* Systematic Functional Characterization of Candidate

Causal Genes for Type 2 Diabetes Risk Variants. *Diabetes* **65**, 3805–3811 (2016).

209. Tu, Z. *et al.* Integrating siRNA and protein–protein interaction data to identify an expanded insulin signaling network. *Genome Res.* **19**, 1057–1067 (2009).
210. Haney, S. *et al.* RNAi Screening in Primary Human Hepatocytes of Genes Implicated in Genome-Wide Association Studies for Roles in Type 2 Diabetes Identifies Roles for CAMK1D and CDKAL1, among Others, in Hepatic Glucose Regulation. *PLoS One* **8**, e64946 (2013).
211. Yang, R. *et al.* A genome-wide siRNA screen to identify modulators of insulin sensitivity and gluconeogenesis. *PLoS One* **7**, (2012).
212. Muoio, D. M. & Newgard, C. B. Molecular and metabolic mechanisms of insulin resistance and β-cell failure in type 2 diabetes. *Nat. Rev. Mol. Cell Biol.* **9**, 193–205 (2008).
213. Scott, R. A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* **44**, 991–1005 (2012).
214. Lagou, V. *et al.* Sex-dimorphic genetic effects and novel loci for fasting glucose and insulin variability. *Nat. Commun. 2021 121* **12**, 1–18 (2021).
215. Taylor, R. *et al.* Direct assessment of liver glycogen storage by 13C nuclear magnetic resonance spectroscopy and regulation of glucose homeostasis after a mixed meal in normal subjects. *J. Clin. Invest.* **97**, 126–132 (1996).
216. Jue, T., Rothman, D. L., Tavitian, B. A. & Shulman, R. G. Natural-abundance 13C NMR study of glycogen repletion in human liver and muscle. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 1439–1442 (1989).
217. Goodyear, L. J. *et al.* Glucose Ingestion Causes GLUT4 Translocation in Human Skeletal Muscle. *Diabetes* **45**, 1051–1056 (1996).
218. Kahn, B. B. Dietary Regulation of Glucose Transporter Gene Expression: Tissue Specific Effects in Adipose Cells and Muscle. *J. Nutr.* **124**, 1289S-1295S (1994).
219. Maianu, L., Keller, S. R. & Garvey, W. T. Adipocytes Exhibit Abnormal Subcellular Distribution and Translocation of Vesicles Containing Glucose Transporter 4 and Insulin-Regulated Aminopeptidase in Type 2 Diabetes Mellitus: Implications Regarding Defects in Vesicle Trafficking. *J. Clin. Endocrinol. Metab.* **86**, 5450–5456 (2001).
220. Rothman, D. L. *et al.* Decreased muscle glucose transport/phosphorylation is

an early defect in the pathogenesis of non-insulin-dependent diabetes mellitus. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 983–987 (1995).

221. DeFronzo, R. A. The triumvirate: β-cell, muscle, liver. A collusion responsible for NIDDM. *Diabetes* vol. 37 667–687 (1988).
222. Haigh, J. W. *et al.* Assessment of Glucose Turnover in Normal Man with the Use of a Non-Radioactive Isotopically Labelled Preparation, [6,6-2H]Glucose, as Tracer. *Clin. Sci.* **63**, 437–440 (1982).
223. Friedlander, A. L. *et al.* Training-induced alterations of carbohydrate metabolism in women: women respond differently from men. *J. Appl. Physiol.* **85**, 1175–1186 (1998).
224. DeFronzo, R. A., Tobin, J. D. & Andres, R. Glucose clamp technique: a method for quantifying insulin secretion and resistance. *https://doi.org/10.1152/ajpendo.1979.237.3.E214* **6**, (1979).
225. Otten, J., Ahrén, B. & Olsson, T. Surrogate measures of insulin sensitivity vs the hyperinsulinaemic- euglycaemic clamp: A meta-analysis. *Diabetologia* **57**, 1781–1788 (2014).
226. Herzberg-Schäfer, S. A. *et al.* Evaluation of fasting state-/oral glucose tolerance test-derived measures of insulin release for the detection of genetically impaired β-cell function. *PLoS One* **5**, e14194 (2010).
227. Antoniolli, L. P., Nedel, B. L., Pazinato, T. C., de Andrade Mesquita, L. & Gerchman, F. Accuracy of insulin resistance indices for metabolic syndrome: a cross-sectional study in adults. *Diabetol. Metab. Syndr.* **10**, 65 (2018).
228. Matsuda, M. & DeFronzo, R. A. Insulin sensitivity indices obtained from oral glucose tolerance testing: Comparison with the euglycemic insulin clamp. *Diabetes Care* **22**, 1462–1470 (1999).
229. Matthews, D. R. *et al.* Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetol. 1985 287* **28**, 412–419 (1985).
230. Katz, A. *et al.* Quantitative insulin sensitivity check index: a simple, accurate method for assessing insulin sensitivity in humans. *J. Clin. Endocrinol. Metab.* **85**, 2402–2410 (2000).
231. Abdul-Ghani, M. A., Jenkinson, C. P., Richardson, D. K., Tripathy, D. & DeFronzo, R. A. Insulin secretion and action in subjects with impaired fasting glucose and impaired glucose tolerance: Results from the veterans administration genetic epidemiology study. *Diabetes* **55**, 1430–1435 (2006).

232. Hoffman, R. P. Indices of insulin action calculated from fasting glucose and insulin reflect hepatic, not peripheral, insulin sensitivity in African-American and Caucasian adolescents. *Pediatr. Diabetes* **9**, 57–61 (2008).
233. Wallace, T. M., Levy, J. C. & Matthews, D. R. Use and abuse of HOMA modeling. *Diabetes Care* **27**, 1487–1495 (2004).
234. Stumvoll, M., Van Haeften, T., Fritsche, A. & Gerich, J. Oral glucose tolerance test indexes for insulin sensitivity and secretion based on various availabilities of sampling times [11]. *Diabetes Care* vol. 24 796–797 (2001).
235. Stumvoll, M. *et al.* Use of the oral glucose tolerance test to assess insulin release and insulin sensitivity. *Diabetes Care* **23**, 295–301 (2000).
236. Moltke, I. *et al.* A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature* **512**, 190–193 (2014).
237. Sano, H. *et al.* Insulin-stimulated phosphorylation of a Rab GTPase-activating protein regulates GLUT4 translocation. *J. Biol. Chem.* **278**, 14599–14602 (2003).
238. Walford, G. A. *et al.* Genome-wide association study of the modified stumvoll insulin sensitivity index identifies BCL2 and FAM19A2 as novel insulin sensitivity loci. *Diabetes* **65**, 3200–3211 (2016).
239. Dimas, A. S. *et al.* Impact of Type 2 Diabetes Susceptibility Variants on Quantitative Glycemic Traits Reveals Mechanistic Heterogeneity. *Diabetes* **63**, 2158–2171 (2014).
240. Lindsay, T. *et al.* Descriptive epidemiology of physical activity energy expenditure in UK adults (The Fenland study). *Int. J. Behav. Nutr. Phys. Act.* **16**, (2019).
241. Hills, S. A. *et al.* The EGIR-RISC Study (the European group for the study of insulin resistance: Relationship between insulin sensitivity and cardiovascular disease risk): I. Methodology and Objectives. *Diabetologia* **47**, 566–570 (2004).
242. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
243. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
244. Lim, E. T. *et al.* Distribution and Medical Impact of Loss-of-Function Variants in the Finnish Founder Population. *PLoS Genet.* **10**, e1004494 (2014).
245. FinnGen. FinnGen Documentation of R3 release. https://finngen.gitbook.io/documentation/ (2020).

246. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Tsegaselassie Work.* **12**, 1192–1212 (2014).
247. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
248. Scott, R. A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* **44**, 991–1005 (2012).
249. GitHub - explodecomputer/random-metal: Adding random effects model to the METAL software. https://github.com/explodecomputer/random-metal.
250. Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet. Epidemiol.* **33**, 79–86 (2009).
251. Vujkovic, M. *et al.* Discovery of 318 novel loci for type-2 diabetes and related micro- and macrovascular outcomes among 1.4 million participants in a multi-ethnic meta-analysis. *medRxiv* 19012690 (2019) doi:10.1101/19012690.
252. Pulit, S. L., Karaderi, T. & Lindgren, C. M. Sexual dimorphisms in genetic loci linked to body fat distribution. *Biosci. Rep.* **37**, BSR20160184 (2017).
253. Wang, Q. *et al.* Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nat. 2021 5977877* **597**, 527–532 (2021).
254. Bulik-Sullivan, B. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet. 2015 473* **47**, 291–295 (2015).
255. Vujkovic, M. *et al.* Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat. Genet. 2020 527* **52**, 680–691 (2020).
256. Pulit, S. L. *et al.* Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).
257. Lotta, L. A. *et al.* Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat Genet* **49**, 17–26 (2017).
258. Leland Taylor, D. *et al.* Integrative analysis of gene expression, DNA methylation, physiological traits, and genetic variation in human skeletal muscle. *Proc. Natl. Acad. Sci. U. S. A.* 10883–10888 (2019).
259. Viñuela, A. *et al.* Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. *Nat. Commun. 2020 111* **11**, 1–

14 (2020).
260. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621 (2018).
261. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–329 (2015).
262. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
263. Zhu, C. H. *et al.* Cellular senescence in human myoblasts is overcome by human telomerase reverse transcriptase and cyclin-dependent kinase 4: Consequences in aging muscle and therapeutic strategies for muscular dystrophies. *Aging Cell* **6**, 515–523 (2007).
264. Fogarty, M. P., Cannon, M. E., Vadlamudi, S., Gaulton, K. J. & Mohlke, K. L. Identification of a Regulatory Variant That Binds FOXA1 and FOXA2 at the CDC123/CAMK1D Type 2 Diabetes GWAS Locus. *PLoS Genet.* **10**, (2014).
265. Roman, T. S. *et al.* A type 2 diabetes-associated functional regulatory variant in a pancreatic islet enhancer at the ADCY5 locus. in *Diabetes* vol. 66 2521–2530 (Diabetes, 2017).
266. Williams, A. L. *et al.* Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* **506**, 97–101 (2014).
267. Machiela, M. J. & Chanock, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31**, 3555–3557 (2015).
268. Sajuthi, S. P. *et al.* Mapping adipose and muscle tissue expression quantitative trait loci in African Americans to identify genes for type 2 diabetes and obesity. *Hum. Genet.* **135**, 869–880 (2016).
269. Keildson, S. *et al.* Expression of Phosphofructokinase in Skeletal Muscle Is Influenced by Genetic Variation and Associated With Insulin Sensitivity. *Diabetes* **63**, 1154–1165 (2014).
270. Mason, C. C. *et al.* Bimodal distribution of RNA expression levels in human skeletal muscle tissue. *BMC Genomics* **12**, (2011).
271. Jordens, I., Molle, D., Xiong, W., Keller, S. R. & McGraw, T. E. Insulin-regulated aminopeptidase is a key regulator of GLUT4 trafficking by controlling the sorting of GLUT4 from endosomes to specialized insulin-regulated vesicles. *Mol. Biol. Cell* **21**, 2034–2044 (2010).
272. Sano, H. *et al.* Rab10, a target of the AS160 Rab GAP, is required for insulin-

stimulated translocation of GLUT4 to the adipocyte plasma membrane. *Cell Metab.* **5**, 293–303 (2007).

273. Klip, A., McGraw, T. E. & James, D. E. Thirty sweet years of GLUT4. **294**, 11369–11381 (2019).
274. Stenbit, A. E. *et al.* GLUT4 heterozygous knockout mice develop muscle insulin resistance and diabetes. *Nat. Med.* **3**, 1096–1101 (1997).
275. Ng, N. H. J. *et al.* Tissue-Specific Alteration of Metabolic Pathways Influences Glycemic Regulation. *bioRxiv* **29**, 790618 (2019).
276. Sim, N.-L. *et al.* SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40**, W452–W457 (2012).
277. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
278. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
279. Gual, P., Le Marchand-Brustel, Y. & Tanti, J. F. Positive and negative regulation of insulin signaling through IRS-1 phosphorylation. *Biochimie* **87**, 99–109 (2005).
280. Barroso, I. Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature* **402**, 880–883.
281. Li, Q. *et al.* The Protein Phosphatase 1 Complex Is a Direct Target of AKT that Links Insulin Signaling to Hepatic Glycogen Deposition. *Cell Rep.* **28**, 3406-3422.e7 (2019).
282. Agius, L. Role of glycogen phosphorylase in liver glycogen metabolism. *Mol. Aspects Med.* **46**, 34–45 (2015).
283. Yoon, M. S. The Role of Mammalian Target of Rapamycin (mTOR) in Insulin Signaling. *Nutrients* **9**, (2017).
284. Kuo, T. *et al.* Identification of C2CD4A as a human diabetes susceptibility gene with a role in β cell insulin secretion. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20033–20042 (2019).
285. Bonnefond, A. & Froguel, P. Disentangling the Role of Melatonin and its Receptor MTNR1B in Type 2 Diabetes: Still a Long Way to Go? *Curr. Diab. Rep.* **17**, 122 (2017).
286. Kahn, B. B. Adipose Tissue, Inter-Organ Communication, and the Path to Type

2 Diabetes: The 2016 Banting Medal for Scientific Achievement Lecture. *Diabetes* **68**, 3–14 (2019).

287. Petersen, K. F. *et al.* The role of skeletal muscle insulin resistance in the pathogenesis of the metabolic syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12587–12594 (2007).
288. Larsen, J. K. *et al.* Illumination of the Endogenous Insulin-Regulated TBC1D4 Interactome in Human Skeletal Muscle. *Diabetes* **71**, 906–920 (2022).
289. Koh, H. C. E. *et al.* Heterogeneity in insulin-stimulated glucose uptake among different muscle groups in healthy lean people and people with obesity. *Diabetologia* **64**, 1158–1168 (2021).
290. Honka, M. J. *et al.* Insulin-stimulated glucose uptake in skeletal muscle, adipose tissue and liver: a positron emission tomography study. *Eur. J. Endocrinol.* **178**, (2018).
291. Virtanen, K. A. *et al.* Human adipose tissue glucose uptake determined using [18F]-fluoro-deoxy-glucose ([18F]FDG) and PET in combination with microdialysis. *Diabetologia* **44**, 2171–2179 (2001).
292. Defronzo, R. A. From the triumvirate to the ominous octet: A new paradigm for the treatment of type 2 diabetes mellitus. in *Diabetes* vol. 58 773–795 (American Diabetes Association, 2009).
293. Virtanen, K. A. *et al.* Increased fat mass compensates for insulin resistance in abdominal obesity and type 2 diabetes: A positron-emitting tomography study. *Diabetes* **54**, 2720–2726 (2005).
294. Ferrannini, E. *et al.* Adipose tissue and skeletal muscle insulin-mediated glucose uptake in insulin resistance: Role of blood flow and diabetes. *Am. J. Clin. Nutr.* **108**, 749–758 (2018).
295. Merz, K. E. & Thurmond, D. C. Role of skeletal muscle in insulin resistance and glucose uptake. *Compr. Physiol.* **10**, 785–809 (2020).
296. DeFronzo, R. A. & Tripathy, D. Skeletal Muscle Insulin Resistance Is the Primary Defect in Type 2 Diabetes. *Diabetes Care* **32**, S157–S163 (2009).
297. Coelho, M., Oliveira, T. & Fernandes, R. Biochemistry of adipose tissue: An endocrine organ. *Archives of Medical Science* vol. 9 191–200 (2013).
298. Hou, J. C. & Pessin, J. E. Ins (endocytosis) and outs (exocytosis) of GLUT4 trafficking. *Current Opinion in Cell Biology* vol. 19 466–473 (2007).
299. Katz, E. B., Stenbit, A. E., Hatton, K., Depinho, R. & Charron, M. J. Cardiac and adipose tissue abnormalities but not diabetes in mice deficient in GLUT4.

Nature **377**, 151–155 (1995).

300. Stenbit, A. E. *et al.* Diverse effects of Glut 4 ablation on glucose uptake and glycogen synthesis in red and white skeletal muscle. *J. Clin. Invest.* **98**, 629 (1996).
301. Rossetti, L. *et al.* Impact of Glucose Transporter Type 4 on In Vivo Insulin Action Peripheral but Not Hepatic Insulin Resistance in Mice with One Disrupted Allele of the Glucose Transporter Type 4 (GLUT4) Gene GLUT4 • insulin resistance • glucose transport • hepatic glucose . *J. Clin. Invest* **100**, 1831–1839 (1997).
302. Zisman, A. *et al.* Targeted disruption of the glucose transporter 4 selectively in muscle causes insulin resistance and glucose intolerance. *Nat. Med. 2000 68* **6**, 924–928 (2000).
303. Carvalho, E., Kotani, K., Peroni, O. D. & Kahn, B. B. Adipose-specific overexpression of GLUT4 reverses insulin resistance and diabetes in mice lacking GLUT4 selectively in muscle. *Am. J. Physiol. - Endocrinol. Metab.* **289**, 551–561 (2005).
304. James, D. E., Burleigh, K. M. & Kraegen, E. W. Time Dependence of Insulin Action in Muscle and Adipose Tissue in the Rat In Vivo: An Increasing Response in Adipose Tissue with Time. *Diabetes* **34**, 1049–1054 (1985).
305. Abel, E. D. *et al.* Adipose-selective targeting of the GLUT4 gene impairs insulin action in muscle and liver. *Nat. 2001 4096821* **409**, 729–733 (2001).
306. Handberg, A., Vaag, A., Damsbo, P., Beck-Nielsen, H. & Vinten, J. Expression of insulin regulatable glucose transporters in skeletal muscle from Type 2 (non-insulin-dependent) diabetic patients. *Diabetologia* **33**, 625–627 (1990).
307. Pedersen, O. *et al.* Evidence against altered expression of GLUT1 or GLUT4 in skeletal muscle of patients with obesity for NIDDM. *Diabetes* **39**, 865–870 (1990).
308. Tremblay, F., Lavigne, C., Jacques, H. & Marette, A. Defective Insulin-Induced GLUT4 Translocation in Skeletal Muscle of High Fat-Fed Rats Is Associated with Alterations in Both Akt/Protein Kinase B and Atypical Protein Kinase C (ζ/λ) Activities. *Diabetes* **50**, 1901–1910 (2001).
309. Cong, L. N. *et al.* Physiological role of AKT in insulin-stimulated translocation of GLUT4 in transfected rat adipose cells. *Mol. Endocrinol.* **11**, 1881–1890 (1997).
310. Peck, G. R. *et al.* Interaction of the Akt substrate, AS160, with the glucose

transporter 4 vesicle marker protein, insulin-regulated aminopeptidase. *Mol. Endocrinol.* **20**, 2576–2583 (2006).

311. Mîinea, C. P. *et al.* AS160, the Akt substrate regulating GLUT4 translocation, has a functional Rab GTPase-activating protein domain. *Biochem. J.* **391**, (2005).
312. Sadacca, L. A., Bruno, J., Wen, J., Xiong, W. & McGraw, T. E. Specialized sorting of GLUT4 and its recruitment to the cell surface are independently regulated by distinct Rabs. *Mol. Biol. Cell* **24**, 2544–2557 (2013).
313. Chen, Y. *et al.* Rab10 and myosin-va mediate insulin-stimulated GLUT4 storage vesicle translocation in adipocytes. *J. Cell Biol.* **198**, 545–560 (2012).
314. Bruss, M. D., Arias, E. B., Lienhard, G. E. & Cartee, G. D. Increased phosphorylation of Akt substrate of 160 kDa (AS160) in rat skeletal muscle in response to insulin or contractile activity. *Diabetes* **54**, 41–50 (2005).
315. Lee, J. Y., Chiu, Y. H., Asara, J. & Cantley, L. C. Inhibition of PI3K binding to activators by serine phosphorylation of PI3K regulatory subunit p85α Src homology-2 domains. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 14157–14162 (2011).
316. Fazakerley, D. J., Koumanov, F. & Holman, G. D. GLUT4 On the move. *Biochem. J.* **479**, 445–462 (2022).
317. Karylowski, O., Zeigerer, A., Cohen, A. & McGraw, T. E. GLUT4 Is Retained by an Intracellular Cycle of Vesicle Formation and Fusion with Endosomes. *Mol. Biol. Cell* **15**, 870 (2004).
318. Li, D. T. *et al.* GLUT4 storage vesicles: Specialized organelles for regulated trafficking. *Yale Journal of Biology and Medicine* vol. 92 453–470 (2019).
319. Brumfield, A. *et al.* Insulin-promoted mobilization of GLUT4 from a perinuclear storage site requires RAB10. *Mol. Biol. Cell* **32**, 57–73 (2021).
320. Huang, S. & Czech, M. P. The GLUT4 Glucose Transporter. *Cell Metab.* **5**, 237–252 (2007).
321. Martin, O. J., Lee, A. & McGraw, T. E. GLUT4 distribution between the plasma membrane and the intracellular compartments is maintained by an insulin-modulated bipartite dynamic mechanism. *J. Biol. Chem.* **281**, 484–490 (2006).
322. Stöckli, J., Fazakerley, D. J. & James, D. E. GLUT4 exocytosis. *Journal of Cell Science* vol. 124 4147–4159 (2011).
323. Pessin, J. E., Thurmond, D. C., Elmendorf, J. S., Coker, K. J. & Okada, S. Molecular basis of insulin-stimulated GLUT4 vesicle trafficking. Location! Location! Location! *Journal of Biological Chemistry* vol. 274 2593–2596 (1999).

324. Martin, O. J., Lee, A. & McGraw, T. E. GLUT4 distribution between the plasma membrane and the intracellular compartments is maintained by an insulin-modulated bipartite dynamic mechanism. *J. Biol. Chem.* **281**, 484–490 (2006).
325. Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. *Brief. Bioinform.* **14**, 144–61 (2013).
326. Sun, W. *et al.* A Transcriptomic Analysis Reveals Novel Patterns of Gene Expression During 3T3-L1 Adipocyte Differentiation. *Front. Mol. Biosci.* **7**, 249 (2020).
327. Kohn, A. D., Summers, S. A., Birnbaum, M. J. & Roth, R. A. Expression of a constitutively active Akt Ser/Thr kinase in 3T3-L1 adipocytes stimulates glucose uptake and glucose transporter 4 translocation. *J. Biol. Chem.* **271**, 31372–31378 (1996).
328. Garza, L. A. & Birnbaum, M. J. Insulin-responsive aminopeptidase trafficking in 3T3-L1 adipocytes. *J. Biol. Chem.* **275**, 2560–2567 (2000).
329. Knutson, V. P. & Balba, Y. 3T3-L1 adipocytes as a cell culture model of insulin resistance [2]. *In Vitro Cellular and Developmental Biology - Animal* vol. 33 77–81 (1997).
330. Tucker, D. F. *et al.* Isolation of state-dependent monoclonal antibodies against the 12-transmembrane domain glucose transporter 4 using virus-like particles. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4990–E4999 (2018).
331. James, D. E., Strube, M. & Muecdler, M. Molecular cloning and characterization of an insulin-regulatable glucose transporter. *Nature* **338**, 83–87 (1989).
332. Ng, Y., Ramm, G., Lopez, J. A. & James, D. E. Rapid activation of Akt2 is sufficient to stimulate GLUT4 translocation in 3T3-L1 adipocytes. *Cell Metab.* **7**, 348–356 (2008).
333. Stöckli, J. *et al.* Regulation of Glucose Transporter 4 Translocation by the Rab Guanosine Triphosphatase-Activating Protein AS160/TBC1D4: Role of Phosphorylation and Membrane Association. *Mol. Endocrinol.* **22**, 2703 (2008).
334. Kanai, F. *et al.* Insulin-stimulated GLUT4 translocation is relevant to the phosphorylation of IRS-1 and the activity of PI3-kinase. *Biochem. Biophys. Res. Commun.* **195**, 762–768 (1993).
335. Keller, S. R., Scott, H. M., Mastick, C. C., Aebersold, R. & Lienhard, G. E. Cloning and Characterization of a Novel Insulin-regulated Membrane Aminopeptidase from Glut4 Vesicles. *J. Biol. Chem.* **270**, 23612–23618 (1995).

336. Chi, N. W. & Lodish, H. F. Tankyrase is a golgi-associated mitogen-activated protein kinase substrate that interacts with IRAP in GLUT4 vesicles. *J. Biol. Chem.* **275**, 38437–38444 (2000).
337. Guo, H. L. *et al.* The Axin/TNKS complex interacts with KIF3A and is required for insulin-stimulated GLUT4 translocation. *Cell Res.* **22**, 1246–1257 (2012).
338. Hook, S. C. *et al.* TBC1D1 interacting proteins, VPS13A and VPS13C, regulate GLUT4 homeostasis in C2C12 myotubes. *Sci. Rep.* **10**, 1–14 (2020).
339. Yeh, T. Y. J., Sbodio, J. I., Tsun, Z. Y., Luo, B. & Chi, N. W. Insulin-stimulated exocytosis of GLUT4 is enhanced by IRAP and its partner tankyrase. *Biochem. J.* **402**, 279–290 (2007).
340. Akram, M. Citric Acid Cycle and Role of its Intermediates in Metabolism. *Cell Biochemistry and Biophysics* vol. 68 475–478 (2014).
341. Liu, H. *et al.* ALKBH5-mediated m6A demethylation of GLUT4 mRNA promotes glycolysis and resistance to HER2-targeted therapy in breast cancer. *Cancer Res.* (2022) doi:10.1158/0008-5472.CAN-22-0800.
342. Kopecki, Z. & Cowin, A. J. Flightless I: an actin-remodelling protein and an important negative regulator of wound repair. *Int. J. Biochem. Cell Biol.* **40**, 1415–1419 (2008).
343. Lee, Y.-H., Campbell, H. D. & Stallcup, M. R. Developmentally Essential Protein Flightless I Is a Nuclear Receptor Coactivator with Actin Binding Activity. *Mol. Cell. Biol.* **24**, 2103 (2004).
344. Strudwick, X. L. & Cowin, A. J. Multifunctional Roles of the Actin-Binding Protein Flightless I in Inflammation, Cancer and Wound Healing. *Frontiers in Cell and Developmental Biology* vol. 8 1394 (2020).
345. Van Rompay, A. R., Johansson, M. & Karlsson, A. Phosphorylation of deoxycytidine analog monophosphates by UMP-CMP kinase: Molecular characterization of the human enzyme. *Mol. Pharmacol.* **56**, 562–569 (1999).
346. Hsu, C. H., Liou, J. Y., Dutschman, G. E. & Cheng, Y. C. Phosphorylation of cytidine, deoxycytidine, and their analog monophosphates by human UMP/CMP kinase is differentially regulated by ATP and magnesium. *Mol. Pharmacol.* **67**, 806–814 (2005).
347. Revelo, X. S. *et al.* Nucleic Acid-Targeting Pathways Promote Inflammation in Obesity-Related Insulin Resistance. *Cell Rep.* **16**, 717–730 (2016).
348. Balasubramanian, R., Robaye, B., Boeynaems, J. M. & Jacobson, K. A. Enhancement of glucose uptake in mouse skeletal muscle cells and adipocytes

by P2Y6 receptor agonists. *PLoS One* **9**, e116203 (2014).
349. Ko, M. S. *et al.* Overexpression of DRG2 suppresses the growth of Jurkat T cells but does not induce apoptosis. *Arch. Biochem. Biophys.* **422**, 137–144 (2004).
350. Jang, S. H., Kim, A. R., Park, N. H., Park, J. W. & Han, I. S. DRG2 regulates G2/M progression via the cyclin B1-Cdk1 complex. *Mol. Cells* **39**, 699–704 (2016).
351. Park, B. S. *et al.* Developmentally regulated GTP-binding protein-2 regulates adipocyte differentiation. *Biochem. Biophys. Res. Commun.* **578**, 1–6 (2021).
352. Degrandmaison, J. *et al.* In vivo mapping of a GPCR interactome using knockin mice. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 13105–13116 (2020).
353. Degrandmaison, J. *et al.* In vivo mapping of a GPCR interactome using knockin mice. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 13105–13116 (2020).
354. Chen, Y. *et al.* Rab10 and myosin-va mediate insulin-stimulated GLUT4 storage vesicle translocation in adipocytes. *J. Cell Biol.* **198**, 545–560 (2012).
355. Niu, Y. *et al.* Structural basis of inhibition of the human SGLT2–MAP17 glucose transporter. *Nature* **601**, 280–284 (2022).
356. Coady, M. J. *et al.* MAP17 Is a Necessary Activator of Renal Na+/Glucose Cotransporter SGLT2. *J. Am. Soc. Nephrol.* **28**, 85–93 (2017).
357. Shang, P. *et al.* The amino acid transporter SLC36A4 regulates the amino acid pool in retinal pigmented epithelial cells and mediates the mechanistic target of rapamycin, complex 1 signaling. *Aging Cell* **16**, 349–359 (2017).
358. Franco, M. *et al.* The Adaptor Protein Tom1L1 Is a Negative Regulator of Src Mitogenic Signaling Induced by Growth Factors. *Mol. Cell. Biol.* **26**, 1932–1947 (2006).
359. Heublein, S. *et al.* Proton-assisted amino-acid transporters are conserved regulators of proliferation and amino-acid-dependent mTORC1 activation. *Oncogene* **29**, 4068–4079 (2010).
360. Katoh, Y., Imakagura, H., Futatsumori, M. & Nakayama, K. Recruitment of clathrin onto endosomes by the Tom1-Tollip complex. *Biochem. Biophys. Res. Commun.* **341**, 143–149 (2006).
361. Wang, T., Liu, N. S., Seet, L. F. & Hong, W. The emerging role of VHS domain-containing Tom1, Tom1L1 and Tom1L2 in membrane trafficking. *Traffic* **11**, 1119–1128 (2010).
362. Franco, M. *et al.* The Adaptor Protein Tom1L1 Is a Negative Regulator of Src

Mitogenic Signaling Induced by Growth Factors. *Mol. Cell. Biol.* **26**, 1932 (2006).

363. Farhang-Fallah, J. *et al.* The Pleckstrin Homology (PH) Domain-Interacting Protein Couples the Insulin Receptor Substrate 1 PH Domain to Insulin Signaling Pathways Leading to Mitogenesis and GLUT4 Translocation. *Mol. Cell. Biol.* **22**, 7325 (2002).
364. Wang, A. L. *et al.* Role of FAM18B in diabetic retinopathy. *Mol. Vis.* **20**, 1146 (2014).
365. Gendre, D. *et al.* Conserved Arabidopsis ECHIDNA protein mediates trans-Golgi-network trafficking and cell elongation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 8048–8053 (2011).
366. Gendre, D. *et al.* Conserved Arabidopsis ECHIDNA protein mediates trans-Golgi-network trafficking and cell elongation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 8048–8053 (2011).
367. Stein, I. S., Gottfried, A., Zimmermann, J. & Fischer von Mollard, G. TVP23 interacts genetically with the yeast SNARE VTI1 and functions in retrograde transport from the early endosome to the late Golgi. *Biochem. J.* **419**, 229–236 (2009).
368. Jumpertz, R., Thearle, M. S., Bunt, J. C. & Krakoff, J. Assessment of Non-insulin Mediated Glucose Uptake: Association with Body Fat and Glycemic Status. *Metabolism.* **59**, 1396 (2010).
369. Lauritzen, H. P. M. M. & Schertzer, J. D. Measuring GLUT4 translocation in mature muscle fibers. *Am. J. Physiol. - Endocrinol. Metab.* **299**, 169–179 (2010).
370. Börgeson, E., Boucher, J. & Hagberg, C. E. Of mice and men: Pinpointing species differences in adipose tissue biology. *Frontiers in Cell and Developmental Biology* vol. 10 (2022).
371. Emont, M. P. *et al.* A single-cell atlas of human and mouse white adipose tissue. *Nature* **603**, 926–933 (2022).
372. Ahlqvist, E., Prasad, R. B. & Groop, L. Subtypes of Type 2 Diabetes Determined From Clinical Parameters. *Diabetes* **69**, 2086–2093 (2020).
373. Tobin, M. D., Minelli, C., Burton, P. R. & Thompson, J. R. Commentary: Development of Mendelian randomization: From hypothesis test to 'Mendelian deconfounding'. *International Journal of Epidemiology* vol. 33 26–29 (2004).
374. Smith, G. D. & Ebrahim, S. Mendelian randomization: Prospects, potentials,

and limitations. *International Journal of Epidemiology* vol. 33 30–42 (2004).

375. Day, N. *et al. EPIC-Norfolk: study design and characteristics of the cohort*. *BHIISh Journalof Cancer* http://www.srl.cam.ac.uk/epic/publications/day_bjc_1999_clean.pdf (1999).
376. Langenberg, C. *et al.* Design and cohort description of the InterAct Project: An examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. *Diabetologia* **54**, 2272–2282 (2011).
377. Pietzner, M. *et al.* Plasma metabolites to profile pathways in noncommunicable disease multimorbidity. *Nat. Med. 2021 273* **27**, 471–479 (2021).
378. Consortium, the H. R. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
379. Steen, J., Loeys, T., Moerkerke, B. & Vansteelandt, S. Medflex: An R package for flexible mediation analysis using natural effect models. *J. Stat. Softw.* **76**, 1–46 (2017).
380. Ritchie, S. C. *et al.* Quality control and removal of technical variation of NMR metabolic biomarker data in ~120,000 UK Biobank participants. *medRxiv* **9**, 2021.09.24.21264079 (2021).
381. Loh, P. R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet. 2018 507* **50**, 906–908 (2018).
382. Lotta, L. A. *et al.* A cross-platform approach identifies genetic regulators of human metabolism and health. *Nat. Genet. 2021 531* **53**, 54–64 (2021).
383. Wittemans, L. B. L. *et al.* Assessing the causal association of glycine with risk of cardio-metabolic diseases. *Nat. Commun. 2019 101* **10**, 1–13 (2019).
384. Wishart, D. S. *et al.* HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
385. Ritchie, S. C. *et al.* Integrative analysis of the plasma proteome and polygenic risk of cardiometabolic diseases. *Anna C. Calkin* **18**, 22.
386. Bragg, F. *et al.* Predictive value of circulating NMR metabolic biomarkers for type 2 diabetes risk in the UK Biobank study. *BMC Med.* **20**, 1–12 (2022).
387. Wurtz, P. *et al.* Branched-Chain and Aromatic Amino Acids Are Predictors of Insulin Resistance in Young Adults. *Diabetes Care* **36**, 648–655 (2013).
388. Liu, J. *et al.* Metabolomics based markers predict type 2 diabetes in a 14-year follow-up study. *Metabolomics* **13**, 1–11 (2017).
389. Adam, D. THE CAUSATION DETECTOR. *Nature* vol. 576 196–199 (2019).

390. Flannick, J. Data-driven type 2 diabetes patient clusters predict metabolic surgery outcomes. *The Lancet Diabetes and Endocrinology* vol. 10 150–151 (2022).
391. Zaghlool, S. B. *et al.* Metabolic and proteomic signatures of type 2 diabetes subtypes in an Arab population. *Nat. Commun.* **13**, 1–17 (2022).
392. Wesolowska-Andersen, A. *et al.* Four groups of type 2 diabetes contribute to the etiological and clinical heterogeneity in newly diagnosed individuals: An IMI DIRECT study. *Cell Reports Med.* **3**, (2022).
393. Deutsch, A. J., Ahlqvist, E. & Udler, M. S. Phenotypic and genetic classification of diabetes. *Diabetologia* vol. 65 1758–1769 (2022).
394. Wagner, R. *et al.* Pathophysiology-based subphenotyping of individuals at elevated risk for type 2 diabetes. *Nat. Med.* **27**, 49–57 (2021).
395. White, P. J. *et al.* Branched-chain amino acid restriction in Zucker-fatty rats improves muscle insulin sensitivity by enhancing efficiency of fatty acid oxidation and acyl-glycine export. *Mol. Metab.* **5**, 538–551 (2016).
396. Lutchmansingh, F. K. *et al.* Glutathione metabolism in type 2 diabetes and its relationship with microvascular complications and glycemia. **13**, e0198626 (2018).
397. Calabrese, V. *et al.* Oxidative stress, glutathione status, sirtuin and cellular stress response in type 2 diabetes. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1822**, 729–736 (2012).
398. Leonetti, S., Herzog, R. I., Caprio, S., Santoro, N. & Tricò, D. Glutamate–Serine–Glycine Index: A Novel Potential Biomarker in Pediatric Non-Alcoholic Fatty Liver Disease. *Children* **7**, 270 (2020).
399. Gaggini, M. *et al.* Altered amino acid concentrations in NAFLD: Impact of obesity and insulin resistance. *Hepatology* **67**, 145–158 (2018).
400. Mardinoglu, A. *et al.* Plasma Mannose Levels Are Associated with Incident Type 2 Diabetes and Cardiovascular Disease. *Cell Metab.* **26**, 281–283 (2017).
401. Lee, S. *et al.* Integrated Network Analysis Reveals an Association between Plasma Mannose Levels and Insulin Resistance. *Cell Metab.* **24**, 172–184 (2016).
402. Sone, H. *et al.* Physiological changes in circulating mannose levels in normal, glucose-intolerant, and diabetic subjects. *Metabolism.* **52**, 1019–1027 (2003).
403. Vanweert, F. *et al.* A randomized placebo-controlled clinical trial for pharmacological activation of BCAA catabolism in patients with type 2

diabetes. *Nat. Commun. 2022 131* **13**, 1–9 (2022).

404. Vanweert, F. *et al.* Elevated Plasma Branched-Chain Amino Acid Levels Correlate with Type 2 Diabetes-Related Metabolic Disturbances. *J. Clin. Endocrinol. Metab.* **106**, E1827–E1836 (2021).
405. Wurtz, P. *et al.* Branched-chain and aromatic amino acidsare predictors of insulinresistance in young adults. *Diabetes Care* **36**, 648–655 (2013).
406. Walford, G. A. *et al.* Metabolite profiles of diabetes incidence and intervention response in the diabetes prevention program. *Diabetes* **65**, 1424–1433 (2016).
407. Palmer, N. D. *et al.* Metabolomic Profile Associated With Insulin Resistance and Conversion to Diabetes in the Insulin Resistance Atherosclerosis Study. *J. Clin. Endocrinol. Metab.* **100**, E463–E468 (2015).
408. Huffman, K. M. *et al.* Relationships between circulating metabolic intermediates and insulin action in overweight to obese, inactive men and women. *Diabetes Care* **32**, 1678–1683 (2009).
409. Newgard, C. B. *et al.* A Branched-Chain Amino Acid-Related Metabolic Signature that Differentiates Obese and Lean Humans and Contributes to Insulin Resistance. *Cell Metab.* **9**, 311–326 (2009).
410. Lotta, L. A. *et al.* Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis. *PLoS Med.* **13**, (2016).
411. Mahajan, A. *et al.* Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat. Genet.* **50**, 559–571 (2018).
412. Zeggini, E., Gloyn, A. L., Barton, A. C. & Wain, L. V. Translational genomics and precision medicine: Moving from the lab to the clinic. *Science (80-.).* **365**, 1409–1413 (2019).
413. Udler, M. S., McCarthy, M. I., Florez, J. C. & Mahajan, A. Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine. *Endocr. Rev.* **40**, 1500–1520 (2019).
414. Lam, B. Y. H. *et al.* MC3R links nutritional state to childhood growth and the timing of puberty. *Nat. 2021 5997885* **599**, 436–441 (2021).
415. Stuart, T. *et al.* Comprehensive integration of single cell data. doi:10.1101/460147.
416. Harno, E., Ramamoorthy, T. G., Coll, A. P. & White, A. POMC: The physiological power of hormone processing. *Physiol. Rev.* **98**, 2381–2430

(2018).
417. Steiner, D. F. The proprotein convertases. *Curr. Opin. Chem. Biol.* **2**, 31–39 (1998).
418. Seidah, N. G., Rochemont, J., Hamelin, J., Benjannet, S. & Chretien, M. The missing fragment of the pro-sequence of human pro-opiomelanocortin: Sequence and evidence for C-terminal amidation. *Biochem. Biophys. Res. Commun.* **102**, 710–716 (1981).
419. Pritchard, L. E., Turnbull, A. V. & White, A. Pro-opiomelanocortin processing in the hypothalamus: impact on melanocortin signalling and obesity. *J. Endocrinol.* **172**, 411–421 (2002).
420. Butler, A. A. The melanocortin system and energy balance. *Peptides* vol. 27 281–290 (2006).
421. Benjannet, S., Rondeau, N., Day, R., Chrétien, M. & Seidah, N. G. PC1 and PC2 are proprotein convertases capable of cleaving proopiomelanocortin at distinct pairs of basic residues. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 3564–3568 (1991).
422. Birch, N. P., Estivariz, F. E., Bennett, H. P. J. & Lob, Y. P. Differential glycosylation of N-POMC1-77 regulates the production of gamma 3-MSH by purified pro-opiomelanocortin converting enzyme. A possible mechanism for tissue-specific processing. *FEBS Lett.* **290**, 191–194 (1991).
423. Glauder, J., Ragg, H., Rauch, J. & Engels, J. W. Human peptidylglycine α-amidating monooxygenase: cDNA, cloning and functional expression of a truncated form in COS cells. *Biochem. Biophys. Res. Commun.* **169**, 551–558 (1990).
424. Ollmann, M. M. *et al.* Antagonism of Central Melanocortin receptors in vitro and in vivo by agouti-related protein. *Science (80-.).* **278**, 135–138 (1997).
425. Shutter, J. R. *et al.* Hypothalamic expression of ART, a novel gene related to agouti, is up-regulated in obese and diabetic mutant mice. *Genes Dev.* **11**, 593–602 (1997).
426. Wilson, B. D. *et al.* Structure and function of ASP, the human homolog of the mouse agouti gene. *Hum. Mol. Genet.* **4**, 223–230 (1995).
427. Kwon, H. Y. *et al.* Molecular structure and chromosomal mapping of the human homolog of the agouti gene. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 9760–9764 (1994).
428. Smith, S. R. *et al.* Agouti Expression in Human Adipose TissueFunctional

Consequences and Increased Expression in Type 2 Diabetes. *Diabetes* **52**, 2914–2922 (2003).

429. Yeo, G. S. H. *et al.* The melanocortin pathway and energy homeostasis: From discovery to obesity therapy. *Mol. Metab.* **48**, 101206 (2021).
430. Fridmanis, D., Roga, A. & Klovins, J. ACTH receptor (MC2R) specificity: What do we know about underlying molecular mechanisms? *Front. Endocrinol. (Lausanne).* **8**, 13 (2017).
431. Ni, X.-P., Pearce, D., Butler, A. A., Cone, R. D. & Humphreys, M. H. Genetic disruption of γ-melanocyte–stimulating hormone signaling leads to salt-sensitive hypertension in the mouse. *J. Clin. Invest.* **111**, 1251–1258 (2003).
432. Haskell-Luevano, C., Holder, J. R., Monck, E. K. & Bauzo, R. M. Characterization of melanocortin NDP-MSH agonist peptide fragments at the mouse central and peripheral melanocortin receptors. *J. Med. Chem.* **44**, 2247–2252 (2001).
433. Sawyer, T. K. *et al.* 4-Norleucine, 7-D-phenylalanine-alpha-melanocyte-stimulating hormone: a highly potent alpha-melanotropin with ultralong biological activity. *Proc. Natl. Acad. Sci.* **77**, 5754–5758 (1980).
434. Wolf Horrell, E. M., Boulanger, M. C. & D'Orazio, J. A. Melanocortin 1 receptor: Structure, function, and regulation. *Front. Genet.* **7**, 95 (2016).
435. Rees, J. L. The Genetics of Sun Sensitivity in Humans. *Am. J. Hum. Genet.* **75**, 739–751 (2004).
436. Robbins, L. S. *et al.* Pigmentation phenotypes of variant extension locus alleles result from point mutations that alter MSH receptor function. *Cell* **72**, 827–834 (1993).
437. Chida, D. *et al.* Melanocortin 2 receptor is required for adrenal gland development, steroidogenesis, and neonatal gluconeogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18205–18210 (2007).
438. Chen, W. *et al.* Exocrine gland dysfunction in MC5-R-deficient mice: evidence for coordinated regulation of exocrine gland function by melanocortin peptides. *Cell* **91**, 789–798 (1997).
439. Cone, R. D. Anatomy and regulation of the central melanocortin system. **8**, (2005).
440. Yeo, G. S. H. Genetics of obesity: can an old dog teach us new tricks? *Diabetologia* vol. 60 778–783 (2017).
441. Oswal, A. & Yeo, G. S. H. The leptin melanocortin pathway and the control of

body weight: Lessons from human and murine genetics. *Obesity Reviews* vol. 8 293–306 (2007).

442. Cummings, D. E. & Schwartz, M. W. Melanocortins and body weight: a tale of two receptors. *Nat. Genet.* **26**, 8–9 (2000).
443. Cowley, M. A. *et al.* Leptin activates anorexigenic POMC neurons through a neural network in the arcuate nucleus. *Nature* **411**, 480–484 (2001).
444. Morton, G. J. & Schwartz, M. W. The NPY=AgRP neuron and energy homeostasis. *Int. J. Obes.* **25**, 56–62 (2001).
445. Varela, L. & Horvath, T. L. Leptin and insulin pathways in POMC and AgRP neurons that modulate energy balance and glucose homeostasis. *EMBO Rep.* **13**, 1079 (2012).
446. Xu, A. W., Ste-Marie, L., Kaelin, C. B. & Barsh, G. S. Inactivation of signal transducer and activator of transcription 3 in proopiomelanocortin (Pomc) neurons causes decreased Pomc expression, mild obesity, and defects in compensatory refeeding. *Endocrinology* **148**, 72–80 (2007).
447. Ernst, M. B. *et al.* Enhanced Stat3 Activation in POMC Neurons Provokes Negative Feedback Inhibition of Leptin and InsulinSignaling in Obesity. *J. Neurosci.* **29**, 11582–11593 (2009).
448. Gurzov, E. N., Stanley, W. J., Pappas, E. G., Thomas, H. E. & Gough, D. J. The JAK/STAT pathway in obesity and diabetes. *FEBS J.* **283**, 3002–3015 (2016).
449. Yeo, G. S. H. *et al.* Mutations in the human melanocortin-4 receptor gene associated with severe familial obesity disrupts receptor function through multiple molecular mechanisms. doi:10.1093/hmg/ddg057.
450. Eves, P. C. & Haycock, J. W. Melanocortin signalling mechanisms. *Adv. Exp. Med. Biol.* **681**, 19–28 (2010).
451. Hepler, J. R. & Gilman, A. G. G proteins. *Trends Biochem. Sci.* **17**, 383–387 (1992).
452. Podyma, B. *et al.* The stimulatory G protein G s α is required in melanocortin 4 receptor-expressing cells for normal energy balance, thermogenesis, and glucose metabolism. *J. Biol. Chem.* **293**, 10993–11005 (2018).
453. Glas, E., Mückter, H., Gudermann, T. & Breit, A. Exchange factors directly activated by cAMP mediate melanocortin 4 receptor-induced gene expression. *Sci. Rep.* **6**, (2016).
454. Brouwers, B. *et al.* Human MC4R variants affect endocytosis, trafficking and

dimerization revealing multiple cellular mechanisms involved in weight regulation. *Cell Rep.* **34**, 108862 (2021).

455. Jackson, R. S. *et al.* Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nat. Genet.* **16**, 303–306 (1997).
456. O'Rahilly, S. *et al.* Impaired Processing of Prohormones Associated with Abnormalities of Glucose Homeostasis and Adrenal Function. *N. Engl. J. Med.* **333**, 1386–1391 (1995).
457. Jackson, R. S. *et al.* Small-intestinal dysfunction accompanies the complex endocrinopathy of human proprotein convertase 1 deficiency. *J. Clin. Invest.* **112**, 1550–1560 (2003).
458. Farooqi, I. S. *et al.* Hyperphagia and early-onset obesity due to a novel homozygous missense mutation in prohormone convertase 1/3. in *Journal of Clinical Endocrinology and Metabolism* vol. 92 3369–3373 (Oxford Academic, 2007).
459. Frank, G. R. *et al.* Severe obesity and diabetes insipidus in a patient with PCSK1 deficiency. *Mol. Genet. Metab.* **110**, 191–194 (2013).
460. Nead, K. T. *et al.* Contribution of common non-synonymous variants in PCSK1 to body mass index variation and risk of obesity: a systematic review and meta-analysis with evidence from up to 331 175 individuals. *Hum. Mol. Genet.* **24**, 3582–3594 (2015).
461. Zhu, X. *et al.* Disruption of PC1/3 expression in mice causes dwarfism and multiple neuroendocrine peptide processing defects. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10293–10298 (2002).
462. Lloyd, D. J., Bohan, S. & Gekakis, N. Obesity, hyperphagia and increased metabolic efficiency in Pc1 mutant mice. *Hum. Mol. Genet.* **15**, 1884–1893 (2006).
463. Stijnen, P., Ramos-Molina, B., O'Rahilly, S. & Creemers, J. W. M. PCSK1 Mutations and Human Endocrinopathies: From Obesity to Gastrointestinal Disorders. *Endocr. Rev.* **37**, 347–371 (2016).
464. Blüher, M. & Mantzoros, C. S. From leptin to other adipokines in health and disease: Facts and expectations at the beginning of the 21st century. *Metabolism* **64**, 131–145 (2015).
465. Sadaf Farooqi, I. & O'Rahilly, S. 20 YEARS OF LEPTIN: Human disorders of leptin action. *J. Endocrinol.* **223**, T63–T70 (2014).

466. Ingalls, A. M., Dickie, M. M. & Snell, G. D. Obese, a new mutation in the house mouse. *J. Hered.* **41**, 317–318 (1950).
467. Coleman, D. L. Obese and diabetes: Two mutant genes causing diabetes-obesity syndromes in mice. *Diabetol. 1978 143* **14**, 141–148 (1978).
468. Zhang, Y. *et al.* Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425–432 (1994).
469. Chehab, F. F., Lim, M. E. & Lu, R. Correction of the sterility defect in homozygous obese female mice by treatment with the human recombinant leptin. *Nat. Genet.* **12**, 318–320 (1996).
470. Halaas, J. L. *et al.* Weight-reducing effects of the plasma protein encoded by the obese gene. *Science (80-.).* **269**, 543–546 (1995).
471. Pelleymounter, M. A. *et al.* Effects of the obese gene product on body weight regulation in ob/ob mice. *Science (80-.).* **269**, 540–543 (1995).
472. Mounzih, K., Lu, R. & Chehab, F. F. Leptin treatment rescues the sterility of genetically obese ob/ob males. *Endocrinology* **138**, 1190–1193 (1997).
473. Montague, C. T. *et al.* Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nat. 1997 3876636* **387**, 903–908 (1997).
474. Strobel, A., Issad, T., Camoin, L., Ozata, M. & Strosberg, A. D. A leptin missense mutation associated with hypogonadism and morbid obesity. *Nat. Genet. 1998 183* **18**, 213–215 (1998).
475. Chen, H. *et al.* Evidence that the diabetes gene encodes the leptin receptor: identification of a mutation in the leptin receptor gene in db/db mice. *Cell* **84**, 491–495 (1996).
476. Lee, G. H. *et al.* Abnormal splicing of the leptin receptor in diabetic mice. *Nature* **379**, 632–635 (1996).
477. Clément, K. *et al.* A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nat. 1998 3926674* **392**, 398–401 (1998).
478. Farooqi, I. S. *et al.* Clinical and Molecular Genetic Spectrum of Congenital Deficiency of the Leptin Receptor. *N. Engl. J. Med.* **356**, 237 (2007).
479. Lahlou, N. *et al.* Soluble leptin receptor in serum of subjects with complete resistance to leptin: relation to fat mass. *Diabetes* **49**, 1347–1352 (2000).
480. Krude, H. *et al.* Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nat. Genet. 1998 192* **19**, 155–157 (1998).
481. Yaswen, L., Diehl, N., Brennan, M. B. & Hochgeschwender, U. Obesity in the

mouse model of pro-opiomelanocortin deficiency responds to peripheral melanocortin. *Nat. Med. 1999 59* **5**, 1066–1070 (1999).

482. Smart, J. L. & Low, M. J. Lack of proopiomelanocortin peptides results in obesity and defective adrenal function but normal melanocyte pigmentation in the murine C57BL/6 genetic background. *Ann. N. Y. Acad. Sci.* **994**, 202–210 (2003).
483. Challis, B. G. *et al.* Mice lacking pro-opiomelanocortin are sensitive to high-fat feeding but respond normally to the acute anorectic effects of peptide-YY3-36. *Proc. Natl. Acad. Sci.* **101**, 4695–4700 (2004).
484. Farooqi, I. S. *et al.* Dominant and recessive inheritance of morbid obesity associate with melanocortin 4 receptor deficiency. *J. Clin. Invest.* **106**, 271–279 (2000).
485. Yeo, G. S. H. *et al. A frameshift mutation in MC4R associated with dominantly inherited human obesity [1]*. *Nature Genetics* vol. 20 111–112 (1998).
486. Huszar, D. *et al.* Targeted Disruption of the Melanocortin-4 Receptor Results in Obesity in Mice. *Cell* **88**, 131–141 (1997).
487. Vaisse, C., Clement, K., Guy-Grand, B. & Froguel, P. A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nat. Genet. 1998 202* **20**, 113–114 (1998).
488. Lotta, L. A. *et al.* Human Gain-of-Function MC4R Variants Show Signaling Bias and Protect against Obesity. *Cell* **177**, 597-607.e9 (2019).
489. Wade, K. H. *et al.* Loss-of-function mutations in the melanocortin 4 receptor in a UK birth cohort. *Nat. Med.* **27**, 1088–1096 (2021).
490. Chen, A. S. *et al.* Inactivation of the mouse melanocortin-3 receptor results in increased fat mass and reduced lean body mass. *Nat. Genet.* **26**, 97–102 (2000).
491. Begriche, K. *et al.* Genetic dissection of the functions of the melanocortin-3 receptor, a seven-transmembrane G-protein-coupled receptor, suggests roles for central and peripheral receptors in energy homeostasis. *J. Biol. Chem.* **286**, 40771–40781 (2011).
492. Butler, A. A. *et al.* A unique metabolic syndrome causes obesity in the melanocortin-3 receptor-deficient mouse. *Endocrinology* **141**, 3518–3521 (2000).
493. Ghamari-Langroudi, M. *et al.* Regulation of energy rheostasis by the melanocortin-3 receptor. *Sci. Adv.* **4**, 866–888 (2018).

494. Abbott, C. R. *et al.* Investigation of the melanocyte stimulating hormones on food intake: Lack of evidence to support a role for the melanocortin-3-receptor. *Brain Res.* **869**, 203–210 (2000).
495. Demidowich, A. P., Jun, J. Y. & Yanovski, J. A. Polymorphisms and mutations in the melanocortin-3 receptor and their relation to human obesity. *Biochimica et Biophysica Acta - Molecular Basis of Disease* vol. 1863 2468–2476 (2017).
496. Yang, Z. & Tao, Y. X. Mutations in Melanocortin-3 Receptor Gene and Human Obesity. in *Progress in Molecular Biology and Translational Science* vol. 140 97–129 (Elsevier B.V., 2016).
497. Girardet, C., Begriche, K., Ptitsyn, A., Koza, R. A. & Butler, A. A. Unravelling the mysterious roles of melanocortin-3 receptors in metabolic homeostasis and obesity using mouse genetics. *Int. J. Obes. Suppl.* **4**, S37–S44 (2014).
498. Lee, B. *et al.* A mouse model for a partially inactive obesity-associated human MC3R variant. *Nat. Commun.* **7**, 1–14 (2016).
499. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700 000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
500. Day, F. R. *et al.* Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat. Genet.* **49**, 834–841 (2017).
501. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013).
502. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
503. S, P. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
504. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
505. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
506. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.* **12**, (2015).
507. Eastwood, S. V. *et al.* Algorithms for the capture and adjudication of prevalent and incident diabetes in UK Biobank. *PLoS One* **11**, e0162388 (2016).

508. Powell, R. *et al.* Development and validation of total and regional body composition prediction equations from anthropometry and single frequency segmental bioelectrical impedance with DEXA. *medRxiv* 2020.12.16.20248330 (2020) doi:10.1101/2020.12.16.20248330.
509. Millard, L. A. C., Davies, N. M., Gaunt, T. R., Smith, G. D. & Tilling, K. Software application profile: PHESANT: A tool for performing automated phenome scans in UK Biobank. *Int. J. Epidemiol.* **47**, 29–35 (2018).
510. Kamat, M. A. *et al.* PhenoScanner V2: An expanded tool for searching human genotype-phenotype associations. *Bioinformatics* **35**, 4851–4853 (2019).
511. Finer, S. *et al.* Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. *Int. J. Epidemiol.* **49**, 20-21l (2020).
512. Pietzner, M. *et al.* Genetic architecture of host proteins involved in SARS-CoV-2 infection. *Nat. Commun. 2020 111* **11**, 1–14 (2020).
513. Gardner, E. J. *et al.* Damaging missense variants in IGF1R implicate a role for IGF-1 resistance in the aetiology of type 2 diabetes. *medRxiv* 2022.03.26.22272972 (2022) doi:10.1101/2022.03.26.22272972.
514. Melvin, A. *et al.* SUN-LB107 Functional Characterisation of Human Heterozygous Non-Synonymous MC3R Variants. *J. Endocr. Soc.* **4**, (2020).
515. Khaw, K. T. *et al.* Combined Impact of Health Behaviours and Mortality in Men and Women: The EPIC-Norfolk Prospective Population Study. *PLOS Med.* **5**, e12 (2008).
516. De Onis, M. *et al.* Development of a WHO growth reference for school-aged children and adolescents. *Bull. World Health Organ.* **85**, 660–667 (2007).
517. Brown, P. I. & Brasel, J. in. *The Malnourished Child: v. 19 (Nestle Nutrition Workshop Series)*. (Nestlé Nutrition Institute and Vevey/Raven Press, 1990).
518. Kühnen, P. *et al.* Proopiomelanocortin Deficiency Treated with a Melanocortin-4 Receptor Agonist. *N. Engl. J. Med.* **375**, 240–246 (2016).
519. Liu, D. J. *et al.* Exome-wide association study of plasma lipids in >300,000 individuals. *Nat. Genet.* **49**, 1758–1766 (2017).
520. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
521. Khetan, S. *et al.* Functional characterization of thousands of type 2 diabetes-associated and chromatin-modulating variants under steady state and endoplasmic reticulum stress. *bioRxiv* 2020.02.12.939348 (2020)

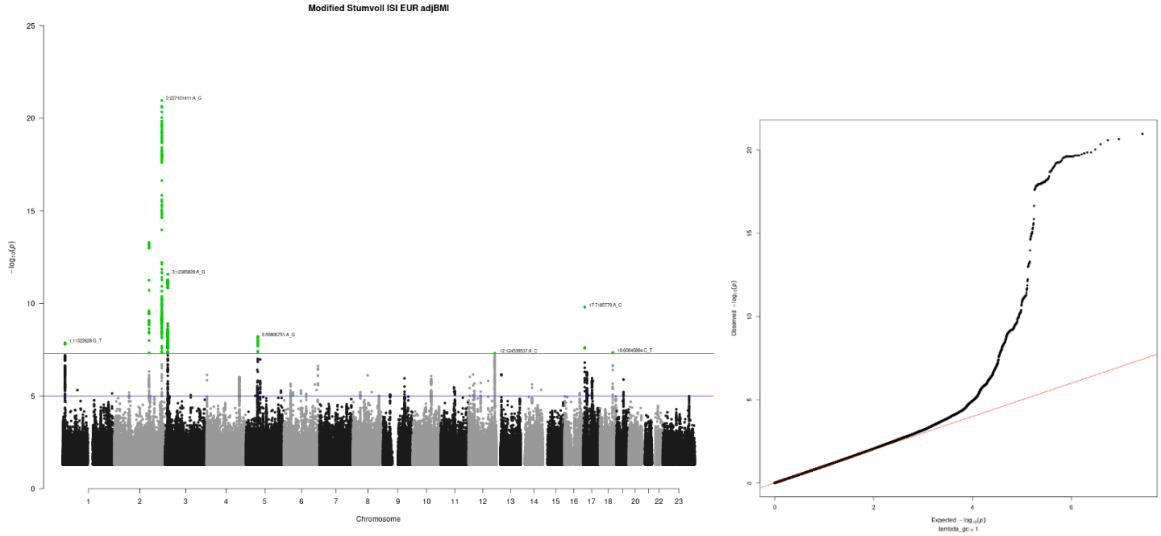
doi:10.1101/2020.02.12.939348.

522. Manousaki, D. *et al.* Toward precision medicine: TBC1D4 disruption is common among the inuit and leads to underdiagnosis of type 2 diabetes. *Diabetes Care* **39**, 1889–1895 (2016).
523. Morze, J. *et al.* Metabolomics and Type 2 Diabetes Risk: An Updated Systematic Review and Meta-analysis of Prospective Cohort Studies. *Diabetes Care* **45**, 1013–1024 (2022).
524. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016).
525. Claussnitzer, M. & Susztak, K. Gaining insight into metabolic diseases from human genetic discoveries. *Trends in Genetics* vol. 37 1081–1094 (2021).
526. Claussnitzer, M. *et al.* A brief history of human disease genetics. *Nature* vol. 577 179–189 (2020).
527. Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *American Journal of Human Genetics* vol. 102 717–730 (2018).
528. Cano-Gamez, E. & Trynka, G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Frontiers in Genetics* vol. 11 424 (2020).
529. Thorpe, L. E. *et al.* Neighborhood Socioeconomic Environment and Risk of Type 2 Diabetes: Associations and Mediation Through Food Environment Pathways in Three Independent Study Samples. *Diabetes Care* **45**, 798–810 (2022).
530. Neuenschwander, M. *et al.* Role of diet in type 2 diabetes incidence: Umbrella review of meta-analyses of prospective observational studies. *The BMJ* vol. 366 (2019).
531. Hill-Briggs, F. *et al.* Social determinants of health and diabetes: A scientific review. *Diabetes Care* vol. 44 258–279 (2021).
532. Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021).
533. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018).
534. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: Beyond biomarkers and towards mechanisms. *Nature Reviews Molecular Cell Biology* vol. 17 451–459 (2016).

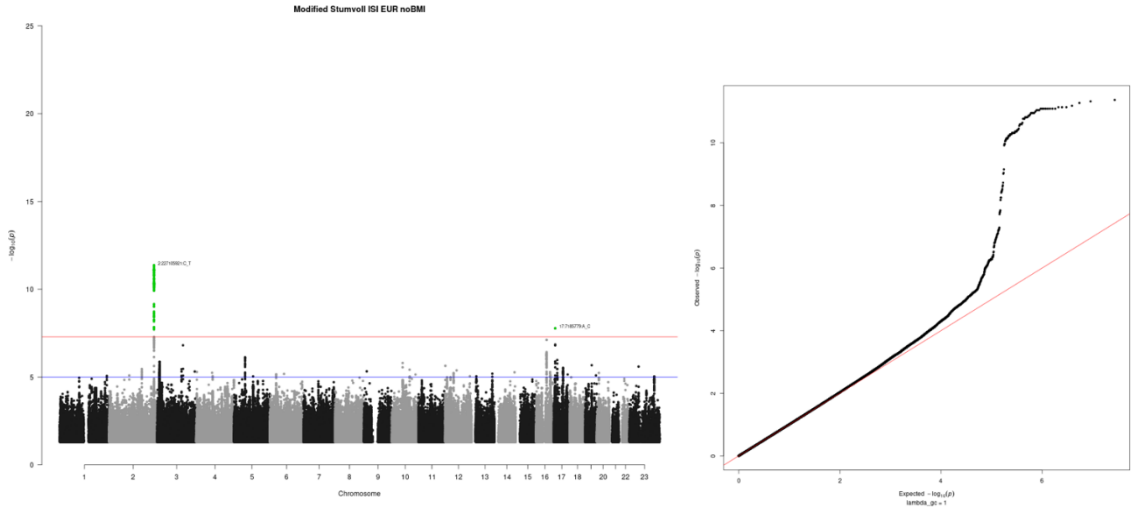
535. Lotta, L. A. *et al.* Integrative genomic analysis implicates limited peripheral adiposestorage capacity in the pathogenesis of human insulin resistance. *Nat. Genet.* **49**, 17 (2017).
536. Lier, N., Gresko, N., Di Chiara, M., Loffing-Cueni, D. & Loffing, J. Immunofluorescent localization of the Rab-GAP protein TBC1D4 (AS160) in mouse kidney. *Histochem. Cell Biol.* **138**, 101–112 (2012).
537. Sinnott-Armstrong, N. *et al.* Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.* **53**, 185–194 (2021).
538. Kim, H. *et al.* High-throughput genetic clustering of type 2 diabetes loci reveals heterogeneous mechanistic pathways of metabolic disease. *Diabetologia* **66**, 495–507 (2022).
539. Nedachi, T. & Kanzaki, M. Regulation of glucose transporters by insulin and extracellular glucose in C2C12 myotubes. *Am. J. Physiol. - Endocrinol. Metab.* **291**, (2006).
540. Al-Khalili, L. *et al.* Insulin action in cultured human skeletal muscle cells during differentiation: Assessment of cell surface GLUT4 and GLUT1 content. *Cell. Mol. Life Sci.* **60**, 991–998 (2003).
541. Chanon, S. *et al.* Glucose uptake measurement and response to insulin stimulation in in vitro cultured human primary myotubes. *J. Vis. Exp.* **2017**, (2017).
542. Kundu, K. *et al.* Genetic associations at regulatory phenotypes improve fine-mapping of causal variants for 12 immune-mediated diseases. *Nat. Genet. 2022 543* **54**, 251–262 (2022).
543. Yu, F. *et al.* Variant to function mapping at single-cell resolution through network propagation. *Nat. Biotechnol.* **40**, 1644–1653 (2022).
544. Wang, G. *et al.* Integration of single-cell multiomic measurements across disease states with genetics identifies mechanisms of beta cell dysfunction in type 2 diabetes. *bioRxiv* 2022.12.31.522386 (2023) doi:10.1101/2022.12.31.522386.

Appendix

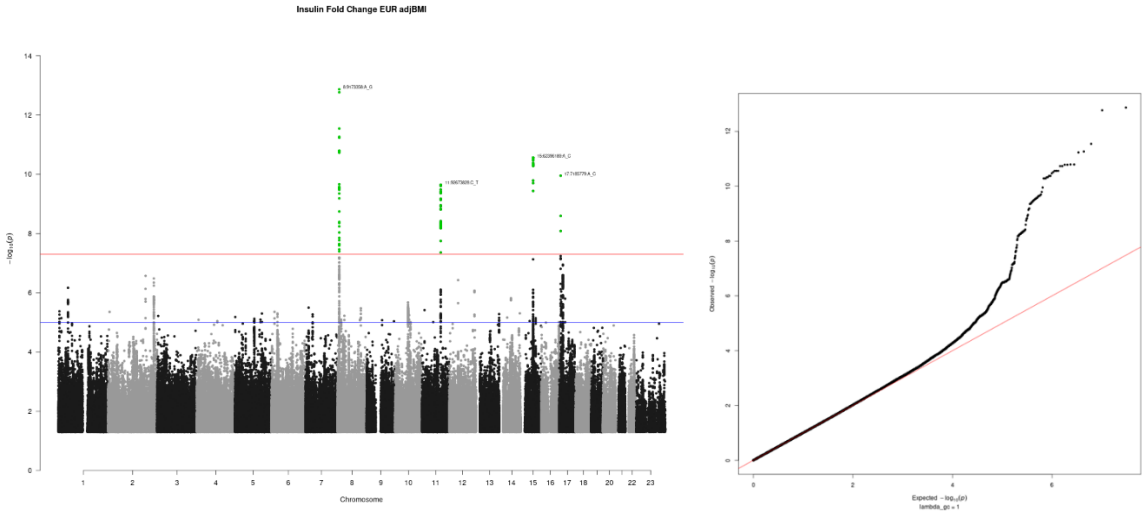
Supplementary Material: Chapter 2



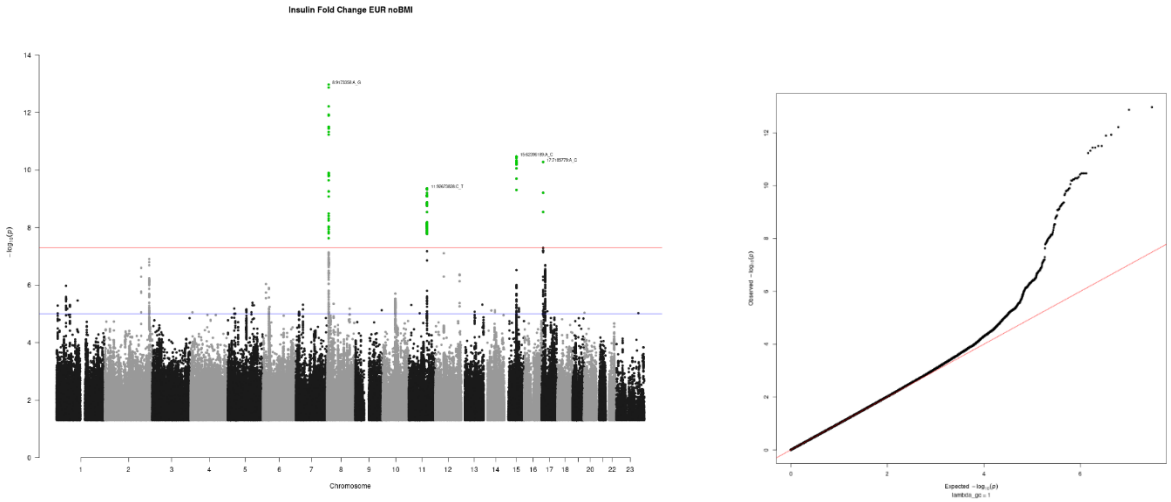
Supplementary Figure S2.1: Manhattan and QQ plots for European-only meta-analysis of Modified Stumvoll ISI adjusted for BMI



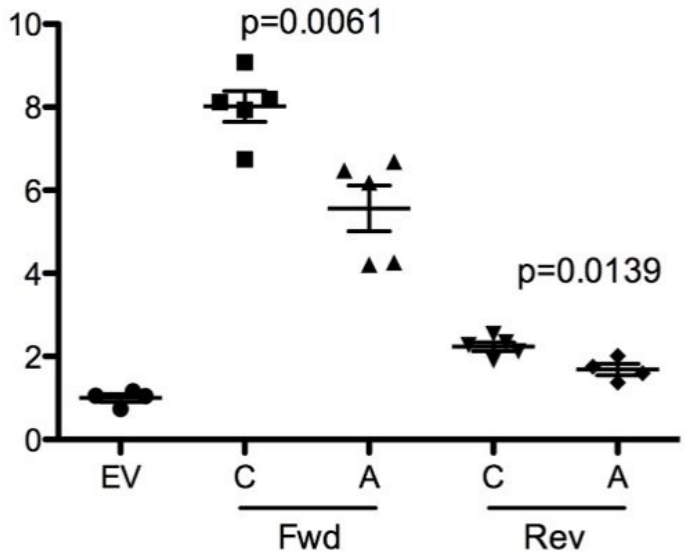
Supplementary Figure S2.2: Manhattan and QQ plots for European-only meta-analysis of Modified Stumvoll ISI unadjusted for BMI



Supplementary Figure S2.3: Manhattan and QQ plots for European-only meta-analysis of insulin fold change adjusted for BMI



Supplementary Figure S2.4: Manhattan and QQ plots for European-only meta-analysis of insulin fold change unadjusted for BMI



Supplementary Figure S2.5: rs117643180 exhibits allelic differences in transcriptional activity

229-bp fragments flanking rs117643180-C or rs117643180-A were cloned upstream of a minimal promoter driving luciferase expression in the forward and reverse orientations with respect to the promoter. Values represent fold-change of firefly luciferase/*Renilla* activity normalized to empty pGL4.23 vector in undifferentiated LHCN-M2 myoblasts. Error bars represent the SEM of four or five independent clones tested in duplicate wells. P-values are calculated from two-sided t-tests.

Supplementary Table S2.1: Association of Genetic Risk scores for IFC and MS ISI, and their comprising SNPs with relevant traits in UK Biobank (Unrelated individuals of European ancestry).

Trait	Estimate	SE	pval	GRS
iDEXA legs fat mass	0.339518511	0.018558	9.93E-75	ISI_adjBMI
iDEXA legs fat mass	0.459920326	0.026256	1.15E-68	ISI_noBMI
iDEXA gynoid fat mass	0.327922733	0.018978	7.20E-67	ISI_adjBMI
iDEXA gynoid fat mass	0.442930622	0.026849	4.08E-61	ISI_noBMI
IDEXA body fat %	0.24980749	0.01662	4.80E-51	ISI_adjBMI
IDEXA body fat %	0.338973847	0.023513	4.22E-47	ISI_noBMI
Type 2 Diabetes	-0.280019554	0.020261	1.97E-43	ISI_adjBMI
iDEXA total fat mass	0.283862195	0.020683	7.44E-43	ISI_adjBMI
Type 2 Diabetes	-0.384255918	0.028666	5.83E-41	ISI_noBMI
iDEXA arms fat mass	0.255545727	0.019474	2.50E-39	ISI_adjBMI
iDEXA total fat mass	0.378575788	0.029261	2.82E-38	ISI_noBMI
iDEXA subcutaneous adipose	0.252606869	0.020025	1.79E-36	ISI_adjBMI
WHRadjBMI (ivn transformed)	-0.253446248	0.020516	4.74E-35	ISI_adjBMI
iDEXA arms fat mass	0.339900626	0.027551	5.82E-35	ISI_noBMI
WHRadjBMI (ivn transformed)	-0.354237499	0.029026	3.00E-34	ISI_noBMI
iDEXA subcutaneous adipose	0.337159684	0.02833	1.19E-32	ISI_noBMI
HbA1C	-0.228787256	0.020508	6.78E-29	ISI_adjBMI
iDEXA trunk fat mass	0.227493932	0.020619	2.67E-28	ISI_adjBMI
HbA1C	-0.316158906	0.029019	1.23E-27	ISI_noBMI
iDEXA android fat mass	0.213487277	0.020281	6.57E-26	ISI_adjBMI
iDEXA trunk fat mass	0.299588552	0.029171	9.70E-25	ISI_noBMI
weight	0.185086275	0.018479	1.30E-23	ISI_adjBMI
iDEXA android fat mass	0.28039961	0.028692	1.49E-22	ISI_noBMI
weight	0.242081393	0.026143	2.06E-20	ISI_noBMI
HbA1C	0.24416009	0.027252	3.28E-19	IFC_adjBMI
Random Glucose	-0.194254409	0.021796	5.02E-19	ISI_adjBMI
HbA1C	0.240931825	0.027208	8.40E-19	IFC_noBMI
Random Glucose	-0.262247519	0.030833	1.82E-17	ISI_noBMI
BMI	0.172332034	0.020403	3.02E-17	ISI_adjBMI
iDEXA visceral adipose	0.129720198	0.016693	7.81E-15	ISI_adjBMI
BMI	0.221300261	0.028867	1.78E-14	ISI_noBMI
WHR (ivn transformed)	-0.206526329	0.028988	1.05E-12	ISI_noBMI
iDEXA visceral adipose	0.166030606	0.023616	2.06E-12	ISI_noBMI
WHR (ivn transformed)	-0.140742377	0.02049	6.48E-12	ISI_adjBMI
Type 2 Diabetes	0.17282906	0.026936	1.40E-10	IFC_adjBMI
Type 2 Diabetes	0.171263015	0.026893	1.91E-10	IFC_noBMI
iDEXA arms lean mass	0.082228702	0.014077	5.18E-09	IFC_adjBMI
iDEXA arms lean mass	0.08191603	0.014054	5.59E-09	IFC_noBMI
WHRadjBMI (ivn transformed)	-0.158153952	0.027274	6.69E-09	IFC_adjBMI
WHRadjBMI (ivn transformed)	-0.157048204	0.027231	8.06E-09	IFC_noBMI
height	0.082148782	0.014288	8.96E-09	ISI_adjBMI
height	0.115352349	0.020214	1.15E-08	ISI_noBMI

Supplementary Table S2.2: Association of Genetic Risk scores for IFC and MS ISI, and their comprising SNPs with relevant traits in Fenland (Unrelated Europeans).
P < 0.05

Trait Name	Estimate	SE	pval	GRS
2 h Insulin	-0.919228404	0.174771	1.48E-07	MS_ISI_noBMI
2 h Insulin	-0.648625471	0.125712	2.53E-07	MS_ISI_adjBMI
2 h Glucose	-0.73299355	0.17213	2.08E-05	MS_ISI_noBMI
2 h Glucose	-0.484146177	0.123822	9.30E-05	MS_ISI_adjBMI
HDL	0.488777472	0.135606	0.000315	FCI_noBMI
HDL	0.486870212	0.136533	0.000365	FCI_adjBMI
Total cholesterol	0.500771054	0.145154	0.000563	FCI_noBMI
Total cholesterol	0.499852645	0.146145	0.000629	FCI_adjBMI
Fasting glucose	0.491135966	0.145368	0.000732	FCI_adjBMI
Fasting glucose	0.479471089	0.144386	0.000902	FCI_noBMI
LDL	0.481949008	0.146694	0.001022	FCI_noBMI
LDL	0.481591976	0.147692	0.001115	FCI_adjBMI
2 h Glucose	0.487221239	0.149614	0.001132	FCI_noBMI
2 h Glucose	0.48691039	0.150634	0.001232	FCI_adjBMI
ApoA1	0.453689823	0.153238	0.00308	FCI_noBMI
Adiponectin	0.355541548	0.120652	0.003221	MS_ISI_adjBMI
ApoA1	0.452266179	0.15425	0.003378	FCI_adjBMI
NEFAs	0.462920789	0.161587	0.004184	FCI_noBMI
HDL	0.320070078	0.112521	0.004458	MS_ISI_adjBMI
NEFAs	0.461166167	0.162653	0.004591	FCI_adjBMI
BMI	0.349324709	0.123524	0.004695	MS_ISI_adjBMI
HDL	0.422043429	0.156438	0.006993	MS_ISI_noBMI
Adiponectin	0.448899754	0.167729	0.00746	MS_ISI_noBMI
Leptin	0.325976741	0.121828	0.007474	MS_ISI_adjBMI
BMI	0.456344057	0.17174	0.007894	MS_ISI_noBMI
Fasting Insulin	-0.414430908	0.162575	0.010819	FCI_noBMI
Fasting Insulin	-0.416072172	0.163646	0.011027	FCI_adjBMI
Fasting Insulin	-0.474700833	0.187019	0.011162	MS_ISI_noBMI
Fasting Insulin	-0.3379293	0.134475	0.011994	MS_ISI_adjBMI
ALT	-0.404967268	0.162489	0.012712	MS_ISI_noBMI
Leptin	0.420084403	0.16936	0.013145	MS_ISI_noBMI
ALT	-0.27986615	0.116884	0.01667	MS_ISI_adjBMI
2 h Insulin	0.318476085	0.153016	0.037435	FCI_adjBMI
2 h Insulin	0.315345343	0.151975	0.03802	FCI_noBMI
ApoA1	0.250950067	0.126651	0.047582	MS_ISI_adjBMI

Supplementary Table S2.3: LD score regression results for IFC adjBMI with cardiometabolic traits in cohorts of European ancestry

Genetic correlations were calculated using python software ldsc (P < 0.05). p2 = trait 2, rg = genetic correlation, se = standard error of rg, p = p-value for rg; h2_obs, h2_obs_se = observed scale h2 for trait 2 and standard error, h2_int, h2_int_se = single-trait LD Score regression intercept for trait 2 and standard error, gcov_int, gcov_int_se = cross-trait LD Score regression intercept and standard error

Phenotype 2 (phenotype 1 is IFC adjBMI)	p2 source of LD score *	rg	se	z	p	h2_obs	h2_obs_se	h2_int	h2_int_se	gcov_int	gcov_int_se
IFC_adjBMI	This Study	1	4.58E-06	2.18E+05	0	0.06222	0.009609	0.9721	0.006018	0.9721	0.006018
ISI_adjBMI	This Study	-0.6632	0.06828	-9.714	2.63E-22	0.1116	0.01276	0.9604	0.007132	-0.4262	0.005216
Glu120_adjBMI	Chen et al (2021) EUR	0.7133	0.0893	7.988	1.38E-15	0.1136	0.01409	0.9887	0.007101	0.2608	0.004889
T2D_MVP	Vujkovic et al (2020) EUR	0.2959	0.04379	6.758	1.40E-11	0.07123	0.003362	1.121	0.02979	0.006067	0.005891
UKB_460K.bp_SYSTOLICadjMED		0.2122	0.03743	5.668	1.44E-08	0.2071	0.008566	1.155	0.02788	-0.00602	0.00613
UKB_460K.disease_DIABETES_ANY_DIAGNOSED		0.3142	0.06174	5.088	3.61E-07	0.04401	0.002673	1.056	0.01221	-0.00135	0.005446
UKB_460K.disease_T2D		0.3174	0.06281	5.053	4.35E-07	0.04267	0.002578	1.047	0.01254	-0.00287	0.005493
UKB_460K.bp_DIASTOLICadjMED		0.1966	0.03961	4.962	6.96E-07	0.2045	0.008715	1.178	0.02771	-0.00956	0.006346
UKB_460K.biochemistry_HbA1c		0.2053	0.04582	4.48	7.46E-06	0.168	0.01593	1.249	0.09107	0.001746	0.006255
UKB_460K.disease_CARDIOVASCULAR		0.1931	0.04393	4.396	1.10E-05	0.104	0.004674	1.109	0.02399	-0.00046	0.005727
UKB_460K.body_WHRadjBMI		0.1961	0.04504	4.353	1.34E-05	0.1484	0.008363	1.152	0.03545	0.00275	0.006665
UKB_460K.disease_ENDOCRINE_DIABETES		0.3024	0.06981	4.331	1.48E-05	0.03377	0.002074	1.043	0.01063	-0.00322	0.005257
UKB_460K.biochemistry_AlanineAminotransferase		0.2056	0.04885	4.21	2.56E-05	0.1097	0.008799	1.07	0.02326	0.005036	0.005436
WHRadjBMI		0.2031	0.05068	4.007	6.14E-05	0.1197	0.007113	1.138	0.03716	0.005679	0.007323
PASS_Type_2_Diabetes		0.4415	0.1108	3.984	6.79E-05	0.09308	0.01036	1.014	0.008287	-0.00302	0.006019

UKB_460K.biochemistry_TotalProtein		0.1736	0.04565	3.802	0.000143	0.1467	0.008762	1.229	0.04232	-0.00113	0.005841
PASS_Coronary_Artery_Disease		0.3611	0.1022	3.533	0.00041	0.07323	0.009442	1.025	0.008843	-0.01232	0.005776
UKB_460K.biochemistry_Creatinine		-0.1356	0.03871	-3.504	0.000459	0.1974	0.01332	1.194	0.05108	0.01223	0.006492
UKB_460K.disease_HYPERTENSION_DIAGNOSED		0.1482	0.04292	3.453	0.000555	0.1244	0.005213	1.11	0.02313	-0.00254	0.005928
UKB_460K.biochemistry_GammaGlutamyltransferase		0.1571	0.05047	3.113	0.001853	0.1277	0.01357	1.16	0.05391	0.007209	0.005929
FI_adjBMI	Chen et al (2021) EUR	0.2981	0.0963	3.096	0.001962	0.1011	0.01004	0.9785	0.0111	-0.02343	0.006258
UKB_460K.biochemistry_Calcium		0.1337	0.04521	2.957	0.003109	0.1302	0.01607	1.085	0.03282	0.002742	0.005415
UKB_460K.biochemistry_AlkalinePhosphatase		0.165	0.05776	2.858	0.004268	0.2622	0.04668	1.35	0.1347	-0.01568	0.008481
UKB_460K.biochemistry_AspartateAminotransferase		0.1118	0.04525	2.471	0.01349	0.108	0.008457	1.097	0.02716	0.004262	0.005828
UKB_460K.biochemistry_Albumin		0.1266	0.05382	2.352	0.01868	0.1326	0.008282	1.157	0.03995	0.004615	0.006098
UKB_460K.biochemistry_Glucose		0.136	0.05853	2.324	0.02015	0.06313	0.01037	1.05	0.02351	-0.00435	0.004999
UKB_460K.biochemistry_Triglycerides		0.1231	0.05428	2.268	0.02332	0.1825	0.03048	1.123	0.07744	0.01084	0.00731
UKB_460K.biochemistry_SHBG		-0.1278	0.05662	-2.258	0.02395	0.2303	0.0509	1.175	0.08969	-0.0177	0.007051
UKB_460K.body_BMIz		-0.07826	0.03695	-2.118	0.03418	0.2741	0.008921	1.09	0.02347	-0.00515	0.005737
UKB_460K.biochemistry_CreactiveProtein		0.1119	0.05469	2.047	0.04069	0.09395	0.01886	1.126	0.0912	-0.00441	0.005392

(*) if source of LD score is not indicated in column this has been obtained from precomputed LD Scores provided by LDSC (Alkes Lab; see methods). Otherwise, the reference provided is the GWAS summary statistics used to calculate the LD scores using LDSC.

Supplementary Table S2.4: LD score regression results for ISI adjBMI with cardiometabolic traits in cohorts of European ancestry

Genetic correlations were calculated using python software ldsc (P < 0.05).p2 = trait 2, rg = genetic correlation, se = standard error of rg, p = p-value for rg; h2_obs, h2_obs_se = observed scale h2 for trait 2 and standard error, h2_int, h2_int_se = single-trait LD Score regression intercept for trait 2 and standard error, gcov_int, gcov_int_se = cross-trait LD Score regression intercept and standard error

p2	p2 source of LD score *	rg	se	z	p	h2_obs	h2_obs_se	h2_int	h2_int_se	gcov_int	gcov_int_se
ISI_adjBMI	This Study	1	4.83E-06	2.07E+05	0	0.1118	0.01271	0.9603	0.007094	0.9603	0.007094
FI_adjBMI	Chen et al (2021)	-0.8337	0.0471	-17.7	4.06E-70	0.1012	0.01006	0.9783	0.01109	-0.1948	0.007066
Glu120_adjBMI	Chen et al (2021)	-0.82	0.04801	-17.08	2.12E-65	0.1137	0.01409	0.9886	0.007079	-0.3643	0.005478
WHRadjBMI	Pulit et al (2019)	-0.4456	0.03778	-11.79	4.27E-32	0.1198	0.007118	1.137	0.03714	-0.01436	0.008813
UKB_460K.body_WHRadjBMI		-0.4241	0.03917	-10.83	2.51E-27	0.1485	0.008355	1.151	0.03539	-0.01244	0.008659
T2D_MVP	Vujkovic et al (2020)	-0.3912	0.03746	-10.44	1.59E-25	0.07125	0.003362	1.121	0.02981	-0.02598	0.008125
UKB_460K.bp_SYSTOLICadjMED		-0.3109	0.02982	-10.43	1.84E-25	0.2071	0.008563	1.155	0.02783	0.00118	0.007158
IFC_adjBMI	This Study	-0.6632	0.06828	-9.714	2.63E-22	0.06224	0.009619	0.9721	0.006025	-0.4262	0.005216
UKB_460K.biochemistry_AlanineAminotransferase		-0.3571	0.04109	-8.691	3.59E-18	0.1097	0.008801	1.07	0.02325	-0.01263	0.006897
UKB_460K.disease_T2D		-0.3993	0.04656	-8.576	9.80E-18	0.04268	0.002584	1.047	0.01257	-0.00665	0.005745
UKB_460K.disease_DIABETES_ANY_DIAGNOSED		-0.3916	0.04633	-8.451	2.89E-17	0.04402	0.002678	1.056	0.01224	-0.00974	0.005921
UKB_460K.bp_DIASTOLICadjMED		-0.2698	0.03267	-8.258	1.48E-16	0.2045	0.008707	1.178	0.02759	0.001327	0.007686
UKB_460K.biochemistry_HbA1c		-0.2727	0.03493	-7.806	5.88E-15	0.1681	0.01592	1.248	0.09095	-0.01012	0.007497
UKB_460K.disease_ENDOCRINE_DIABETES		-0.398	0.05171	-7.697	1.39E-14	0.03376	0.002076	1.043	0.01065	-0.00287	0.005689
UKB_460K.biochemistry_Glucose		-0.2992	0.04038	-7.41	1.27E-13	0.06316	0.01037	1.05	0.02349	-0.00604	0.00569
UKB_460K.disease_CARDIOVASCULAR		-0.2597	0.03607	-7.199	6.05E-13	0.104	0.004662	1.109	0.02391	-0.00418	0.006576

UKB_460K.biochemistry_HDLcholesterol		0.2606	0.03899	6.683	2.33E-11	0.2454	0.02614	1.513	0.273	0.01942	0.008682
UKB_460K.biochemistry_GammaGlutamyltransferase		-0.2444	0.03734	-6.545	5.95E-11	0.1278	0.01357	1.159	0.05381	3.51E-05	0.005802
UKB_460K.body_BMIz		0.1882	0.03073	6.123	9.16E-10	0.2743	0.008926	1.089	0.02346	0.003199	0.006753
UKB_460K.biochemistry_TotalProtein		-0.2205	0.03666	-6.016	1.79E-09	0.1468	0.00875	1.229	0.04233	-0.01017	0.006589
UKB_460K.disease_HYPERTENSION_DIAGNOSED		-0.2079	0.03504	-5.934	2.96E-09	0.1244	0.005198	1.11	0.02306	-0.00328	0.006936
UKB_460K.biochemistry_Triglycerides		-0.2688	0.04602	-5.841	5.18E-09	0.1825	0.03047	1.123	0.07736	-0.02679	0.008505
UKB_460K.biochemistry_IGF1		-0.1796	0.03362	-5.343	9.17E-08	0.2657	0.01972	1.177	0.05282	-0.0014	0.007015
UKB_460K.biochemistry_Calcium		-0.197	0.03751	-5.25	1.52E-07	0.1302	0.01606	1.085	0.03274	-0.00726	0.006607
UKB_460K.biochemistry_SHBG		0.2296	0.04379	5.243	1.58E-07	0.2304	0.05094	1.175	0.08948	0.02627	0.00912
UKB_460K.biochemistry_Albumin		-0.214	0.04101	-5.217	1.82E-07	0.1327	0.008278	1.156	0.03988	-0.01017	0.007435
UKB_460K.biochemistry_ApolipoproteinA		0.1979	0.04064	4.869	1.12E-06	0.1979	0.02355	1.448	0.201	0.0108	0.007678
PASS_BMI1		0.2154	0.0454	4.745	2.08E-06	0.1433	0.007331	0.7674	0.008256	0.009283	0.005062
UKB_460K.biochemistry_AspartateAminotransferase		-0.1679	0.04051	-4.143	3.42E-05	0.1079	0.008458	1.097	0.02713	-0.01041	0.006512
PASS_Type_2_Diabetes		-0.3294	0.08223	-4.006	6.18E-05	0.0931	0.01036	1.014	0.008269	-0.01929	0.005736
PASS_Coronary_Artery_Disease		-0.2503	0.0703	-3.56	0.000371	0.07319	0.009439	1.025	0.008846	-0.00087	0.005655
UKB_460K.biochemistry_AlkalinePhosphatase		-0.1264	0.03783	-3.341	0.000835	0.2622	0.04667	1.35	0.1345	0.007314	0.007145

(*) if source of LD score is not indicated in column this has been obtained from precomputed LD Scores provided by LDSC (Alkes Lab; see methods). Otherwise, the reference provided is the GWAS summary statistics used to calculate the LD scores using LDSC.

Supplementary Material Chapter 3

Supplementary Table S3.1: Protein coding genes taken forward for screening at loci associated with insulin fold change

Human Gene Name	Ensembl ENSG ID	Ensembl ENSMUS ID	Ensembl Gene Name Mouse	Ensembl Gene Name Human
ALKBH5	ENSG00000091542	ENSMUSG00000042650	*Alkbh5*	*ALKBH5*
ATPAF2	ENSG00000171953	ENSMUSG00000042709	*Atpaf2*	*ATPAF2*
BICC1	ENSG00000122870	ENSMUSG00000014329	*Bicc1*	*BICC1*
CAST	ENSG00000153113	ENSMUSG00000021585	*Cast*	*CAST*
CMPK1	ENSG00000162368	ENSMUSG00000028719	*Cmpk1*	*CMPK1*
DRG2	ENSG00000108591	ENSMUSG00000020537	*Drg2*	*DRG2*
ERAP1	ENSG00000164307	ENSMUSG00000021583	*Erap1*	*ERAP1*
ERI1	ENSG00000104626	ENSMUSG00000031527	*Eri1*	*ERI1*
FAM13C	ENSG00000148541	ENSMUSG00000043259	*Fam13c*	*FAM13C*
FLII	ENSG00000177731	ENSMUSG00000002812	*Flii*	*FLII*
GID4	ENSG00000141034	ENSMUSG00000018415	*Gid4*	*GID4*
IRS1	ENSG00000169047	ENSMUSG00000055980	*Irs1*	*IRS1*
KIAA1731	ENSG00000166004	ENSMUSG00000046111	*Cep295*	*CEP295*
LLGL1	ENSG00000131899	ENSMUSG00000020536	*Llgl1*	*LLGL1*
LNPEP	ENSG00000113441	ENSMUSG00000023845	*Lnpep*	*LNPEP*
LRRC48	ENSG00000171962	ENSMUSG00000056598	*Drc3*	*DRC3*
MFHAS1	ENSG00000147324	ENSMUSG00000070056	*Mfhas1*	*MFHAS1*
PDZK1IP1	ENSG00000162366	ENSMUSG00000028716	*Pdzk1ip1*	*PDZK1IP1*
PPP1R3B	ENSG00000173281	ENSMUSG00000046794	*Ppp1r3b*	*PPP1R3B*
PRPSAP2	ENSG00000141127	ENSMUSG00000020528	*Prpsap2*	*PRPSAP2*
RIOK2	ENSG00000058729	ENSMUSG00000116564	*Riok2*	*RIOK2*
SHMT1	ENSG00000176974	ENSMUSG00000020534	*Shmt1*	*SHMT1*
SLC2A4	ENSG00000181856	ENSMUSG00000018566	*Slc2a4*	*SLC2A4*
SLC36A4	ENSG00000180773	ENSMUSG00000043885	*Slc36a4*	*SLC36A4*
SMCO4	ENSG00000166002	ENSMUSG00000058173	*Smco4*	*SMCO4*
SMCR7	ENSG00000100335	ENSMUSG00000022412	*Mief1*	*MIEF1*
SMCR8	ENSG00000176994	ENSMUSG00000049323	*Smcr8*	*SMCR8*
TAL1	ENSG00000162367	ENSMUSG00000028717	*Tal1*	*TAL1*
TFAM	ENSG00000108064	ENSMUSG00000003923	*Tfam*	*TFAM*
TNKS	ENSG00000173273	ENSMUSG00000031529	*Tnks*	*TNKS*
TOM1L2	ENSG00000175662	ENSMUSG00000000538	*Tom1l2*	*TOM1L2*
TOP3A	ENSG00000177302	ENSMUSG00000002814	*Top3a*	*TOP3A*
TRABD2B	ENSG00000269113	ENSMUSG00000070867	*Trabd2b*	*TRABD2B*
TRIM16L	ENSG00000108448	ENSMUSG00000047821	*Trim16*	*TRIM16L*
TVP23B	ENSG00000171928	ENSMUSG00000014177	*Tvp23b*	*TVP23B*
VPS13C	ENSG00000129003	ENSMUSG00000035284	*Vps13c*	*VPS13C*

Supplementary Table S3.2 - Horizon Discovery Dharmacon siGenome siRNA pools used for knockdown

Pool Catalog Number	Duplex Catalog Number	Gene Symbol	GENE ID	Gene Accession	GI Number
M-058321-01	D-058321-01	*Alkbh5*	268420	NM_172943	169234735
M-058321-01	D-058321-02	*Alkbh5*	268420	NM_172943	169234735
M-058321-01	D-058321-03	*Alkbh5*	268420	NM_172943	169234735
M-058321-01	D-058321-04	*Alkbh5*	268420	NM_172943	169234735
M-051768-00	D-051768-01	*Atpaf2*	246782	NM_145427	21703851
M-051768-00	D-051768-02	*Atpaf2*	246782	NM_145427	21703851
M-051768-00	D-051768-03	*Atpaf2*	246782	NM_145427	21703851
M-051768-00	D-051768-04	*Atpaf2*	246782	NM_145427	21703851
M-045420-01	D-045420-01	*Bicc1*	83675	NM_031397	118130749
M-045420-01	D-045420-02	*Bicc1*	83675	NM_031397	118130749
M-045420-01	D-045420-03	*Bicc1*	83675	NM_031397	118130749
M-045420-01	D-045420-04	*Bicc1*	83675	NM_031397	118130749
M-043048-01	D-043048-01	*Cast*	12380	NM_009817	145966843
M-043048-01	D-043048-02	*Cast*	12380	NM_009817	145966843
M-043048-01	D-043048-03	*Cast*	12380	NM_009817	145966843
M-043048-01	D-043048-04	*Cast*	12380	NM_009817	145966843
M-063389-01	D-063389-01	*Cmpk1*	66588	NM_025647	165377064
M-063389-01	D-063389-03	*Cmpk1*	66588	NM_025647	165377064
M-063389-01	D-063389-04	*Cmpk1*	66588	NM_025647	165377064
M-063389-01	D-063389-17	*Cmpk1*	66588	NM_025647	165377064
M-047866-00	D-047866-01	*Drg2*	13495	NM_021354	10946677
M-047866-00	D-047866-02	*Drg2*	13495	NM_021354	10946677
M-047866-00	D-047866-03	*Drg2*	13495	NM_021354	10946677
M-047866-00	D-047866-04	*Drg2*	13495	NM_021354	10946677
M-059768-01	D-059768-01	*Erap1*	80898	NM_030711	146149238
M-059768-01	D-059768-02	*Erap1*	80898	NM_030711	146149238
M-059768-01	D-059768-03	*Erap1*	80898	NM_030711	146149238
M-059768-01	D-059768-04	*Erap1*	80898	NM_030711	146149238
M-059530-01	D-059530-02	*Eri1*	67276	NM_026067	165905604
M-059530-01	D-059530-03	*Eri1*	67276	NM_026067	165905604
M-059530-01	D-059530-04	*Eri1*	67276	NM_026067	165905604

M-059530-01	D-059530-17	*Eri1*	67276	NM_026067	165905604
M-057268-00	D-057268-01	*Fam13c*	71721	NM_024244	31343632
M-057268-00	D-057268-02	*Fam13c*	71721	NM_024244	31343632
M-057268-00	D-057268-03	*Fam13c*	71721	NM_024244	31343632
M-057268-00	D-057268-04	*Fam13c*	71721	NM_024244	31343632
M-050081-00	D-050081-01	*Flii*	14248	NM_022009	11528489
M-050081-00	D-050081-02	*Flii*	14248	NM_022009	11528489
M-050081-00	D-050081-03	*Flii*	14248	NM_022009	11528489
M-050081-00	D-050081-04	*Flii*	14248	NM_022009	11528489
M-042620-00	D-042620-01	*Gid4*	66771	NM_025757	34328468
M-042620-00	D-042620-02	*Gid4*	66771	NM_025757	34328468
M-042620-00	D-042620-03	*Gid4*	66771	NM_025757	34328468
M-042620-00	D-042620-04	*Gid4*	66771	NM_025757	34328468
M-040503-01	D-040503-01	*Irs1*	16367	NM_010570	161086993
M-040503-01	D-040503-03	*Irs1*	16367	NM_010570	161086993
M-040503-01	D-040503-04	*Irs1*	16367	NM_010570	161086993
M-040503-01	D-040503-17	*Irs1*	16367	NM_010570	161086993
M-052952-01	D-052952-02	*Cep295*	319675	NM_176976	161377457
M-052952-01	D-052952-03	*Cep295*	319675	NM_176976	161377457
M-052952-01	D-052952-17	*Cep295*	319675	NM_176976	161377457
M-052952-01	D-052952-18	*Cep295*	319675	NM_176976	161377457
M-047129-00	D-047129-01	*Llgl1*	16897	NM_008502	6678695
M-047129-00	D-047129-02	*Llgl1*	16897	NM_008502	6678695
M-047129-00	D-047129-03	*Llgl1*	16897	NM_008502	6678695
M-047129-00	D-047129-04	*Llgl1*	16897	NM_008502	6678695
M-065506-01	D-065506-01	*Lnpep*	240028	NM_172827	146198621
M-065506-01	D-065506-02	*Lnpep*	240028	NM_172827	146198621
M-065506-01	D-065506-03	*Lnpep*	240028	NM_172827	146198621
M-065506-01	D-065506-04	*Lnpep*	240028	NM_172827	146198621
M-047941-01	D-047941-01	*Lrrc48*	74665	NM_029044	142379069
M-047941-01	D-047941-02	*Lrrc48*	74665	NM_029044	142379069
M-047941-01	D-047941-03	*Lrrc48*	74665	NM_029044	142379069
M-047941-01	D-047941-04	*Lrrc48*	74665	NM_029044	142379069
M-061049-01	D-061049-01	*Mfhas1*	52065	NM_001081279	124486880
M-061049-01	D-061049-02	*Mfhas1*	52065	NM_001081279	124486880
M-061049-01	D-061049-03	*Mfhas1*	52065	NM_001081279	124486880
M-061049-01	D-061049-04	*Mfhas1*	52065	NM_001081279	124486880

M-055151-01	D-055151-01	*Pdzk1ip1*	67182	NM_026018	141802816
M-055151-01	D-055151-02	*Pdzk1ip1*	67182	NM_026018	141802816
M-055151-01	D-055151-03	*Pdzk1ip1*	67182	NM_026018	141802816
M-055151-01	D-055151-04	*Pdzk1ip1*	67182	NM_026018	141802816
M-057396-01	D-057396-01	*Ppp1r3b*	244416	NM_177741	51036623
M-057396-01	D-057396-02	*Ppp1r3b*	244416	NM_177741	51036623
M-057396-01	D-057396-03	*Ppp1r3b*	244416	NM_177741	51036623
M-057396-01	D-057396-04	*Ppp1r3b*	244416	NM_177741	51036623
M-053452-00	D-053452-01	*Prpsap2*	212627	NM_144806	21450168
M-053452-00	D-053452-02	*Prpsap2*	212627	NM_144806	21450168
M-053452-00	D-053452-03	*Prpsap2*	212627	NM_144806	21450168
M-053452-00	D-053452-04	*Prpsap2*	212627	NM_144806	21450168
M-056209-00	D-056209-01	*Riok2*	67045	NM_025934	13385427
M-056209-00	D-056209-02	*Riok2*	67045	NM_025934	13385427
M-056209-00	D-056209-03	*Riok2*	67045	NM_025934	13385427
M-056209-00	D-056209-04	*Riok2*	67045	NM_025934	13385427
M-044898-01	D-044898-01	*Shmt1*	20425	NM_009171	67846102
M-044898-01	D-044898-02	*Shmt1*	20425	NM_009171	67846102
M-044898-01	D-044898-04	*Shmt1*	20425	NM_009171	67846102
M-044898-01	D-044898-17	*Shmt1*	20425	NM_009171	67846102
M-040063-01	D-040063-01	*Slc2a4*	20528	NM_009204	118026924
M-040063-01	D-040063-02	*Slc2a4*	20528	NM_009204	118026924
M-040063-01	D-040063-03	*Slc2a4*	20528	NM_009204	118026924
M-040063-01	D-040063-04	*Slc2a4*	20528	NM_009204	118026924
M-064246-01	D-064246-01	*Slc36a4*	234967	NM_172289	141802265
M-064246-01	D-064246-02	*Slc36a4*	234967	NM_172289	141802265
M-064246-01	D-064246-03	*Slc36a4*	234967	NM_172289	141802265
M-064246-01	D-064246-04	*Slc36a4*	234967	NM_172289	141802265
M-042029-01	D-042029-01	*Smco4*	170748	NM_133214	113865950
M-042029-01	D-042029-04	*Smco4*	170748	NM_133214	113865950
M-042029-01	D-042029-17	*Smco4*	170748	NM_133214	113865950
M-042029-01	D-042029-18	*Smco4*	170748	NM_133214	113865950
M-064405-01	D-064405-01	*Mief1*	239555	NM_178719	142347394
M-064405-01	D-064405-02	*Mief1*	239555	NM_178719	142347394
M-064405-01	D-064405-03	*Mief1*	239555	NM_178719	142347394
M-064405-01	D-064405-04	*Mief1*	239555	NM_178719	142347394
M-057233-01	D-057233-01	*Smcr8*	237782	NM_175491	146198630

M-057233-01	D-057233-02	*Smcr8*	237782	NM_175491	146198630
M-057233-01	D-057233-03	*Smcr8*	237782	NM_175491	146198630
M-057233-01	D-057233-04	*Smcr8*	237782	NM_175491	146198630
M-044207-01	D-044207-01	*Tal1*	21349	NM_011527	133892330
M-044207-01	D-044207-02	*Tal1*	21349	NM_011527	133892330
M-044207-01	D-044207-03	*Tal1*	21349	NM_011527	133892330
M-044207-01	D-044207-04	*Tal1*	21349	NM_011527	133892330
M-058896-01	D-058896-01	*Tfam*	21780	NM_009360	145966787
M-058896-01	D-058896-02	*Tfam*	21780	NM_009360	145966787
M-058896-01	D-058896-03	*Tfam*	21780	NM_009360	145966787
M-058896-01	D-058896-04	*Tfam*	21780	NM_009360	145966787
M-056040-01	D-056040-01	*Tnks*	21951	NM_175091	87239973
M-056040-01	D-056040-02	*Tnks*	21951	NM_175091	87239973
M-056040-01	D-056040-03	*Tnks*	21951	NM_175091	87239973
M-056040-01	D-056040-04	*Tnks*	21951	NM_175091	87239973
M-054478-01	D-054478-01	*Tom1l2*	216810	NM_001039093	84875525
M-054478-01	D-054478-02	*Tom1l2*	216810	NM_001039093	84875525
M-054478-01	D-054478-17	*Tom1l2*	216810	NM_001039093	84875525
M-054478-01	D-054478-18	*Tom1l2*	216810	NM_001039093	84875525
M-047552-01	D-047552-01	*Top3a*	21975	NM_009410	142344076
M-047552-01	D-047552-02	*Top3a*	21975	NM_009410	142344076
M-047552-01	D-047552-03	*Top3a*	21975	NM_009410	142344076
M-047552-01	D-047552-04	*Top3a*	21975	NM_009410	142344076
M-162295-00	D-162295-09	*Trabd2b*	666048	NM_001085549	148234944
M-162295-00	D-162295-10	*Trabd2b*	666048	NM_001085549	148234944
M-162295-00	D-162295-11	*Trabd2b*	666048	NM_001085549	148234944
M-162295-00	D-162295-12	*Trabd2b*	666048	NM_001085549	148234944
M-045755-01	D-045755-01	*Trim16*	94092	NM_053169	70778827
M-045755-01	D-045755-02	*Trim16*	94092	NM_053169	70778827
M-045755-01	D-045755-03	*Trim16*	94092	NM_053169	70778827
M-045755-01	D-045755-04	*Trim16*	94092	NM_053169	70778827
M-063504-02	D-063504-01	*Tvp23b*	67510	NM_026210	146134903
M-063504-02	D-063504-02	*Tvp23b*	67510	NM_026210	146134903
M-063504-02	D-063504-03	*Tvp23b*	67510	NM_026210	146134903
M-063504-02	D-063504-04	*Tvp23b*	67510	NM_026210	146134903
M-053177-01	D-053177-01	*Vps13c*	320528	NM_177184	122114536
M-053177-01	D-053177-02	*Vps13c*	320528	NM_177184	122114536

M-053177-01	D-053177-03	*Vps13c*	320528	NM_177184	122114536
M-053177-01	D-053177-04	*Vps13c*	320528	NM_177184	122114536
M-040174-01	D-040174-01	*Tbc1d4*	210789	NM_001081278	163644269
M-040174-01	D-040174-02	*Tbc1d4*	210789	NM_001081278	163644269
M-040174-01	D-040174-03	*Tbc1d4*	210789	NM_001081278	163644269
M-040174-01	D-040174-04	*Tbc1d4*	210789	NM_001081278	163644269
M-040709-01	D-040709-01	*Akt1*	11651	NM_009652	133891722
M-040709-01	D-040709-03	*Akt1*	11651	NM_009652	133891722
M-040709-01	D-040709-04	*Akt1*	11651	NM_009652	133891722
M-040709-01	D-040709-18	*Akt1*	11651	NM_009652	133891722
M-040782-01	D-040782-01	*Akt2*	11652	NM_007434	158937313
M-040782-01	D-040782-02	*Akt2*	11652	NM_007434	158937313
M-040782-01	D-040782-03	*Akt2*	11652	NM_007434	158937313
M-040782-01	D-040782-04	*Akt2*	11652	NM_007434	158937313
M-040862-01	D-040862-01	*Rab10*	19325	NM_016676	142352796
M-040862-01	D-040862-02	*Rab10*	19325	NM_016676	142352796
M-040862-01	D-040862-03	*Rab10*	19325	NM_016676	142352796
M-040862-01	D-040862-04	*Rab10*	19325	NM_016676	142352796
D-001206-13	D-001210-01	siGENOME Non-targeting Control	0		0
D-001206-13	D-001210-02	siGENOME Non-targeting Control	0		0
D-001206-13	D-001210-03	siGENOME Non-targeting Control	0		0
D-001206-13	D-001210-04	siGENOME Non-targeting Control	0		0
D-001810-10	D-001810-01	ON-TARGETplus Non-targeting Control	0		0
D-001810-10	D-001810-02	ON-TARGETplus Non-targeting Control	0		0
D-001810-10	D-001810-03	ON-TARGETplus Non-targeting Control	0		0
D-001810-10	D-001810-04	ON-TARGETplus Non-targeting Control	0		0
L-040862-01-0005	J-040862-09	ON-TARGETplus Rab10	19325	NM_016676	142352796
L-040862-01-0005	J-040862-10	ON-TARGETplus Rab10	19325	NM_016676	142352796
L-040862-01-0005	J-040862-11	ON-TARGETplus Rab10	19325	NM_016676	142352796
L-040862-01-0005	J-040862-12	ON-TARGETplus Rab10	19325	NM_016676	142352796

Supplementary Table S3.3 - Effect of knockdown of targets of interest on Glucose Transport in wildtype 3T3-L1 adipocytes upon stimulation with 0.5nM insulin.

Data in this table was generated by Dr Daniel Fazakerley.

Insulin Dose (nM)		0	0.5	0	0.5		
siRNA KD	**N**	**Glucose transport (pmole/mg/min)**	**Glucose transport (pmole/mg/min)**	**t-test P value**	**t-test P value**	**NT control compare**	**Control?**
NT control A	3	49.75893	160.3982	NA	NA	A	NT control
Rab10	3	38.81059	78.35885	0.48072	0.00459	A	Positive Control
Llgl1	3	45.33971	115.084	0.574155	0.03006	A	
Flii	3	37.97219	119.4091	0.466168	0.037968	A	
Drg2	3	37.32225	90.21184	0.046223	0.022039	A	
Pdzk1ip1	3	40.23898	215.2586	0.341216	0.016515	A	
Slc36a4	3	42.67205	82.04302	0.6312	0.020431	A	
Vps13c	3	27.4984	101.2159	0.107012	0.031156	A	
Alkbh5	3	31.45314	55.79248	0.122756	0.001419	A	
Akt1/2	3	45.67779	68.66265	0.829253	0.005679	A	Positive Control
Cmpk1	3	52.21355	90.18453	0.829739	0.00735	A	
Tal1	3	42.59834	136.6786	0.001599	0.027223	A	
Tnks	3	49.15795	208.2936	0.960156	0.010722	A	
Tvp23b	3	43.31148	77.84615	0.858559	0.042204	A	
Tom1l2	3	16.94871	40.76831	0.024932	0.00516	A	
Bicc1	3	21.35208	81.15639	0.041854	0.01505	A	
Slc2a4	3	33.49529	81.85042	0.09036	0.005227	A	Positive Control
NT control b	3	35.83996	112.5332	NA	NA	B	NT control
Rab10	3	36.68385	73.02634	0.858988	0.014622	B	Positive Control
Tvp23b	3	21.94021	81.57117	0.021737	0.157995	B	no effect control
Prpsap2	3	53.34755	138.2689	0.156767	0.307902	B	no effect control
Atpaf2	3	63.70846	115.7141	0.025163	0.696101	B	no effect control
Trim16	3	40.24848	101.4391	0.17274	0.101588	B	no effect control
Cep295	3	32.95481	116.2322	0.475514	0.66931	B	no effect control
Fam13c	3	42.12253	143.9908	0.467873	0.309035	B	no effect control
Shmt1	3	34.42249	117.8719	0.699092	0.618672	B	no effect control
Smco4	3	55.63223	105.9474	0.278616	0.76559	B	no effect control
Akt1/2	3	35.51852	65.73061	0.945447	0.010049	B	Positive Control
Top3a	3	28.72903	120.1383	0.253692	0.672205	B	no effect control
Ppp1r3b	3	31.68659	113.8883	0.361259	0.939779	B	no effect control

Supplementary Table S3.4 - GLUT4 screen results in 3T3-L1 wildtype cells (N = 3)

		Surface/Total GLUT4				Surface GLUT4/ Nuclei#				totalGLUT4/Nuclei #				TfR/Nuclei#			
				t-test vs NT ctrl				t-test vs NT ctrl				t-test vs NT ctrl				t-test vs NT ctrl	
siRNA	Insulin (nM)	Mean	SE	P-value	FDR	Mean	SE	P-value	FDR	Mean	SE	P-value	FDR	Mean	SE	P-value	FDR
Akt1/2	0	0.552	0.190	NA	NA	0.452	0.194	NA	NA	1.018	0.057	NA	NA	0.783	0.144	NA	NA
Alkbh5	0	0.284	0.157	NA	NA	0.203	0.077	NA	NA	0.968	0.018	NA	NA	0.470	0.050	NA	NA
Atpaf2	0	0.524	0.020	NA	NA	0.398	0.049	NA	NA	1.059	0.024	NA	NA	0.577	0.029	NA	NA
Bicc1	0	0.545	0.192	NA	NA	0.568	0.227	NA	NA	0.923	0.021	NA	NA	0.789	0.137	NA	NA
Cast	0	0.210	0.081	NA	NA	0.148	0.071	NA	NA	0.953	0.048	NA	NA	0.389	0.063	NA	NA
Cep295	0	0.316	0.093	NA	NA	0.382	0.042	NA	NA	1.050	0.067	NA	NA	0.667	0.020	NA	NA
Cmpk1	0	0.314	0.052	NA	NA	0.371	0.126	NA	NA	1.039	0.049	NA	NA	0.654	0.155	NA	NA
Drc3	0	0.485	0.254	NA	NA	0.829	0.390	NA	NA	1.191	0.090	NA	NA	1.016	0.284	NA	NA
Drg2	0	0.258	0.080	NA	NA	0.392	0.111	NA	NA	1.065	0.011	NA	NA	0.790	0.111	NA	NA
Erap1	0	0.371	0.192	NA	NA	0.586	0.234	NA	NA	1.010	0.001	NA	NA	0.750	0.144	NA	NA
Eri1	0	0.315	0.127	NA	NA	0.367	0.188	NA	NA	0.991	0.050	NA	NA	0.484	0.105	NA	NA
Fam13c	0	0.370	0.055	NA	NA	0.220	0.301	NA	NA	0.966	0.032	NA	NA	0.532	0.123	NA	NA
Flii	0	0.779	0.245	NA	NA	1.289	0.300	NA	NA	1.102	0.018	NA	NA	1.447	0.199	NA	NA
Gid4	0	0.106	0.048	NA	NA	0.020	0.159	NA	NA	1.019	0.095	NA	NA	0.338	0.107	NA	NA
Irs1	0	0.276	0.110	NA	NA	0.494	0.094	NA	NA	1.001	0.076	NA	NA	0.735	0.057	NA	NA
Llgl1	0	0.598	0.292	NA	NA	0.749	0.295	NA	NA	1.036	0.035	NA	NA	1.019	0.127	NA	NA
Lnpep	0	0.489	0.123	NA	NA	0.862	0.370	NA	NA	1.149	0.113	NA	NA	0.948	0.241	NA	NA
Mfhas1	0	0.573	0.261	NA	NA	0.944	0.332	NA	NA	1.173	0.157	NA	NA	1.065	0.248	NA	NA
Mief1	0	0.382	0.225	NA	NA	0.561	0.112	NA	NA	1.127	0.011	NA	NA	0.842	0.059	NA	NA
NT ctrl	0	0.302	0.034	NA	NA	0.338	0.075	NA	NA	1.019	0.019	NA	NA	0.664	0.060	NA	NA
OT_Rab10	0	0.217	0.039	NA	NA	-0.194	0.245	NA	NA	0.974	0.022	NA	NA	0.209	0.070	NA	NA
Pdzk1ip1	0	0.332	0.080	NA	NA	0.583	0.309	NA	NA	1.011	0.036	NA	NA	0.754	0.238	NA	NA
Ppp1r3b	0	0.242	0.126	NA	NA	0.370	0.255	NA	NA	1.171	0.133	NA	NA	0.899	0.222	NA	NA
Prpsap2	0	0.456	0.146	NA	NA	0.666	0.222	NA	NA	1.081	0.119	NA	NA	0.825	0.212	NA	NA
Rab10	0	0.284	0.098	NA	NA	0.411	0.149	NA	NA	1.055	0.052	NA	NA	0.562	0.117	NA	NA
Riok2	0	0.504	0.220	NA	NA	0.710	0.267	NA	NA	1.126	0.039	NA	NA	0.896	0.234	NA	NA
Shmt1	0	0.363	0.084	NA	NA	0.448	0.160	NA	NA	0.936	0.023	NA	NA	0.861	0.145	NA	NA
Slc2a4	0	1.053	0.154	NA	NA	0.232	0.127	NA	NA	0.730	0.056	NA	NA	1.024	0.112	NA	NA
Slc36a4	0	0.313	0.114	NA	NA	0.265	0.280	NA	NA	1.008	0.044	NA	NA	0.977	0.193	NA	NA
Smco4	0	0.404	0.194	NA	NA	0.656	0.087	NA	NA	1.058	0.065	NA	NA	0.824	0.035	NA	NA
Smcr8	0	0.460	0.222	NA	NA	0.584	0.273	NA	NA	0.978	0.016	NA	NA	0.902	0.144	NA	NA

Tal1	0	0.221	0.104	NA	NA	0.196	0.023	NA	NA	1.033	0.023	NA	NA	0.500	0.058	NA	NA
Tbc1d4	0	0.454	0.102	NA	NA	0.425	0.106	NA	NA	1.011	0.020	NA	NA	0.909	0.092	NA	NA
Tfam	0	0.351	0.085	NA	NA	0.338	0.110	NA	NA	1.073	0.012	NA	NA	0.585	0.115	NA	NA
Tnks	0	0.262	0.073	NA	NA	0.194	0.144	NA	NA	1.011	0.051	NA	NA	0.566	0.038	NA	NA
Tom1l2	0	0.520	0.134	NA	NA	0.763	0.147	NA	NA	1.055	0.071	NA	NA	0.845	0.099	NA	NA
Top3a	0	0.296	0.093	NA	NA	0.474	0.064	NA	NA	1.089	0.034	NA	NA	0.790	0.037	NA	NA
Trabd2b	0	0.242	0.143	NA	NA	0.005	0.101	NA	NA	0.962	0.019	NA	NA	0.355	0.062	NA	NA
Trim16	0	0.386	0.179	NA	NA	0.632	0.201	NA	NA	1.054	0.039	NA	NA	0.903	0.108	NA	NA
Tvp23b	0	0.125	0.020	NA	NA	-0.041	0.165	NA	NA	0.971	0.055	NA	NA	0.309	0.178	NA	NA
Vps13c	0	0.210	0.126	NA	NA	0.081	0.182	NA	NA	0.860	0.069	NA	NA	0.312	0.179	NA	NA
Akt1/2	0.5	0.717	0.129	1.36E-05	1.54E-04	0.629	0.119	3.66E-03	1.68E-02	0.974	0.039	9.75E-01	9.94E-01	0.986	0.077	9.94E-01	9.94E-01
Alkbh5	0.5	0.608	0.070	1.67E-05	1.54E-04	0.473	0.082	1.98E-03	1.14E-02	0.952	0.034	6.68E-01	8.59E-01	0.753	0.053	6.68E-02	1.23E-01
Atpaf2	0.5	1.182	0.152	1.49E-01	2.53E-01	1.005	0.158	9.07E-01	9.07E-01	1.087	0.082	2.03E-02	8.49E-02	0.872	0.147	3.66E-01	4.68E-01
Bicc1	0.5	1.010	0.149	7.64E-01	7.64E-01	0.957	0.300	8.61E-01	8.85E-01	0.922	0.030	2.96E-01	5.92E-01	1.148	0.134	2.06E-01	3.06E-01
Cast	0.5	0.829	0.018	3.35E-02	7.01E-02	0.791	0.091	2.38E-01	3.53E-01	0.953	0.046	6.72E-01	8.59E-01	0.686	0.063	1.85E-02	4.74E-02
Cep295	0.5	1.122	0.148	4.05E-01	4.77E-01	1.051	0.170	6.91E-01	8.35E-01	0.996	0.081	6.33E-01	8.59E-01	1.090	0.094	4.18E-01	5.11E-01
Cmpk1	0.5	0.572	0.038	3.31E-06	5.08E-05	0.611	0.150	2.35E-02	5.68E-02	1.014	0.055	4.05E-01	6.97E-01	0.840	0.182	2.50E-01	3.39E-01
Drc3	0.5	1.357	0.340	1.41E-03	4.64E-03	1.440	0.397	6.12E-03	2.56E-02	1.116	0.094	3.58E-03	2.74E-02	1.191	0.201	1.10E-01	1.88E-01
Drg2	0.5	0.640	0.014	6.37E-05	4.19E-04	0.619	0.139	2.66E-02	6.11E-02	0.988	0.009	7.66E-01	9.30E-01	0.960	0.161	8.35E-01	8.53E-01
Erap1	0.5	1.004	0.082	7.21E-01	7.37E-01	1.213	0.046	1.67E-01	2.75E-01	0.994	0.039	6.69E-01	8.59E-01	1.089	0.044	4.22E-01	5.11E-01
Eri1	0.5	0.876	0.037	9.84E-02	1.81E-01	0.810	0.172	2.86E-01	4.10E-01	0.943	0.045	5.30E-01	7.86E-01	0.647	0.146	7.97E-03	3.33E-02
Fam13c	0.5	1.106	0.073	5.04E-01	5.65E-01	0.903	0.170	6.18E-01	7.89E-01	0.934	0.067	4.24E-01	6.97E-01	0.820	0.077	1.91E-01	2.93E-01
Flii	0.5	1.264	0.083	2.35E-02	5.68E-02	1.691	0.134	2.41E-05	3.70E-04	1.055	0.050	9.47E-02	2.56E-01	1.696	0.053	5.54E-08	2.55E-06
Gid4	0.5	0.821	0.125	2.76E-02	6.26E-02	0.652	0.072	4.36E-02	9.56E-02	0.960	0.082	7.88E-01	9.30E-01	0.693	0.157	2.17E-02	5.21E-02
Irs1	0.5	0.487	0.046	4.56E-08	2.10E-06	0.588	0.102	1.63E-02	4.41E-02	0.927	0.073	3.45E-01	6.61E-01	0.832	0.101	2.24E-01	3.21E-01
Llgl1	0.5	1.273	0.240	1.84E-02	4.71E-02	1.277	0.237	7.76E-02	1.37E-01	1.015	0.036	3.96E-01	6.97E-01	1.533	0.130	2.37E-05	5.44E-04
Lnpep	0.5	1.399	0.081	3.09E-04	1.58E-03	1.838	0.487	3.99E-07	1.84E-05	1.118	0.121	3.11E-03	2.74E-02	1.407	0.218	1.06E-03	9.77E-03
Mfhas1	0.5	1.171	0.222	1.83E-01	2.81E-01	1.390	0.404	1.44E-02	4.41E-02	1.058	0.059	8.41E-02	2.56E-01	1.237	0.275	5.01E-02	9.60E-02
Mief1	0.5	0.905	0.178	1.73E-01	2.80E-01	1.312	0.322	4.78E-02	9.56E-02	1.120	0.051	2.85E-03	2.74E-02	1.353	0.113	4.24E-03	1.95E-02
NT ctrl	0.5	1.040	0.024	NA	NA	0.986	0.057	NA	NA	0.973	0.017	NA	NA	0.987	0.052	NA	NA
OT_Rab10	0.5	0.635	0.036	7.53E-04	3.07E-03	0.268	0.221	3.42E-04	3.01E-03	0.957	0.008	7.88E-01	9.30E-01	0.544	0.070	4.15E-03	1.95E-02
Pdzk1ip1	0.5	1.445	0.177	5.12E-05	3.93E-04	1.649	0.160	6.98E-05	8.03E-04	0.983	0.044	8.42E-01	9.44E-01	1.278	0.190	2.26E-02	5.21E-02
Ppp1r3b	0.5	0.927	0.040	2.56E-01	3.46E-01	1.041	0.229	7.35E-01	8.35E-01	1.133	0.134	1.12E-03	1.72E-02	1.313	0.209	1.09E-02	3.51E-02
Prpsap2	0.5	0.933	0.115	2.82E-01	3.70E-01	1.307	0.086	5.17E-02	9.91E-02	1.110	0.066	5.38E-03	3.54E-02	1.135	0.055	2.45E-01	3.39E-01
Rab10	0.5	0.752	0.072	9.72E-05	5.59E-04	0.888	0.099	4.24E-01	5.74E-01	1.034	0.048	9.33E-02	2.56E-01	0.841	0.049	1.23E-01	1.94E-01
Riok2	0.5	1.221	0.167	6.63E-02	1.27E-01	1.411	0.275	1.01E-02	3.32E-02	1.088	0.059	1.89E-02	8.49E-02	1.217	0.252	7.12E-02	1.26E-01
Shmt1	0.5	0.980	0.061	5.44E-01	5.86E-01	0.932	0.125	7.44E-01	8.35E-01	0.886	0.061	7.42E-02	2.44E-01	1.316	0.131	1.02E-02	3.51E-02
Slc2a4	0.5	1.133	0.079	3.45E-01	4.17E-01	0.253	0.073	1.17E-05	2.70E-04	0.679	0.066	4.20E-09	1.93E-07	1.436	0.180	4.77E-04	7.32E-03
Slc36a4	0.5	0.742	0.038	2.79E-03	8.01E-03	0.555	0.225	9.33E-03	3.30E-02	0.942	0.035	5.17E-01	7.86E-01	1.313	0.214	1.09E-02	3.51E-02

Smco4	0.5	0.919	0.124	2.22E-01	3.19E-01	0.952	0.139	8.38E-01	8.85E-01	0.970	0.079	9.41E-01	9.94E-01	1.056	0.091	5.89E-01	6.45E-01
Smcr8	0.5	1.173	0.094	1.76E-01	2.80E-01	1.158	0.137	2.94E-01	4.10E-01	0.963	0.047	8.30E-01	9.44E-01	1.354	0.038	4.18E-03	1.95E-02
Tal1	0.5	0.832	0.070	3.61E-02	7.23E-02	0.790	0.122	2.35E-01	3.53E-01	1.012	0.048	4.21E-01	6.97E-01	0.846	0.078	2.70E-01	3.55E-01
Tbc1d4	0.5	1.065	0.073	7.09E-01	7.37E-01	0.944	0.119	7.19E-01	8.35E-01	0.973	0.032	9.94E-01	9.94E-01	1.198	0.107	1.74E-02	4.71E-02
Tfam	0.5	1.152	0.030	2.55E-01	3.46E-01	1.313	0.238	4.72E-02	9.56E-02	1.093	0.075	1.43E-02	7.32E-02	1.080	0.127	4.65E-01	5.41E-01
Tnks	0.5	1.553	0.149	3.56E-07	8.18E-06	1.427	0.094	7.65E-03	2.93E-02	0.940	0.060	4.90E-01	7.77E-01	1.079	0.152	4.71E-01	5.41E-01
Tom1l2	0.5	0.788	0.059	1.11E-02	3.00E-02	0.952	0.156	8.39E-01	8.85E-01	1.025	0.073	2.87E-01	5.92E-01	1.016	0.156	8.20E-01	8.53E-01
Top3a	0.5	1.108	0.113	4.89E-01	5.62E-01	1.239	0.111	1.25E-01	2.13E-01	1.076	0.031	3.48E-02	1.34E-01	1.185	0.099	1.21E-01	1.94E-01
Trabd2b	0.5	0.944	0.207	3.35E-01	4.16E-01	0.680	0.166	6.40E-02	1.18E-01	0.896	0.048	1.14E-01	2.91E-01	0.668	0.138	1.26E-02	3.61E-02
Trim16	0.5	0.937	0.156	3.00E-01	3.84E-01	1.112	0.083	4.42E-01	5.80E-01	1.000	0.043	5.84E-01	8.39E-01	1.262	0.060	3.11E-02	6.81E-02
Tvp23b	0.5	0.715	0.075	1.12E-03	3.97E-03	0.396	0.177	3.93E-04	3.01E-03	0.903	0.041	1.52E-01	3.69E-01	0.590	0.204	1.98E-03	1.30E-02
Vps13c	0.5	0.705	0.079	7.77E-04	3.07E-03	0.425	0.237	7.46E-04	4.90E-03	0.808	0.086	7.81E-04	1.72E-02	0.558	0.243	8.38E-04	9.64E-03
Akt1/2	100	1.188	0.112	NA	NA	1.090	0.155	NA	NA	1.045	0.071	NA	NA	1.314	0.142	NA	NA
Alkbh5	100	1.281	0.179	NA	NA	1.279	0.224	NA	NA	0.983	0.056	NA	NA	1.071	0.153	NA	NA
Atpaf2	100	1.696	0.160	NA	NA	1.569	0.183	NA	NA	1.032	0.044	NA	NA	1.001	0.149	NA	NA
Bicc1	100	1.825	0.268	NA	NA	1.649	0.326	NA	NA	0.899	0.042	NA	NA	1.309	0.216	NA	NA
Cast	100	1.560	0.268	NA	NA	1.429	0.201	NA	NA	0.939	0.032	NA	NA	0.773	0.060	NA	NA
Cep295	100	1.987	0.262	NA	NA	1.783	0.262	NA	NA	0.973	0.077	NA	NA	1.250	0.124	NA	NA
Cmpk1	100	1.063	0.085	NA	NA	1.104	0.069	NA	NA	0.976	0.036	NA	NA	1.024	0.136	NA	NA
Drc3	100	2.232	0.404	NA	NA	2.115	0.389	NA	NA	1.061	0.111	NA	NA	1.255	0.242	NA	NA
Drg2	100	1.445	0.241	NA	NA	1.510	0.237	NA	NA	1.008	0.049	NA	NA	1.346	0.269	NA	NA
Erap1	100	1.692	0.259	NA	NA	2.040	0.104	NA	NA	1.042	0.040	NA	NA	1.391	0.159	NA	NA
Eri1	100	1.729	0.081	NA	NA	1.663	0.298	NA	NA	0.948	0.069	NA	NA	0.818	0.136	NA	NA
Fam13c	100	1.616	0.099	NA	NA	1.337	0.195	NA	NA	0.894	0.076	NA	NA	0.865	0.079	NA	NA
Flii	100	2.000	0.197	NA	NA	2.408	0.297	NA	NA	1.048	0.028	NA	NA	2.112	0.208	NA	NA
Gid4	100	1.467	0.202	NA	NA	1.238	0.299	NA	NA	0.961	0.087	NA	NA	0.783	0.199	NA	NA
Irs1	100	1.446	0.187	NA	NA	1.553	0.178	NA	NA	0.948	0.066	NA	NA	1.158	0.073	NA	NA
Llgl1	100	1.666	0.108	NA	NA	1.533	0.040	NA	NA	1.001	0.036	NA	NA	1.538	0.042	NA	NA
Lnpep	100	2.171	0.219	NA	NA	2.299	0.256	NA	NA	1.055	0.106	NA	NA	1.430	0.123	NA	NA
Mfhas1	100	2.191	0.079	NA	NA	2.296	0.262	NA	NA	1.064	0.102	NA	NA	1.469	0.263	NA	NA
Mief1	100	1.917	0.100	NA	NA	2.155	0.071	NA	NA	1.083	0.041	NA	NA	1.427	0.031	NA	NA
NT ctrl	100	1.724	0.065	NA	NA	1.644	0.069	NA	NA	0.969	0.020	NA	NA	1.165	0.061	NA	NA
OT_Rab10	100	1.210	0.021	NA	NA	0.803	0.318	NA	NA	0.934	0.067	NA	NA	0.753	0.178	NA	NA
Pdzk1ip1	100	2.263	0.561	NA	NA	2.210	0.263	NA	NA	0.927	0.063	NA	NA	1.262	0.048	NA	NA
Ppp1r3b	100	1.561	0.279	NA	NA	1.805	0.106	NA	NA	1.079	0.103	NA	NA	1.527	0.134	NA	NA
Prpsap2	100	1.605	0.030	NA	NA	1.997	0.022	NA	NA	1.086	0.026	NA	NA	1.352	0.070	NA	NA
Rab10	100	1.371	0.123	NA	NA	1.479	0.139	NA	NA	1.020	0.046	NA	NA	1.013	0.069	NA	NA
Riok2	100	2.029	0.143	NA	NA	2.238	0.134	NA	NA	1.062	0.038	NA	NA	1.375	0.141	NA	NA
Shmt1	100	1.561	0.137	NA	NA	1.606	0.289	NA	NA	0.930	0.035	NA	NA	1.681	0.344	NA	NA

Slc2a4	100	1.420	0.087	NA	NA	0.463	0.089	NA	NA	0.678	0.051	NA	NA	1.692	0.227	NA	NA
Slc36a4	100	1.397	0.069	NA	NA	1.205	0.124	NA	NA	0.937	0.010	NA	NA	1.541	0.136	NA	NA
Smco4	100	1.529	0.025	NA	NA	1.647	0.070	NA	NA	1.000	0.064	NA	NA	1.273	0.027	NA	NA
Smcr8	100	1.910	0.067	NA	NA	1.792	0.206	NA	NA	0.945	0.035	NA	NA	1.610	0.160	NA	NA
Tal1	100	1.765	0.136	NA	NA	1.688	0.219	NA	NA	1.020	0.087	NA	NA	1.192	0.132	NA	NA
Tbc1d4	100	1.648	0.116	NA	NA	1.437	0.124	NA	NA	0.962	0.031	NA	NA	1.342	0.128	NA	NA
Tfam	100	1.823	0.160	NA	NA	1.999	0.136	NA	NA	1.079	0.047	NA	NA	1.223	0.120	NA	NA
Tnks	100	2.227	0.350	NA	NA	2.254	0.203	NA	NA	0.955	0.051	NA	NA	1.311	0.105	NA	NA
Tom1l2	100	1.412	0.101	NA	NA	1.500	0.100	NA	NA	0.992	0.027	NA	NA	1.225	0.074	NA	NA
Top3a	100	1.875	0.123	NA	NA	2.011	0.100	NA	NA	1.029	0.048	NA	NA	1.520	0.066	NA	NA
Trabd2b	100	1.538	0.049	NA	NA	1.182	0.178	NA	NA	0.907	0.061	NA	NA	0.758	0.183	NA	NA
Trim16	100	1.671	0.081	NA	NA	1.787	0.242	NA	NA	0.963	0.013	NA	NA	1.485	0.147	NA	NA
Tvp23b	100	1.224	0.172	NA	NA	0.999	0.267	NA	NA	0.930	0.053	NA	NA	0.801	0.251	NA	NA
Vps13c	100	1.312	0.070	NA	NA	0.813	0.156	NA	NA	0.784	0.067	NA	NA	0.667	0.185	NA	NA

*OT_Rab10 – Rab10 KD using ON-TARGETplus siRNA pools

Supplementary Table S3.5 - GLUT4 screen results in 3T3-L1 cells stably expressing HA- GLUT4-mRuby (N = 5)

		Surface/Total GLUT4				Surface GLUT4/ Nuclei#				totalGLUT4/Nuclei #				TfR/Nuclei#			
				t-test vs NT ctrl				t-test vs NT ctrl				t-test vs NT ctrl				t-test vs NT ctrl	
siRNA	Insulin (nM)	Mean	SE	P-value	FDR	Mean	SE	P-value	FDR	Mean	SE	P-value	FDR	Mean	SE	P-value	FDR
Akt1/2	0	0.300	0.029	NA	NA	0.325	0.055	NA	NA	1.015	0.023	NA	NA	0.581	0.035	NA	NA
Alkbh5	0	0.196	0.018	NA	NA	0.224	0.037	NA	NA	1.043	0.028	NA	NA	0.492	0.023	NA	NA
Atpaf2	0	0.228	0.018	NA	NA	0.371	0.040	NA	NA	1.110	0.023	NA	NA	0.527	0.026	NA	NA
Bicc1	0	0.452	0.021	NA	NA	0.260	0.053	NA	NA	0.926	0.031	NA	NA	0.594	0.031	NA	NA
Cast	0	0.288	0.031	NA	NA	0.379	0.102	NA	NA	1.067	0.032	NA	NA	0.519	0.055	NA	NA
Cep295	0	0.313	0.023	NA	NA	0.288	0.064	NA	NA	1.012	0.031	NA	NA	0.574	0.050	NA	NA
Cmpk1	0	0.257	0.008	NA	NA	0.368	0.037	NA	NA	1.094	0.022	NA	NA	0.550	0.038	NA	NA
Drc3	0	0.318	0.030	NA	NA	0.401	0.054	NA	NA	1.086	0.018	NA	NA	0.568	0.014	NA	NA
Drg2	0	0.287	0.020	NA	NA	0.295	0.051	NA	NA	1.047	0.031	NA	NA	0.808	0.043	NA	NA
Erap1	0	0.221	0.014	NA	NA	0.279	0.032	NA	NA	1.030	0.027	NA	NA	0.570	0.038	NA	NA
Eri1	0	0.496	0.038	NA	NA	0.445	0.107	NA	NA	0.988	0.036	NA	NA	0.509	0.055	NA	NA
Fam13c	0	0.332	0.026	NA	NA	0.424	0.070	NA	NA	1.082	0.023	NA	NA	0.544	0.035	NA	NA
Flii	0	0.215	0.050	NA	NA	0.993	0.161	NA	NA	1.301	0.052	NA	NA	1.047	0.083	NA	NA
Gid4	0	0.269	0.021	NA	NA	0.296	0.063	NA	NA	1.055	0.024	NA	NA	0.442	0.037	NA	NA
Irs1	0	0.180	0.031	NA	NA	0.241	0.046	NA	NA	1.034	0.024	NA	NA	0.584	0.031	NA	NA
Llgl1	0	0.407	0.035	NA	NA	0.343	0.073	NA	NA	0.964	0.035	NA	NA	0.729	0.069	NA	NA
Lnpep	0	0.618	0.040	NA	NA	0.392	0.115	NA	NA	0.915	0.041	NA	NA	0.643	0.086	NA	NA
Mfhas1	0	0.374	0.030	NA	NA	0.386	0.060	NA	NA	1.000	0.033	NA	NA	0.620	0.046	NA	NA
Mief1	0	0.301	0.021	NA	NA	0.508	0.082	NA	NA	1.134	0.028	NA	NA	0.638	0.057	NA	NA
NT ctrl	0	0.351	0.004	NA	NA	0.293	0.008	NA	NA	1.002	0.005	NA	NA	0.553	0.005	NA	NA
OT_Rab10	0	0.269	0.018	NA	NA	0.257	0.034	NA	NA	1.047	0.020	NA	NA	0.447	0.035	NA	NA
Pdzk1ip1	0	0.354	0.028	NA	NA	0.324	0.056	NA	NA	0.985	0.032	NA	NA	0.604	0.045	NA	NA
Ppp1r3b	0	0.140	0.021	NA	NA	0.292	0.060	NA	NA	1.119	0.029	NA	NA	0.723	0.069	NA	NA
Prpsap2	0	0.482	0.037	NA	NA	0.423	0.050	NA	NA	0.975	0.026	NA	NA	0.607	0.040	NA	NA
Rab10	0	0.346	0.018	NA	NA	0.326	0.021	NA	NA	1.016	0.013	NA	NA	0.451	0.015	NA	NA
Riok2	0	0.267	0.013	NA	NA	0.247	0.031	NA	NA	1.017	0.027	NA	NA	0.582	0.044	NA	NA
Shmt1	0	0.330	0.029	NA	NA	0.246	0.021	NA	NA	0.969	0.022	NA	NA	0.632	0.032	NA	NA
Slc2a4	0	0.729	0.048	NA	NA	0.211	0.035	NA	NA	0.836	0.018	NA	NA	0.933	0.074	NA	NA
Slc36a4	0	0.420	0.021	NA	NA	0.304	0.036	NA	NA	0.990	0.017	NA	NA	0.852	0.072	NA	NA
Smco4	0	-0.032	0.030	NA	NA	0.259	0.056	NA	NA	1.145	0.034	NA	NA	0.509	0.041	NA	NA
Smcr8	0	0.468	0.025	NA	NA	0.302	0.057	NA	NA	0.943	0.022	NA	NA	0.640	0.043	NA	NA

Tal1	0	0.292	0.026	NA	NA	0.118	0.044	NA	NA	0.964	0.022	NA	NA	0.386	0.040	NA	NA
Tbc1d4	0	0.448	0.019	NA	NA	0.529	0.049	NA	NA	1.029	0.019	NA	NA	0.780	0.036	NA	NA
Tfam	0	0.452	0.024	NA	NA	0.428	0.081	NA	NA	1.041	0.031	NA	NA	0.559	0.056	NA	NA
Tnks	0	0.428	0.021	NA	NA	0.332	0.056	NA	NA	1.006	0.015	NA	NA	0.510	0.040	NA	NA
Tom1l2	0	0.296	0.020	NA	NA	0.241	0.051	NA	NA	0.991	0.021	NA	NA	0.467	0.037	NA	NA
Top3a	0	0.364	0.020	NA	NA	0.207	0.051	NA	NA	0.967	0.024	NA	NA	0.529	0.038	NA	NA
Trabd2b	0	0.354	0.028	NA	NA	0.232	0.037	NA	NA	0.997	0.026	NA	NA	0.387	0.019	NA	NA
Trim16	0	0.483	0.046	NA	NA	0.709	0.075	NA	NA	1.079	0.031	NA	NA	0.838	0.042	NA	NA
Tvp23b	0	0.445	0.027	NA	NA	0.454	0.065	NA	NA	1.044	0.024	NA	NA	0.544	0.040	NA	NA
Vps13c	0	0.232	0.048	NA	NA	0.223	0.048	NA	NA	1.010	0.028	NA	NA	0.382	0.027	NA	NA
Akt1/2	0.5	0.605	0.027	1.48E-12	6.81E-11	0.664	0.066	1.40E-02	1.29E-01	1.015	0.025	2.52E-01	5.28E-01	0.862	0.045	4.07E-02	1.67E-01
Alkbh5	0.5	0.738	0.026	7.35E-05	6.76E-04	0.818	0.085	2.77E-01	6.85E-01	1.031	0.033	2.48E-01	5.28E-01	0.912	0.047	2.63E-01	4.32E-01
Atpaf2	0.5	0.958	0.037	7.70E-02	1.83E-01	1.231	0.092	3.76E-01	7.20E-01	1.100	0.022	1.34E-02	8.82E-02	0.936	0.040	3.52E-01	5.17E-01
Bicc1	0.5	1.101	0.044	7.49E-01	8.40E-01	0.897	0.090	4.79E-01	8.48E-01	0.920	0.027	3.43E-01	6.37E-01	1.002	0.042	6.74E-01	7.78E-01
Cast	0.5	1.173	0.038	6.85E-01	8.27E-01	1.348	0.126	1.50E-01	4.91E-01	1.051	0.033	1.22E-01	3.30E-01	1.002	0.065	6.77E-01	7.78E-01
Cep295	0.5	1.171	0.081	7.01E-01	8.27E-01	1.271	0.201	2.83E-01	6.85E-01	1.002	0.037	5.43E-01	8.05E-01	1.246	0.107	1.38E-01	2.54E-01
Cmpk1	0.5	0.732	0.032	5.71E-05	6.56E-04	0.918	0.069	5.42E-01	8.64E-01	1.087	0.022	2.66E-02	1.01E-01	0.881	0.052	1.73E-01	2.94E-01
Drc3	0.5	1.199	0.052	5.01E-01	7.43E-01	1.381	0.086	1.10E-01	4.91E-01	1.062	0.020	7.96E-02	2.44E-01	1.239	0.093	1.52E-01	2.68E-01
Drg2	0.5	0.812	0.022	3.51E-03	1.62E-02	0.858	0.090	4.22E-01	7.76E-01	1.039	0.027	2.35E-01	5.28E-01	1.313	0.051	6.80E-02	1.85E-01
Erap1	0.5	0.944	0.038	5.59E-02	1.83E-01	1.000	0.069	8.27E-01	9.55E-01	0.999	0.024	5.87E-01	8.17E-01	1.112	0.064	6.59E-01	7.78E-01
Eri1	0.5	1.090	0.049	6.61E-01	8.22E-01	0.978	0.129	7.47E-01	9.28E-01	0.953	0.031	7.44E-01	9.25E-01	0.851	0.055	1.11E-01	2.31E-01
Fam13c	0.5	1.184	0.059	6.05E-01	7.73E-01	1.606	0.162	7.67E-03	8.81E-02	1.131	0.028	2.19E-03	3.35E-02	1.190	0.067	2.93E-01	4.65E-01
Flii	0.5	1.076	0.081	5.64E-01	7.63E-01	1.971	0.238	1.20E-05	5.50E-04	1.218	0.048	2.88E-06	1.32E-04	1.630	0.123	8.55E-06	3.93E-04
Gid4	0.5	1.074	0.063	5.52E-01	7.63E-01	1.363	0.166	1.30E-01	4.91E-01	1.086	0.032	2.73E-02	1.01E-01	1.019	0.072	7.73E-01	8.47E-01
Irs1	0.5	0.728	0.064	4.73E-05	6.56E-04	1.065	0.175	9.26E-01	9.88E-01	1.085	0.044	2.85E-02	1.01E-01	1.081	0.078	8.40E-01	8.78E-01
Llgl1	0.5	1.318	0.047	6.09E-02	1.83E-01	1.241	0.127	3.51E-01	7.20E-01	0.943	0.035	6.04E-01	8.17E-01	1.455	0.104	1.88E-03	2.16E-02
Lnpep	0.5	1.460	0.025	9.78E-04	6.42E-03	0.970	0.080	7.20E-01	9.20E-01	0.850	0.027	2.30E-02	1.01E-01	1.175	0.061	3.52E-01	5.17E-01
Mfhas1	0.5	1.045	0.034	3.72E-01	5.90E-01	1.049	0.081	9.88E-01	9.88E-01	0.978	0.029	8.84E-01	9.48E-01	1.066	0.045	9.36E-01	9.36E-01
Mief1	0.5	0.959	0.021	7.88E-02	1.83E-01	1.173	0.068	5.44E-01	8.64E-01	1.089	0.023	2.34E-02	1.01E-01	1.039	0.059	8.97E-01	9.17E-01
NT ctrl	0.5	1.129	0.009	NA	NA	1.041	0.018	NA	NA	0.968	0.006	NA	NA	1.059	0.011	NA	NA
OT_Rab10	0.5	0.810	0.031	1.16E-03	6.66E-03	0.858	0.027	3.70E-01	7.20E-01	1.043	0.018	1.67E-01	4.28E-01	0.840	0.034	9.19E-02	2.11E-01
Pdzk1ip1	0.5	1.305	0.041	8.17E-02	1.83E-01	1.288	0.112	2.47E-01	6.68E-01	0.965	0.034	9.27E-01	9.48E-01	1.294	0.074	6.32E-02	1.85E-01
Ppp1r3b	0.5	0.854	0.031	4.93E-03	2.06E-02	1.001	0.056	8.30E-01	9.55E-01	1.067	0.018	6.55E-02	2.15E-01	1.393	0.090	8.67E-03	5.70E-02
Prpsap2	0.5	1.131	0.058	9.88E-01	9.88E-01	1.007	0.090	8.55E-01	9.60E-01	0.942	0.025	5.96E-01	8.17E-01	1.022	0.054	7.91E-01	8.47E-01
Rab10	0.5	0.951	0.031	1.28E-02	4.89E-02	0.987	0.045	7.05E-01	9.20E-01	1.011	0.015	2.92E-01	5.84E-01	0.885	0.030	7.04E-02	1.85E-01
Riok2	0.5	1.121	0.039	9.07E-01	9.48E-01	1.351	0.135	1.45E-01	4.91E-01	1.052	0.037	1.18E-01	3.30E-01	1.299	0.095	5.79E-02	1.85E-01
Shmt1	0.5	1.074	0.033	5.53E-01	7.63E-01	1.040	0.035	9.78E-01	9.88E-01	0.972	0.021	9.69E-01	9.69E-01	1.286	0.051	7.25E-02	1.85E-01

Slc2a4	0.5	1.168	0.059	7.24E-01	8.33E-01	0.609	0.060	3.76E-02	2.88E-01	0.835	0.013	1.05E-02	8.82E-02	1.588	0.112	3.63E-05	8.35E-04
Slc36a4	0.5	0.962	0.027	8.33E-02	1.83E-01	0.852	0.057	3.55E-01	7.20E-01	0.978	0.017	8.79E-01	9.48E-01	1.427	0.110	3.89E-03	3.58E-02
Smco4	0.5	0.699	0.055	1.34E-05	3.08E-04	1.038	0.105	9.73E-01	9.88E-01	1.098	0.029	1.56E-02	8.96E-02	0.938	0.051	3.59E-01	5.17E-01
Smcr8	0.5	1.328	0.025	4.84E-02	1.71E-01	1.121	0.061	7.17E-01	9.20E-01	0.923	0.023	3.65E-01	6.47E-01	1.272	0.056	9.11E-02	2.11E-01
Tal1	0.5	0.977	0.029	1.16E-01	2.22E-01	0.721	0.035	1.22E-01	4.91E-01	0.925	0.021	3.91E-01	6.67E-01	0.797	0.036	4.35E-02	1.67E-01
Tbc1d4	0.5	1.197	0.032	3.13E-01	5.15E-01	1.209	0.069	2.27E-01	6.54E-01	0.977	0.017	8.40E-01	9.48E-01	1.190	0.047	1.05E-01	2.29E-01
Tfam	0.5	1.112	0.043	8.35E-01	9.15E-01	0.969	0.044	7.16E-01	9.20E-01	0.982	0.024	8.20E-01	9.48E-01	0.973	0.054	5.21E-01	6.66E-01
Tnks	0.5	1.430	0.035	2.79E-03	1.42E-02	1.328	0.043	1.78E-01	5.45E-01	0.976	0.019	9.12E-01	9.48E-01	1.158	0.036	4.26E-01	5.86E-01
Tom1l2	0.5	0.994	0.061	1.60E-01	2.95E-01	0.964	0.059	6.96E-01	9.20E-01	0.975	0.023	9.27E-01	9.48E-01	0.957	0.037	4.42E-01	5.86E-01
Top3a	0.5	1.217	0.018	3.94E-01	6.05E-01	0.988	0.026	7.85E-01	9.50E-01	0.935	0.022	5.01E-01	7.94E-01	1.153	0.038	4.46E-01	5.86E-01
Trabd2b	0.5	1.028	0.033	2.87E-01	4.89E-01	0.909	0.037	5.15E-01	8.64E-01	0.984	0.021	7.87E-01	9.48E-01	0.791	0.027	3.91E-02	1.67E-01
Trim16	0.5	1.268	0.078	1.73E-01	3.07E-01	1.398	0.114	9.28E-02	4.91E-01	1.008	0.028	4.67E-01	7.68E-01	1.295	0.065	6.20E-02	1.85E-01
Tvp23b	0.5	1.120	0.066	8.94E-01	9.48E-01	1.165	0.100	5.70E-01	8.74E-01	1.020	0.023	3.46E-01	6.37E-01	0.987	0.065	5.94E-01	7.39E-01
Vps13c	0.5	0.966	0.050	9.11E-02	1.91E-01	0.847	0.051	3.43E-01	7.20E-01	0.952	0.023	7.30E-01	9.25E-01	0.786	0.023	3.57E-02	1.67E-01
Akt1/2	100	1.022	0.031	NA	NA	1.045	0.082	NA	NA	0.990	0.023	NA	NA	1.001	0.042	NA	NA
Alkbh5	100	1.198	0.038	NA	NA	1.262	0.064	NA	NA	1.015	0.028	NA	NA	1.139	0.033	NA	NA
Atpaf2	100	1.441	0.067	NA	NA	1.689	0.135	NA	NA	1.063	0.025	NA	NA	1.075	0.057	NA	NA
Bicc1	100	1.697	0.068	NA	NA	1.633	0.133	NA	NA	0.946	0.030	NA	NA	1.329	0.055	NA	NA
Cast	100	1.838	0.081	NA	NA	2.193	0.178	NA	NA	1.062	0.031	NA	NA	1.337	0.073	NA	NA
Cep295	100	1.748	0.086	NA	NA	1.883	0.231	NA	NA	0.995	0.037	NA	NA	1.514	0.107	NA	NA
Cmpk1	100	1.328	0.036	NA	NA	1.446	0.095	NA	NA	1.033	0.024	NA	NA	1.053	0.055	NA	NA
Drc3	100	1.697	0.052	NA	NA	1.915	0.088	NA	NA	1.055	0.019	NA	NA	1.447	0.087	NA	NA
Drg2	100	1.480	0.050	NA	NA	1.574	0.096	NA	NA	1.028	0.030	NA	NA	1.786	0.072	NA	NA
Erap1	100	1.570	0.057	NA	NA	1.692	0.095	NA	NA	0.998	0.023	NA	NA	1.400	0.078	NA	NA
Eri1	100	1.683	0.076	NA	NA	1.578	0.209	NA	NA	0.939	0.038	NA	NA	1.080	0.074	NA	NA
Fam13c	100	1.516	0.047	NA	NA	1.690	0.089	NA	NA	1.053	0.014	NA	NA	1.195	0.027	NA	NA
Flii	100	1.756	0.127	NA	NA	2.956	0.343	NA	NA	1.225	0.049	NA	NA	2.037	0.136	NA	NA
Gid4	100	1.614	0.058	NA	NA	1.704	0.118	NA	NA	1.018	0.023	NA	NA	1.151	0.065	NA	NA
Irs1	100	1.430	0.059	NA	NA	1.758	0.152	NA	NA	1.054	0.034	NA	NA	1.481	0.070	NA	NA
Llgl1	100	1.762	0.038	NA	NA	1.481	0.089	NA	NA	0.897	0.029	NA	NA	1.570	0.076	NA	NA
Lnpep	100	1.889	0.069	NA	NA	1.410	0.148	NA	NA	0.851	0.029	NA	NA	1.458	0.088	NA	NA
Mfhas1	100	1.609	0.067	NA	NA	1.862	0.166	NA	NA	1.023	0.035	NA	NA	1.472	0.071	NA	NA
Mief1	100	1.659	0.018	NA	NA	1.830	0.066	NA	NA	1.052	0.021	NA	NA	1.289	0.045	NA	NA
NT ctrl	100	1.606	0.011	NA	NA	1.477	0.022	NA	NA	0.950	0.005	NA	NA	1.267	0.011	NA	NA
OT_Rab10	100	1.284	0.026	NA	NA	1.303	0.028	NA	NA	1.020	0.019	NA	NA	1.077	0.046	NA	NA
Pdzk1ip1	100	1.688	0.053	NA	NA	1.598	0.120	NA	NA	0.939	0.030	NA	NA	1.474	0.068	NA	NA
Ppp1r3b	100	1.358	0.038	NA	NA	1.378	0.069	NA	NA	1.009	0.026	NA	NA	1.625	0.096	NA	NA

Prpsap2	100	1.545	0.054	NA	NA	1.341	0.092	NA	NA	0.918	0.023	NA	NA	1.195	0.054	NA	NA
Rab10	100	1.413	0.031	NA	NA	1.353	0.053	NA	NA	0.972	0.015	NA	NA	1.074	0.040	NA	NA
Riok2	100	1.683	0.048	NA	NA	1.827	0.156	NA	NA	1.011	0.034	NA	NA	1.557	0.092	NA	NA
Shmt1	100	1.518	0.068	NA	NA	1.452	0.082	NA	NA	0.952	0.018	NA	NA	1.456	0.054	NA	NA
Slc2a4	100	1.508	0.053	NA	NA	0.778	0.059	NA	NA	0.797	0.014	NA	NA	1.681	0.103	NA	NA
Slc36a4	100	1.419	0.045	NA	NA	1.198	0.070	NA	NA	0.936	0.017	NA	NA	1.662	0.120	NA	NA
Smco4	100	1.440	0.088	NA	NA	1.945	0.203	NA	NA	1.084	0.038	NA	NA	1.259	0.066	NA	NA
Smcr8	100	1.720	0.050	NA	NA	1.313	0.093	NA	NA	0.872	0.022	NA	NA	1.384	0.074	NA	NA
Tal1	100	1.565	0.033	NA	NA	1.199	0.059	NA	NA	0.894	0.020	NA	NA	1.075	0.039	NA	NA
Tbc1d4	100	1.567	0.037	NA	NA	1.558	0.086	NA	NA	0.963	0.018	NA	NA	1.351	0.049	NA	NA
Tfam	100	1.590	0.043	NA	NA	1.359	0.049	NA	NA	0.954	0.024	NA	NA	1.149	0.054	NA	NA
Tnks	100	1.847	0.045	NA	NA	1.726	0.065	NA	NA	0.962	0.018	NA	NA	1.317	0.033	NA	NA
Tom1l2	100	1.576	0.047	NA	NA	1.401	0.055	NA	NA	0.930	0.021	NA	NA	1.148	0.039	NA	NA
Top3a	100	1.726	0.027	NA	NA	1.438	0.068	NA	NA	0.914	0.025	NA	NA	1.397	0.061	NA	NA
Trabd2b	100	1.711	0.053	NA	NA	1.594	0.097	NA	NA	0.965	0.028	NA	NA	1.051	0.038	NA	NA
Trim16	100	1.873	0.085	NA	NA	2.298	0.167	NA	NA	1.059	0.033	NA	NA	1.652	0.070	NA	NA
Tvp23b	100	1.597	0.086	NA	NA	1.592	0.130	NA	NA	0.991	0.022	NA	NA	1.122	0.065	NA	NA
Vps13c	100	1.514	0.068	NA	NA	1.407	0.110	NA	NA	0.944	0.030	NA	NA	1.007	0.036	NA	NA

*OT_Rab10 – Rab10 KD using ON-TARGETplus siRNA pools

Supplementary Table S3.6 - GLUT4 screen results in 3T3-L1 wildtype cells (N = 3) for lipid droplets and nuclei
Average values across all insulin doses were calculated for each siRNA target, as these are not expected to change with differing doses of insulin.

	Lipid Droplets (LDs)				LDs/Nuclei number				Nuclei number			
			t-test siRNA target vs NT				t-test siRNA target vs NT				t-test siRNA target vs NT	
siRNA	**mean**	**se**	**p-value**	**FDR**	**mean**	**se**	**p-value**	**FDR**	**mean**	**se**	**p-value**	**FDR**
Akt1/2	536.924	112.756	0.704	0.852	0.214	0.034	0.334	0.700	2296.417	243.181	0.862	0.944
Alkbh5	996.533	21.164	0.002	0.080	0.336	0.007	0.009	0.084	3244.333	4.702	0.001	0.042
Atpaf2	501.492	138.634	0.525	0.805	0.224	0.032	0.715	0.843	2083.500	373.753	0.243	0.431
Bicc1	580.750	176.575	0.871	0.931	0.238	0.047	0.907	0.922	2277.000	375.972	0.824	0.944
Cast	676.383	34.608	0.416	0.736	0.247	0.009	0.750	0.843	2977.333	50.999	0.021	0.122
Cep295	729.383	12.758	0.231	0.638	0.292	0.004	0.136	0.510	2726.833	22.214	0.155	0.374
Cmpk1	455.483	105.216	0.271	0.638	0.195	0.022	0.164	0.540	2292.167	354.603	0.882	0.944
Drc3	397.666	113.924	0.044	0.249	0.176	0.029	0.012	0.093	2017.889	310.444	0.074	0.208
Drg2	867.950	91.962	0.028	0.249	0.329	0.029	0.015	0.097	2883.167	63.048	0.049	0.186
Erap1	530.392	158.335	0.730	0.852	0.224	0.045	0.712	0.843	2198.250	341.636	0.544	0.782
Eri1	650.217	52.846	0.533	0.805	0.271	0.020	0.341	0.700	2583.333	50.742	0.358	0.589
Fam13c	574.292	85.701	0.922	0.964	0.258	0.021	0.395	0.700	2272.000	217.139	0.805	0.944
Flii	373.308	97.930	0.054	0.249	0.210	0.035	0.379	0.700	1681.667	219.587	0.002	0.042
Gid4	459.958	64.469	0.292	0.638	0.202	0.013	0.254	0.687	2330.333	228.337	0.970	0.970
Irs1	467.558	78.160	0.328	0.638	0.207	0.022	0.333	0.700	2292.000	202.719	0.881	0.944
Llgl1	355.469	100.866	0.011	0.176	0.164	0.028	0.003	0.059	1948.222	277.470	0.028	0.131
Lnpep	682.108	203.355	0.236	0.638	0.259	0.050	0.374	0.700	2458.750	412.803	0.506	0.751
Mfhas1	457.900	140.223	0.197	0.638	0.206	0.039	0.226	0.682	1939.556	292.833	0.025	0.128
Mief1	402.600	123.114	0.050	0.249	0.192	0.036	0.072	0.333	1862.000	281.744	0.007	0.081

NT ctrl	564.599	41.032	NA	NA	0.234	0.011	NA	NA	2322.558	85.731	NA	NA
OT_Rab10	382.758	19.071	0.067	0.280	0.201	0.005	0.237	0.682	1896.333	51.504	0.038	0.158
Pdzk1ip1	526.844	92.711	0.648	0.805	0.222	0.021	0.604	0.841	2341.278	238.781	0.913	0.945
Ppp1r3b	607.950	149.038	0.600	0.805	0.283	0.039	0.039	0.197	2024.944	253.434	0.081	0.208
Prpsap2	463.889	104.176	0.223	0.638	0.220	0.030	0.548	0.841	2007.333	215.762	0.065	0.208
Rab10	498.414	84.956	0.283	0.638	0.223	0.024	0.504	0.827	2098.694	169.037	0.079	0.208
Riok2	519.294	154.140	0.648	0.805	0.230	0.049	0.880	0.920	2090.250	319.542	0.257	0.438
Shmt1	605.411	107.578	0.621	0.805	0.259	0.030	0.292	0.700	2296.833	212.195	0.880	0.944
Slc2a4	566.592	81.990	0.984	0.984	0.244	0.025	0.728	0.843	2399.500	162.190	0.707	0.904
Slc36a4	479.806	76.189	0.305	0.638	0.238	0.024	0.872	0.920	2021.000	140.707	0.077	0.208
Smco4	484.578	100.390	0.333	0.638	0.223	0.031	0.640	0.843	2091.944	169.434	0.177	0.387
Smcr8	569.511	158.893	0.953	0.974	0.247	0.045	0.595	0.841	2102.278	260.711	0.197	0.412
Tal1	613.200	171.435	0.624	0.805	0.237	0.050	0.922	0.922	2499.750	314.917	0.387	0.614
Tbc1D4	421.274	78.114	0.031	0.249	0.180	0.023	0.004	0.059	2158.333	192.858	0.231	0.426
Tfam	484.578	112.318	0.333	0.638	0.240	0.032	0.791	0.867	1897.556	220.659	0.013	0.099
Tnks	663.800	116.344	0.230	0.638	0.312	0.036	0.001	0.044	2118.389	168.936	0.232	0.426
Tom1l2	399.128	78.627	0.045	0.249	0.214	0.028	0.385	0.700	1825.444	146.930	0.004	0.056
Top3a	591.917	162.336	0.783	0.878	0.254	0.048	0.483	0.822	2243.583	311.855	0.700	0.904
Trabd2b	515.583	100.350	0.553	0.805	0.225	0.027	0.699	0.843	2245.333	212.051	0.651	0.880
Trim16	510.039	101.040	0.509	0.805	0.248	0.030	0.570	0.841	1994.889	192.406	0.055	0.195
Tvp23b	637.444	96.004	0.378	0.695	0.270	0.027	0.127	0.510	2365.444	164.128	0.801	0.944
Vps13c	616.517	90.935	0.530	0.805	0.242	0.027	0.751	0.843	2559.056	156.735	0.166	0.382

Supplementary Table S3.7 - Effect of knockdown of targets of interest on Glucose Transport in 3T3-L1 adipocytes upon stimulation with 0.5nM insulin

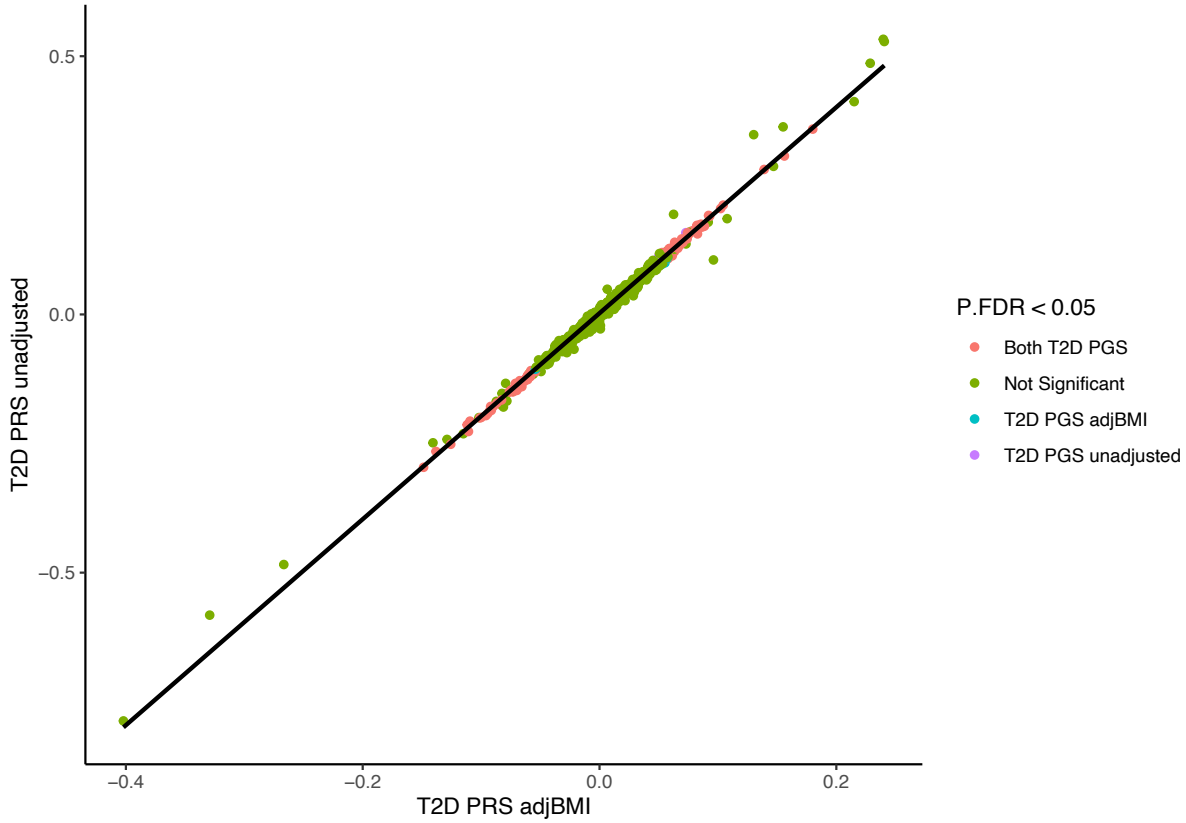
Insulin Dose (nM)		0	0.5	0	0.5		
siRNA KD	N	Glucose transport (pmole/mg/min)	Glucose transport (pmole/mg/min)	t-test P value	t-test P value	NT control compare	Control?
NT control A	3	49.75893	160.3982	NA	NA	A	NT control
Rab10	3	38.81059	78.35885	0.48072	0.00459	A	Positive Control
Llgl1	3	45.33971	115.084	0.574155	0.03006	A	
Flii	3	37.97219	119.4091	0.466168	0.037968	A	
Drg2	3	37.32225	90.21184	0.046223	0.022039	A	
Pdzk1ip1	3	40.23898	215.2586	0.341216	0.016515	A	
Slc36a4	3	42.67205	82.04302	0.6312	0.020431	A	
Vps13c	3	27.4984	101.2159	0.107012	0.031156	A	
Alkbh5	3	31.45314	55.79248	0.122756	0.001419	A	
Akt1/2	3	45.67779	68.66265	0.829253	0.005679	A	Positive Control
Cmpk1	3	52.21355	90.18453	0.829739	0.00735	A	
Tal1	3	42.59834	136.6786	0.001599	0.027223	A	
Tnks	3	49.15795	208.2936	0.960156	0.010722	A	
Tvp23b	3	43.31148	77.84615	0.858559	0.042204	A	
Tom1l2	3	16.94871	40.76831	0.024932	0.00516	A	
Bicc1	3	21.35208	81.15639	0.041854	0.01505	A	
Slc2a4	3	33.49529	81.85042	0.09036	0.005227	A	Positive Control
NT control b	3	35.83996	112.5332	NA	NA	B	NT control
Rab10	3	36.68385	73.02634	0.858988	0.014622	B	Positive Control
Tvp23b	3	21.94021	81.57117	0.021737	0.157995	B	no effect control
Prpsap2	3	53.34755	138.2689	0.156767	0.307902	B	no effect control
Atpaf2	3	63.70846	115.7141	0.025163	0.696101	B	no effect control
Trim16	3	40.24848	101.4391	0.17274	0.101588	B	no effect control
Cep295	3	32.95481	116.2322	0.475514	0.66931	B	no effect control
Fam13c	3	42.12253	143.9908	0.467873	0.309035	B	no effect control
Shmt1	3	34.42249	117.8719	0.699092	0.618672	B	no effect control
Smco4	3	55.63223	105.9474	0.278616	0.76559	B	no effect control
Akt1/2	3	35.51852	65.73061	0.945447	0.010049	B	Positive Control
Top3a	3	28.72903	120.1383	0.253692	0.672205	B	no effect control
Ppp1r3b	3	31.68659	113.8883	0.361259	0.939779	B	no effect control

Supplementary Material Chapter 4

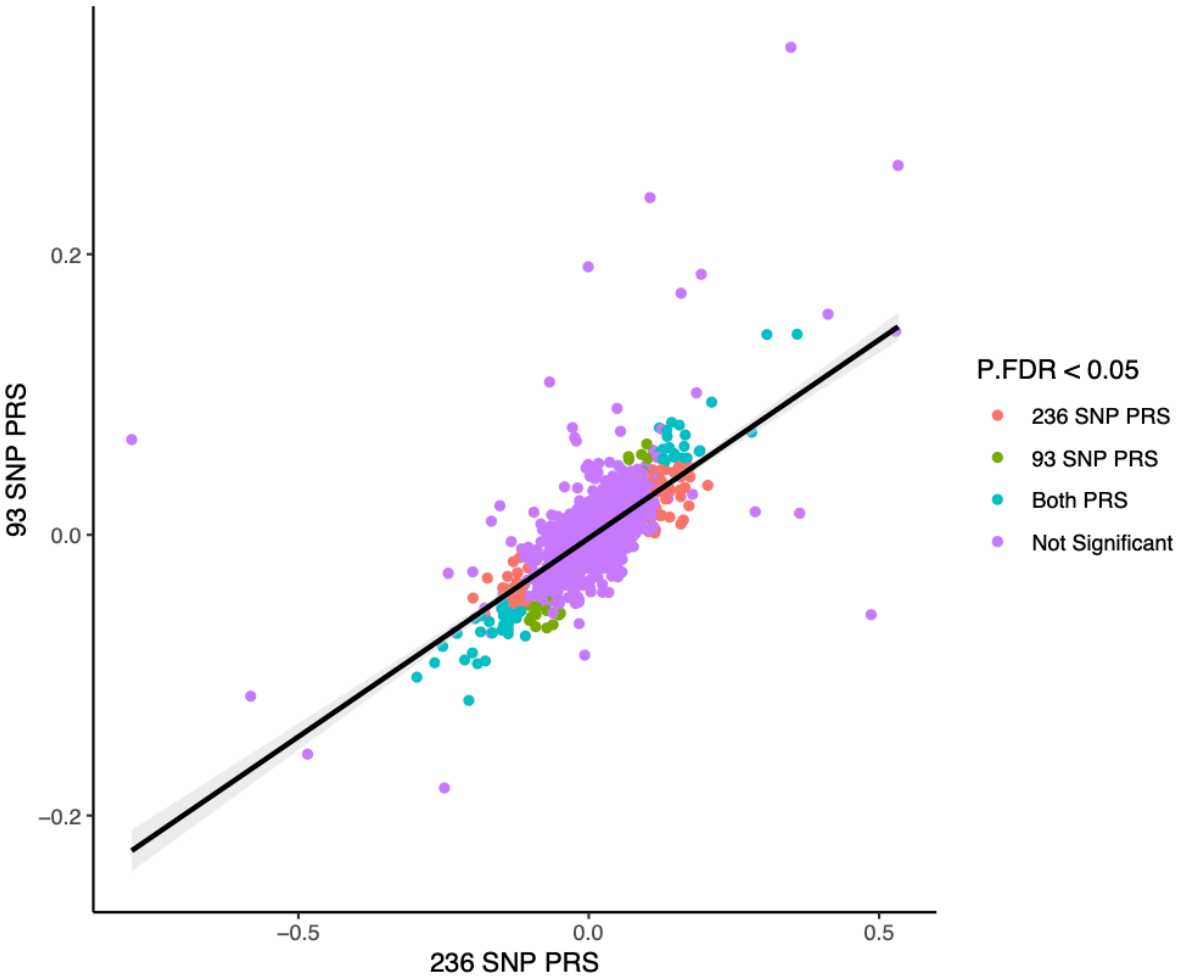
Supplementary Table S4.1 Systematic Review Search Terms used in PubMed

	Search Terms
Observational epidemiological studies #1	("Case-Control Studies"[MeSH] OR (Case-Control Study OR Case-Control Studies) OR (Case-control Study OR Case-control Studies) OR (Studies, Case-Control OR Study, Case-Control) OR (Case-Comparison Studies OR Case Comparison Studies) OR (Case-Comparison Study OR Studies), (Case-Comparison OR Study,Case- Comparison) OR (Case-Referent StudiesOR (Case Referent Studies OR Case- Referent Study)OR (Studies, Case-Referent OR Study, Case-Referent) OR (Case-Referent Studies OR Case Referent Studies) OR (Case-Referent Study OR Studies, Case-Referent OR Study, Case-Referent) OR (Case Control Studies OR Case Control Study OR Studies, Case Control OR Study, Case Control) OR Matched Case-Control Studies OR Case-Control Studies, Matched OR Case-Control Study, Matched OR Matched Case Control Studies OR Matched Case-Control Study OR Studies, Matched Case-Control OR Study, Matched Case-Control OR Nested Case-Control Studies OR Case-Control Studies, Nested OR Case-Control Study, Nested OR Nested Case Control Studies OR Nested Case-Control Study OR Studies, Nested Case-Control OR Study, Nested Case-Control OR "Cross-Sectional Studies"[MeSH] OR Cross-sectional Studies OR Cross Sectional Studies OR Cross Sectional Study OR Cross-Sectional Study OR Studies, Cross-Sectional OR Study, Cross-Sectional OR Cross Sectional Analysis OR Analyses, Cross Sectional OR Cross Sectional Analyses OR Disease Frequency Surveys OR Surveys, Disease Frequency OR Disease Frequency Survey OR Survey, Disease Frequency OR Analysis, Cross- Sectional OR Analyses, Cross-Sectional OR Analysis, Cross Sectional OR Cross- Sectional Analyses OR Cross-Sectional Analysis OR Cross-Sectional Survey OR Cross Sectional Survey OR Cross-Sectional Surveys OR Survey, Cross-Sectional OR Surveys, Cross-Sectional OR Prevalence Studies OR Prevalence Study OR Studies, Prevalence OR Study, Prevalence OR "Cohort Studies"[Mesh] OR Cohort Study OR Studies, Cohort OR Study, Cohort OR Studies, Historical Cohort OR Cohort Studies, Historical Cohort Study, Historical OR Historical Cohort Study OR Study, Historical OR Analysis, Cohort OR Analyses, Cohort OR Cohort Analyses OR Cohort Analysis OR Incidence Studies OR Incidence Study OR Studies, Incidence OR Study, Incidence OR "Incidence"[Mesh] OR "Cohort Studies"[Mesh] OR Case- cohort study OR Nested case-control study OR Survey OR cross-sectional OR case- control)
Type 2 Diabetes #2	"Diabetes mellitus, Type 2"[Mesh] OR NIDDM OR Maturity-Onset Diabetes OR Diabetes Mellitus, Noninsulin-Dependent OR Diabetes Mellitus, Adult-Onset OR Adult-Onset Diabetes Mellitus OR Diabetes Mellitus, Adult Onset OR Diabetes Mellitus, Maturity-Onset OR Diabetes Mellitus, Maturity Onset OR Diabetes Mellitus, Non-Insulin Dependent OR Diabetes Mellitus, Non-Insulin-Dependent OR Non-Insulin-Dependent Diabetes Mellitus OR Diabetes Mellitus, Noninsulin Dependent OR Diabetes Mellitus, Slow-Onset OR Diabetes Mellitus, Slow Onset OR Slow-Onset Diabetes Mellitus OR Diabetes Mellitus, Stable OR Stable Diabetes Mellitus OR Diabetes Mellitus, Type II OR Maturity-Onset Diabetes Mellitus OR Maturity Onset Diabetes Mellitus OR MODY OR Type 2 Diabetes Mellitus OR Noninsulin-Dependent Diabetes Mellitus OR "Type 2 Diabetes"[tiab] OR "Diabetes"[ti]
Metabolomics #3	metabolomics OR metabolome OR metabonomics OR metabolic profiling OR metabolite profiling OR metabolites OR metabolome OR lipid profiling OR amino acid profiling OR lipidomics OR metabolite profiles OR metabolite profile OR "mass spectrometry" [MeSH] OR mass spectrometry OR NMR OR nuclear magnetic resonance OR "NMR spectroscopy" [MeSH] OR LC-MS OR GC-MS
	#1 AND #2 AND #3

Supplementary Figures for Chapter 4

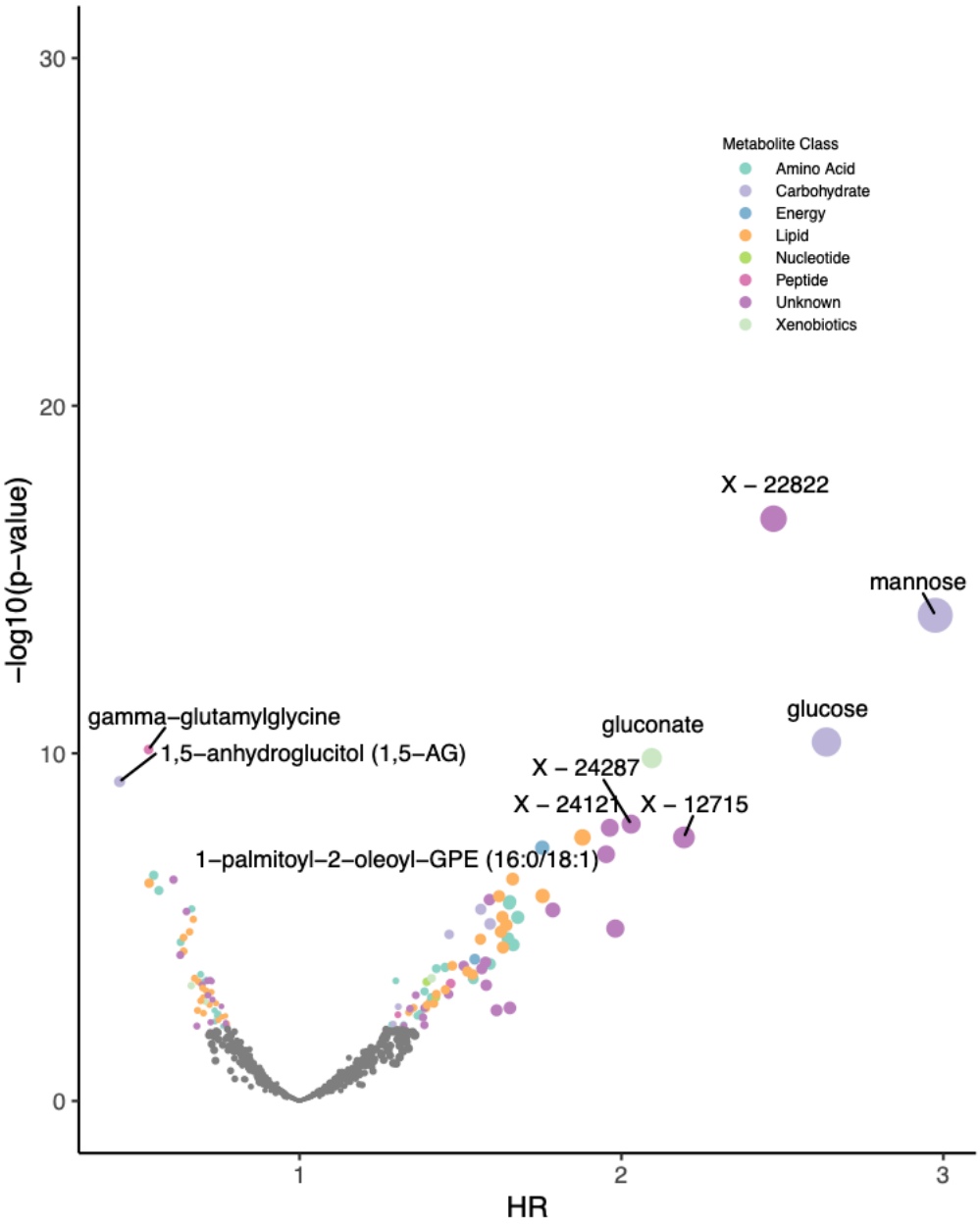


Supplementary Figure S4.1: Comparison of significant PGS-metabolite associations for BMI-unadjusted and adjusted T2D PGS
Beta estimates of metabolites significantly associated with BMI adjusted and unadjusted T2D 236 SNP PGS were compared. Beta estimates for these metabolites are similar across both scores. Significance in either analysis is indicated by colour of the points.



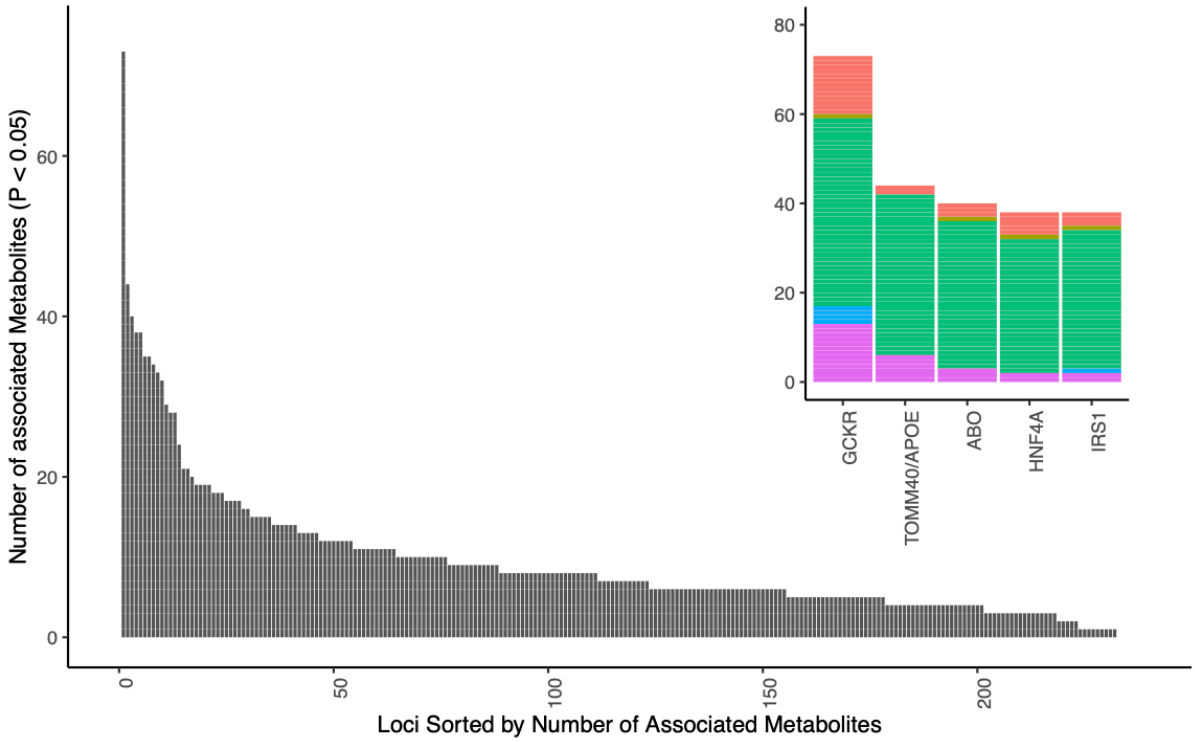
Supplementary Figure S4.2: Comparison of significant PGS-metabolite associations for the 236 and 93 SNP T2D PGS

Beta estimates of metabolites significantly associated with BMI adjusted and unadjusted T2D PGS were compared. Beta estimates for these metabolites are similar across both scores. Significance in either analysis is indicated by colour of the points.

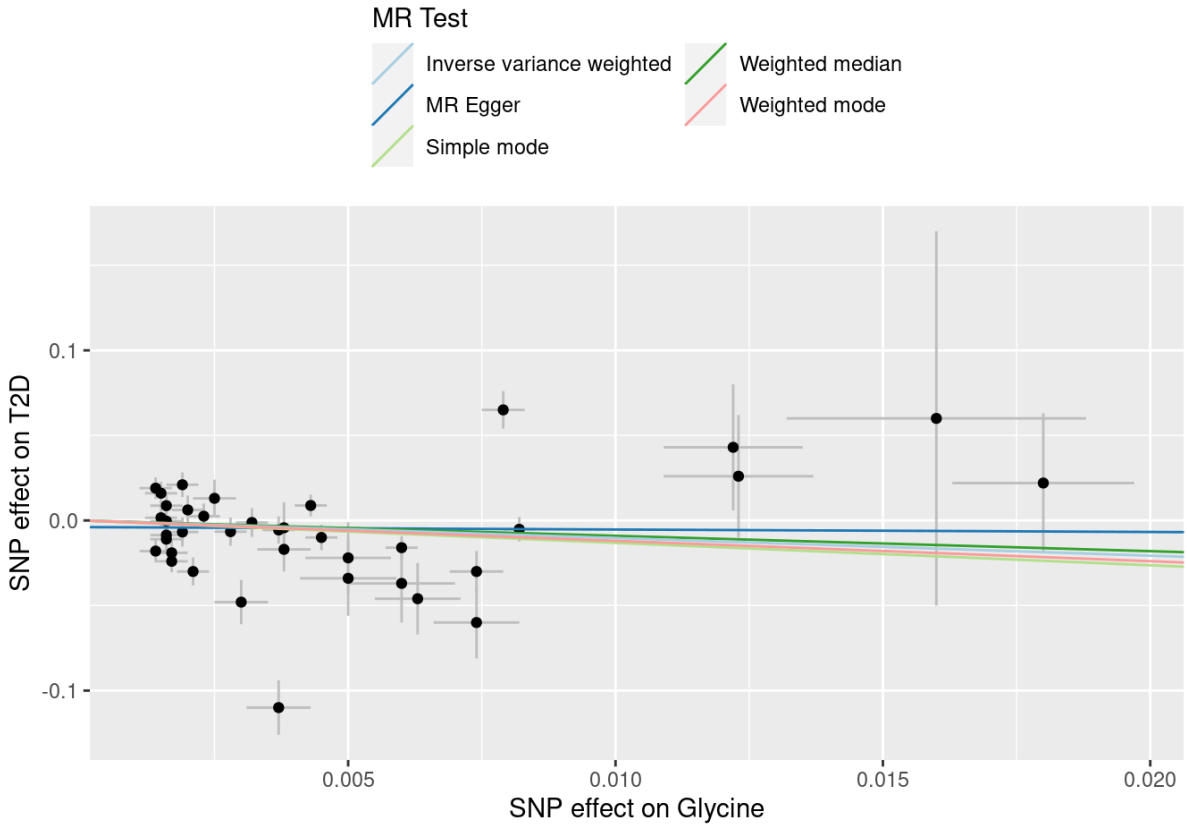


Supplementary Figure S4.3: Metabolomic signature for incident diabetes upon adjustment for established T2D risk factors.

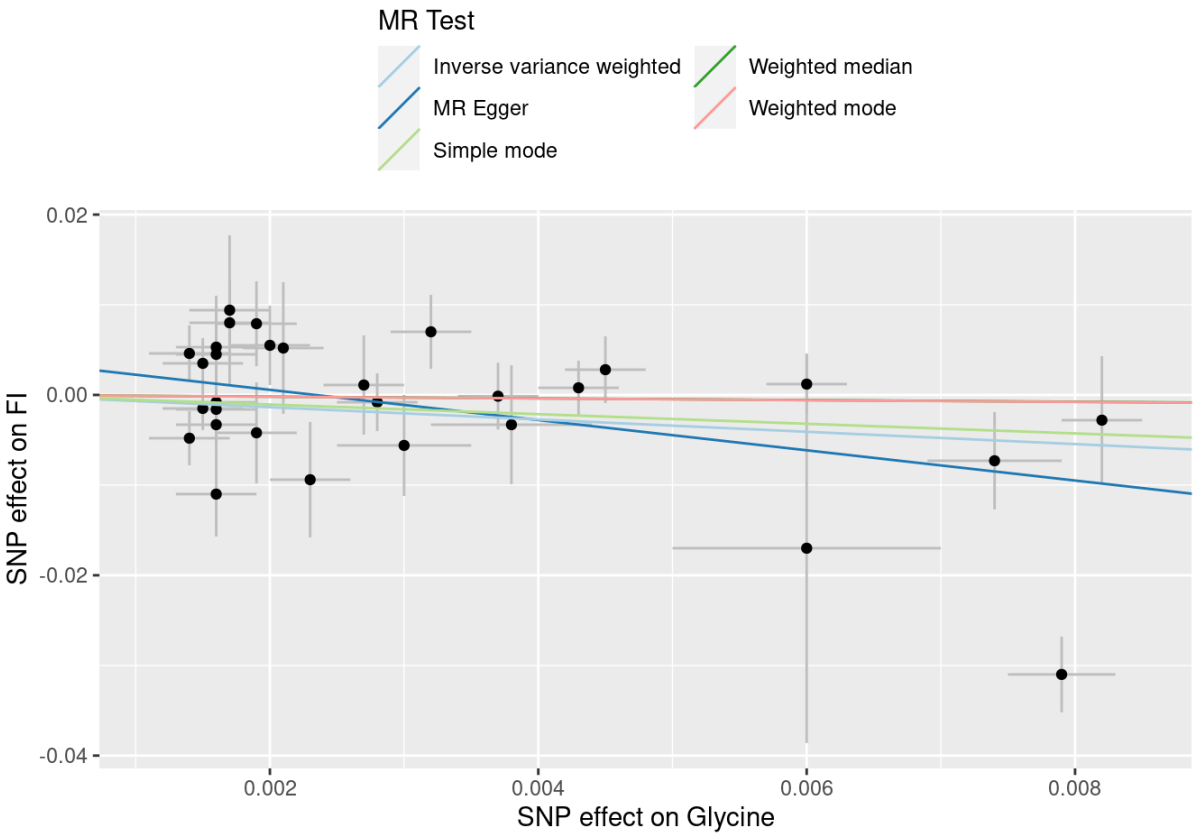
Further adjustment in addition to age and sex for: body mass index (BMI), waist-to-hip ratio (WHR), educational attainment, smoking status, alcohol consumption and physical activity. Coloured points represent metabolite class for significant associations. Non-significant associations are in grey. Point size proportional to HR.



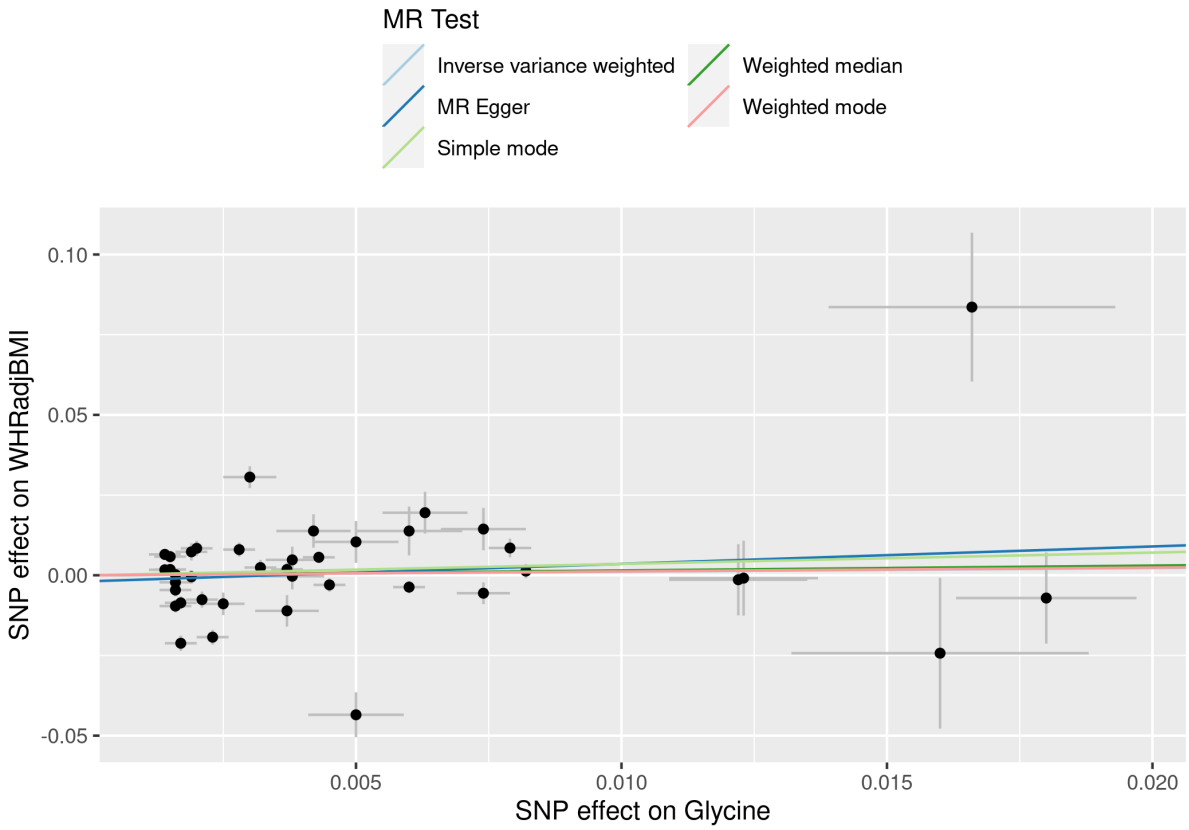
Supplementary Figure S4.4: number of metabolites nominally associated (P < 0.05) with each SNP in the T2D PGS.
Inset figure shows the breakdown of these association by metabolite calls for the 7 SNPs with the most nominally significant metabolite associations (P < 0.05). SNPs are named by their nearest gene as reported by Mahajan *et al* (2018). Coloured points represent metabolites significantly associated with the 91 SNP PGS (FDR < 0.05), with the top 10 associated metabolites labelled with their corresponding names.



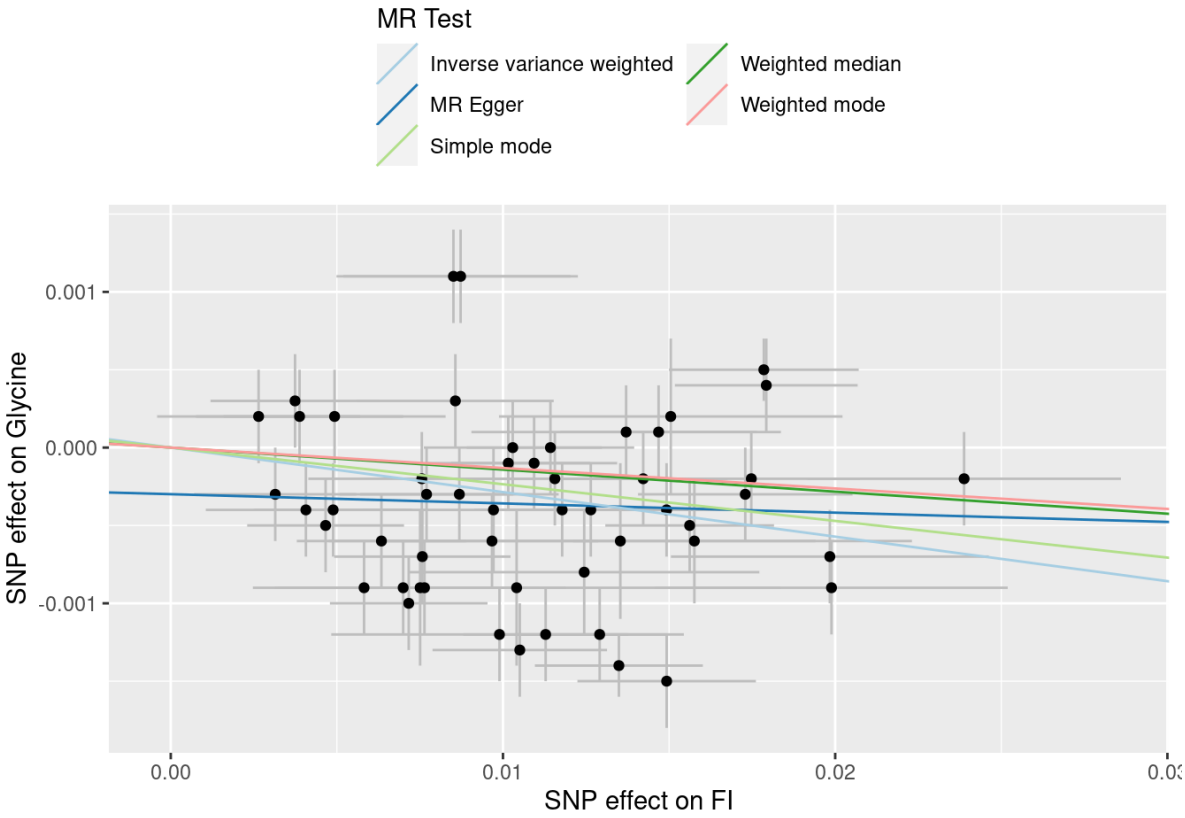
Supplementary Figure S4.5: Dosage plots of effect sizes of glycine variants on standard deviations of glycine levels versus the log odds for T2D unadjusted for BMI



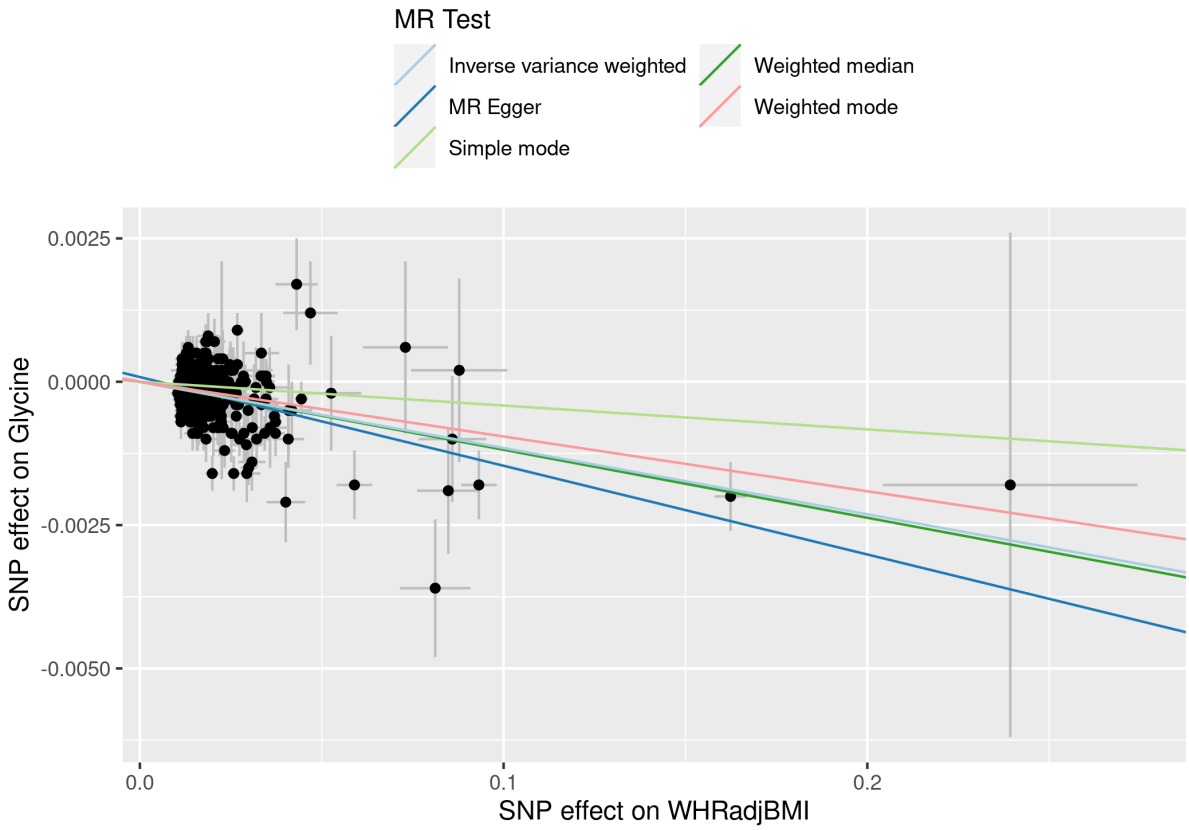
Supplementary Figure S4.6: Dosage plots of effect sizes of glycine variants on standard deviations of glycine levels versus the effect sizes for Fasting insulin unadjusted for BMI



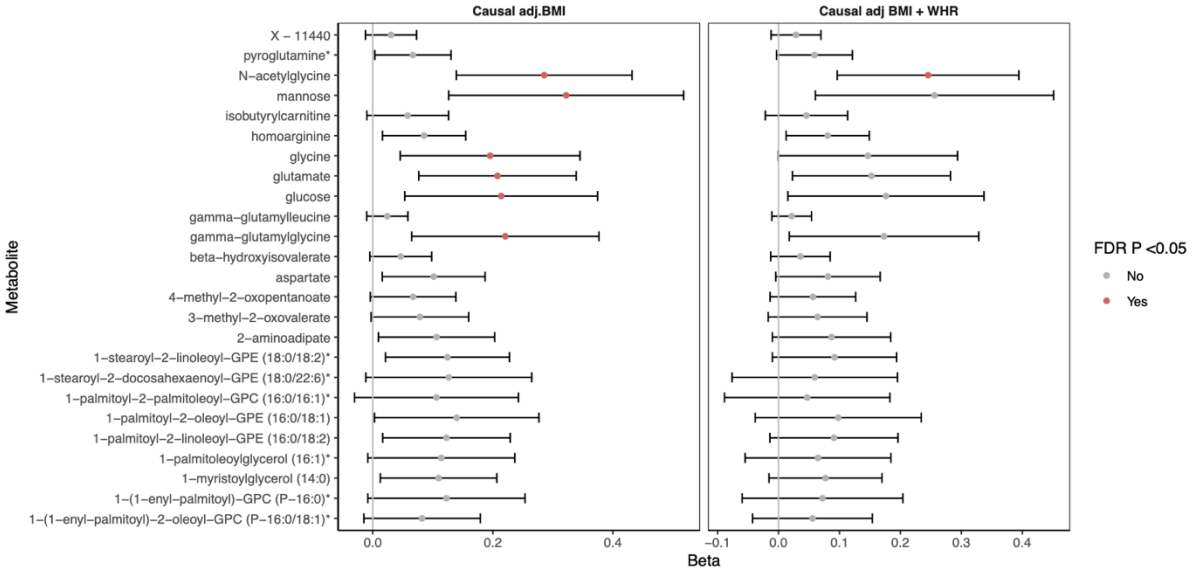
Supplementary Figure S4.7: Dosage plots of effect sizes of glycine variants on standard deviations of glycine levels versus the effect sizes for Waist to hip ratio adjusted for BMI (WHR adjBMI)



Supplementary Figure S4.8: Dosage plots of the effect sizes for fasting insulin unadjusted for BMI verses effect sizes of glycine variants on standard deviations of glycine levels



Supplementary Figure S4.9: Dosage plots of the effect sizes for Waist to hip ratio adjusted for BMI (WHR adjBMI) verses effect sizes of glycine variants on standard deviations of glycine levels



Supplementary Figure S4.10: additional adjustment with WHR attenuates almost all candidate causal mediator metabolites in the mediation framework

Supplementary Table S4.2: Identified studies examining the association of plasma metabolite levels with incident T2D in systematic literature review (up to 12th October 2022)

Publication	Study design	Study population	Incident T2D cases	Follow-up	Measurement technique or platform	Measured metabolite classes	References
Ahola-Olli, 2019	prospective cohort	4 cohorts - YFS, FINRISK-1997, DILGOM,NFBC	392	YFS-10 years, FINNRISK-1997 - 15 years, DILGOM-7.8 years, NFBC 15 years	NMR - Nightingale Health	229 metabolites	Ahola-Olli, A. V., Mustelin, L., Kalimeri, M., Kettunen, J., Jokelainen, J., Auvinen, J., Puukka, K., Havulinna, A. S., Lehtimäki, T., Kähönen, M., Juonala, M., Keinänen-Kiukaanniemi, S., Salomaa, V., Perola, M., Järvelin, M. R., Ala-Korpela, M., Raitakari, O., & Würtz, P. (2019). Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. Diabetologia, 62(12). https://doi.org/10.1007/s00125-019-05001-w
Bragg, 2022	Nested case-cohort	China Kadoorie Biobank	882	8 years	NMR	Targeted 225 metabolites	Bragg, F., Kartsonaki, C., Guo, Y., Holmes, M., Du, H., Yu, C., Pei, P., Yang, L., Jin, D., Chen, Y., Schmidt, D., Avery, D., Lv, J., Chen, J., Clarke, R., Hill, M., Li, L., Millwood, I., & Chen, Z. (2022). Circulating Metabolites and the Development of Type 2 Diabetes in Chinese Adults. *Diabetes Care*, *45*(2), 477–480. https://doi.org/10.2337/DC21-1415
Chen, 2016	Nested case-control and case-subcohort	Shanghai Diabetes Study	51	10 years	UPLC-MS	Amino acids	Chen, T., Ni, Y., Ma, X., Bao, Y., Liu, J., Huang, F., Hu, C., Xie, G., Zhao, A., Jia, W., & Jia, W. (2016). Branched-chain and aromatic amino acid profiles and diabetes risk in Chinese populations. Scientific Reports, 6(1), 20594. https://doi.org/10.1038/srep20594
Chen, 2020	pospective population based study	Hispanic Community Health Study/Study of Latinos (HCHS/SOL)	224	6 years	DiscoveryHD4 platform Metabolon	Untargeted approach (Metabolon discovery HD4) only 43 shingolipids reported	Chen, G. C., Chai, J. C., Yu, B., Michelotti, G. A., Grove, M. L., Fretts, A. M., Daviglus, M. L., Garcia-Bedoya, O. L., Thyagarajan, B., Schneiderman, N., Cai, J., Kaplan, R. C., Boerwinkle, E., & Qi, Q. (2020). Serum sphingolipids and incident diabetes in a US population with high diabetes burden: The Hispanic Community Health Study/Study of Latinos (HCHS/SOL). American Journal of Clinical Nutrition, 112(1), 57–65. https://doi.org/10.1093/ajcn/nqaa114
de Mello, 2015	Prospective cohort	Finnish Diabetes Prevention Study	130	8 years	GC-MS	Non-cholesterol sterols	de Mello, V. D. F., Lindström, J., Eriksson, J. G., Ilanne-Parikka, P., Keinänen-Kiukaanniemi, S., Pihlajamäki, J., Tuomilehto, J., & Uusitupa, M. (2015). Markers of cholesterol metabolism as biomarkers in predicting diabetes in the Finnish Diabetes Prevention Study. Nutrition, Metabolism and Cardiovascular Diseases, 25(7), 635–642. https://doi.org/10.1016/j.numecd.2015.03.012

Publication	Study design	Study population	Incident T2D cases	Follow-up	Measurement technique or platform	Measured metabolite classes	References
Drogan, 2015	Nested case-control	EPIC-Potsdam	300	6 years	UPLC-MS	Untargeted approach	Drogan, D., Dunn, W. B., Lin, W., Buijsse, B., Schulze, M. B., Langenberg, C., Brown, M., Floegel, A., Dietrich, S., Rolandsson, O., Wedge, D. C., Goodacre, R., Forouhi, N. G., Sharp, S. J., Spranger, J., Wareham, N. J., & Boeing, H. (2015). Untargeted Metabolic Profiling Identifies Altered Serum Metabolites of Type 2 Diabetes Mellitus in a Prospective, Nested Case Control Study. Clinical Chemistry, 61(3), 487–497. https://doi.org/10.1373/clinchem.2014.228965
Fall, 2016	Prospective cohort and nested case-cohort	ULSAM, PIVUS TwinGene, KORA S4	358	Not reported	UPLC-MS	Untargeted approach	Fall, T., Salihovic, S., Brandmaier, S., Nowak, C., Ganna, A., Gustafsson, S., Broeckling, C. D., Prenni, J. E., Kastenmüller, G., Peters, A., Magnusson, P. K., Wang-Sattler, R., Giedraitis, V., Berne, C., Gieger, C., Pedersen, N. L., Ingelsson, E., & Lind, L. (2016). Non-targeted metabolomics combined with genetic analyses identifies bile acid synthesis and phospholipid metabolism as being associated with incident type 2 diabetes. Diabetologia, 59(10), 2114–2124. https://doi.org/10.1007/s00125-016-4041-1
Fenske, 2015	Prospective cohort	Study of Health in Pomerania	226	5 years	LC-MS	androstenedione, total tostesterone	Fenske, B., Kische, H., Gross, S., Wallaschofski, H., Völzke, H., Dörr, M., Nauck, M., Keevil, B. G., Brabant, G., & Haring, R. (2015). Endogenous Androgens and Sex Hormone–Binding Globulin in Women and Risk of Metabolic Syndrome and Type 2 Diabetes. The Journal of Clinical Endocrinology & Metabolism, 100(12), 4595–4603. https://doi.org/10.1210/jc.2015-2546
Festa, 2005	Prospective cohort	Insulin resistance atherosclerosis study (IRAS)	130	5.2 years	^{1}H-NMR	Lipoprotein particles	Festa, A., Williams, K., Hanley, A. J. G., Otvos, J. D., Goff, D. C., Wagenknecht, L. E., & Haffner, S. M. (2005). Nuclear Magnetic Resonance Lipoprotein Abnormalities in Prediabetic Subjects in the Insulin Resistance Atherosclerosis Study. Circulation, 111(25), 3465–3472. https://doi.org/10.1161/CIRCULATIONAHA.104.512079
Floegel, 2013	Case-cohort	EPIC-Potsdam	800	7 years	AbsoluteIDQ p150 Biocrates	Hexose, acylcarnitines, amino acids, hexose and phosphatidylcholines	Floegel, A., Stefan, N., Yu, Z., Mühlenbruch, K., Drogan, D., Joost, H.-G., Fritsche, A., Häring, H.-U., Hrabě de Angelis, M., Peters, A., Roden, M., Prehn, C., Wang-Sattler, R., Illig, T., Schulze, M. B., Adamski, J., Boeing, H., & Pischon, T. (2013). Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. Diabetes, 62(2), 639–648. https://doi.org/10.2337/db12-0495

Publication	Study design	Study population	Incident T2D cases	Follow-up	Measurement technique or platform	Measured metabolite classes	References
Forouhi, 2014	Case-cohort	EPIC-InterAct	12,132	7.8 years	GC	Saturated fatty acids in plasma phospholipids	Forouhi, N. G., Koulman, A., Sharp, S. J., Imamura, F., Kröger, J., Schulze, M. B., Crowe, F. L., Huerta, J. M., Guevara, M., Beulens, J. W., van Woudenbergh, G. J., Wang, L., Summerhill, K., Griffin, J. L., Feskens, E. J., Amiano, P., Boeing, H., Clavel-Chapelon, F., Dartois, L., … Wareham, N. J. (2014). Differences in the prospective association between individual plasma phospholipid saturated fatty acids and incident type 2 diabetes: the EPIC-InterAct case-cohort study. The Lancet Diabetes & Endocrinology, 2(10), 810–818. https://doi.org/10.1016/S2213-8587(14)70146-9
Fretts, 2019	12 prospective cohort studies	FORCE consortium	14276 (across the 12 studies)	various	GC	VLSFA: 20:0, 22:0, or 24:0	Fretts, A. M., Imamura, F., Marklund, M., Micha, R., Wu, J. H. Y., Murphy, R. A., Chien, K. L., McKnight, B., Tintle, N., Forouhi, N. G., Qureshi, W. T., Virtanen, J. K., Wong, K., Wood, A. C., Lankinen, M., Rajaobelina, K., Harris, T. B., Djoussé, L., Harris, B., … Lemaitre, R. N. (2019). Associations of circulating very-long-chain saturated fatty acids and incident type 2 diabetes: A pooled analysis of prospective cohort studies. American Journal of Clinical Nutrition, 109(4), 1216–1223. https://doi.org/10.1093/ajcn/nqz005
Hodge, 2009	Case-cohort	Melbourne Collaborative Cohort Study	59	4 years	^{1}H-NMR	Lipoprotein particles	Hodge, A. M., Jenkins, A. J., English, D. R., O'Dea, K., & Giles, G. G. (2009). NMR-determined lipoprotein subclass profile predicts type 2 diabetes. Diabetes Research and Clinical Practice, 83(1), 132–139. https://doi.org/10.1016/j.diabres.2008.11.007
Joyce, 2017	Prospective cohort	Cardiovascular Health Study	112	9.8 years	LC-MS	Testosterone and dihydrotestosterone	Joyce, K. E., Biggs, M. L., Djoussé, L., Ix, J. H., Kizer, J. R., Siscovick, D. S., Shores, M. M., Matsumoto, A. M., & Mukamal, K. J. (2017). Testosterone, Dihydrotestosterone, Sex Hormone-Binding Globulin, and Incident Diabetes Among Older Men: The Cardiovascular Health Study. The Journal of Clinical Endocrinology and Metabolism, 102(1), 33–39. https://doi.org/10.1210/jc.2016-2623
Liu, 2017	Prospective cohort	Erasmus Rucphen Family study	137	11.3 years	LC-MS, NMR	triglycerides, glycerophospholipids, ceramides, amino acids and other low-weight molecules, lipoprotein particle size sub-fractions	Liu, J., Semiz, S., van der Lee, S. J., van der Spek, A., Verhoeven, A., van Klinken, J. B., Sijbrands, E., Harms, A. C., Hankemeier, T., van Dijk, K. W., van Duijn, C. M., & Demirkan, A. (2017). Metabolomics based markers predict type 2 diabetes in a 14-year follow-up study. Metabolomics, 13(9), 104. https://doi.org/10.1007/s11306-017-1239-2

Publication	Study design	Study population	Incident T2D cases	Follow-up	Measurement technique or platform	Measured metabolite classes	References
Lu, 2016	Nested case-control	Singapore Chinese Health Study	197	6 years	LC-MS, GC-MS	Untargeted approach	Lu, Y., Wang, Y., Ong, C. N., Subramaniam, T., Choi, H. W., Yuan, J. M., Koh, W. P., & Pan, A. (2016). Metabolic signatures and risk of type 2 diabetes in a Chinese population: an untargeted metabolomics study using both LC-MS and GC-MS. Diabetologia, 59(11), 2349–2359. https://doi.org/10.1007/s00125-016-4069-2
Lu, 2018	Nested case-control	Singapore Chinese Health Study	160	6 years	HPLC-MS	lysophosphatidylinositols, non-esterified fatty acids, acylcarnitines	Lu, Y., Wang, Y., Zou, L., Liang, X., Ong, C. N., Tavintharan, S., Yuan, J.-M., Koh, W.-P., & Pan, A. (2018). Serum Lipids in Association With Type 2 Diabetes Risk and Prevalence in a Chinese Population. The Journal of Clinical Endocrinology & Metabolism, 103(2), 671–680. https://doi.org/10.1210/jc.2017-02176
Mahendran, 2013	Prospective cohort	METSIM	176	4.5 years	^{1}H-NMR	Glycerol and fatty acids	Mahendran, Y., Cederberg, H., Vangipurapu, J., Kangas, A. J., Soininen, P., Kuusisto, J., Uusitupa, M., Ala-Korpela, M., & Laakso, M. (2013). Glycerol and Fatty Acids in Serum Predict the Development of Hyperglycemia and Type 2 Diabetes in Finnish Men. Diabetes Care, 36(11), 3732–3738. https://doi.org/10.2337/dc13-0800
Merino, 2018	Prospective cohort	Framingham Heart Study Offspring cohort	95	20 years	LC-MS	Amino acids, amines and other polar metabolites	Merino, J., Leong, A., Liu, C.-T., Porneala, B., Walford, G. A., von Grotthuss, M., Wang, T. J., Flannick, J., Dupuis, J., Levy, D., Gerszten, R. E., Florez, J. C., & Meigs, J. B. (2018). Metabolomics insights into early type 2 diabetes pathogenesis and detection in individuals with normal fasting glucose. Diabetologia, 61(6), 1315–1324. https://doi.org/10.1007/s00125-018-4599-x
Molnos, 2018	Prospective cohort, nested case-control	KORA and EPIC-Potsdam	910	7 years	AbsoluteIDQ p150 Biocrates	Metabolite ratios	Molnos, S., Wahl, S., Haid, M., Marelise, E., Eekhoff, W., Pool, R., Floegel, A., Deelen, J., Much, D., Prehn, C., Breier, M., Draisma, H. H., Van Leeuwen, N., Simonis-Bik, A. M. C., Jonsson, A., Gonneke Willemsen, &, Bernigau, W., Wang-Sattler, R., Suhre, K., … Leen, &. (2018). Metabolite ratios as potential biomarkers for type 2 diabetes: a DIRECT study. 12, 28. https://doi.org/10.1007/s00125-017-4436-7
Mora, 2010	Prospective cohort	Women's Health Study	1,687	13 years	^{1}H-NMR	Lipoprotein particles	Mora, S., Otvos, J. D., Rosenson, R. S., Pradhan, A., Buring, J. E., & Ridker, P. M. (2010). Lipoprotein Particle Size and Concentration by Nuclear Magnetic Resonance and Incident Type 2 Diabetes in Women. Diabetes, 59(5), 1153–1160. https://doi.org/10.2337/db09-1114

Publication	Study design	Study population	Incident T2D cases	Follow-up	Measurement technique or platform	Measured metabolite classes	References
Muilwijk, 2020	Nested case-subcohort	Healthy Life in an Urban Setting (HELIUS) study	95	4 years	tandem-MS (amino acids and acylcarnitines). (Glyco)sphingolipids LC-MS/MS	13 sphingolipid species,17 Amino acids; 27 acylcarnitine species	Muilwijk, M., Goorden, S. M. I., Celis-Morales, C., Hof, M. H., Ghauharali-van der Vlugt, K., Beers-Stet, F. S., Gill, J. M. R., Vaz, F. M., & van Valkengoed, I. G. M. (2020). Contributions of amino acid, acylcarnitine and sphingolipid profiles to type 2 diabetes risk among South-Asian Surinamese and Dutch adults. BMJ Open Diabetes Research & Care, 8(1). https://doi.org/10.1136/bmjdrc-2019-001003
Mwinyi, 2017	Nested case-control	CoLaus	251	5 years	LC-MS	Sphingoid bases	Mwinyi, J., Boström, A., Fehrer, I., Othman, A., Waeber, G., Marti-Soler, H., Vollenweider, P., Marques-Vidal, P., Schiöth, H. B., von Eckardstein, A., & Hornemann, T. (2017). Plasma 1-deoxysphingolipids are early predictors of incident type 2 diabetes mellitus. PLOS ONE, 12(5), e0175776. https://doi.org/10.1371/journal.pone.0175776
Othman, 2015	Prospective cohort	CAD patients	32	8 years	LC-MS	Sphingoid bases	Othman, A., Saely, C. H., Muendlein, A., Vonbank, A., Drexel, H., von Eckardstein, A., & Hornemann, T. (2015). Plasma 1-deoxysphingolipids are predictive biomarkers for type 2 diabetes mellitus. BMJ Open Diabetes Research & Care, 3, e000073. https://doi.org/10.1136/bmjdrc-2014-000073
Padberg, 2014	Retrospective study	Blood donors from the Bavarian Red Cross	28	6 years	MxPTM Broad Profiling Platform	Untargeted approach and targeted measurement of eicosanoids	Padberg, I., Peter, E., González-Maldonado, S., Witt, H., Mueller, M., Weis, T., Bethan, B., Liebenberg, V., Wiemer, J., Katus, H. A., Rein, D., & Schatz, P. (2014). A New Metabolomic Signature in Type-2 Diabetes Mellitus and Its Pathophysiology. PLoS ONE, 9(1), e85082. https://doi.org/10.1371/journal.pone.0085082
Palmer, 2015	Prospective cohort	IRAS	76	5 years	MS/MS	Amino acids and acylcarnitines	Palmer, N. D., Stevens, R. D., Antinozzi, P. A., Anderson, A., Bergman, R. N., Wagenknecht, L. E., Newgard, C. B., & Bowden, D. W. (2015). Metabolomic Profile Associated With Insulin Resistance and Conversion to Diabetes in the Insulin Resistance Atherosclerosis Study. The Journal of Clinical Endocrinology & Metabolism, 100(3), E463–E468. https://doi.org/10.1210/jc.2014-2357
Parnell, 2021	Prospective cohort	Boston Puerto Rican Health Study (BPRHS)	66	5 years	Metabolon	Untargeted approach	Parnell, L. D., Noel, S. E., Bhupathiraju, S. N., Smith, C. E., Haslam, D. E., Zhang, X., Tucker, K. L., Ordovas, J. M., & Lai, C. Q. (2021). Metabolite patterns link diet, obesity, and type 2 diabetes in a Hispanic population. *Metabolomics*, *17*(10), 88. https://doi.org/10.1007/s11306-021-01835-x
Peddinti, 2017	Nested case-cohort	Botnia Prospective Study	146	10 years	UHPLC-MS, GC-MS	Untargeted approach	Peddinti, G., Cobb, J., Yengo, L., Froguel, P., Kravić, J., Balkau, B., Tuomi, T., Aittokallio, T., & Groop, L. (2017). Early metabolic markers identify potential targets for the prevention of

Publication	Study design	Study population	Incident T2D cases	Follow-up	Measurement technique or platform	Measured metabolite classes	References
							type 2 diabetes. Diabetologia, 60(9), 1740–1750. https://doi.org/10.1007/s00125-017-4325-0
Qiu, 2016	Nested case-control	Dongfeng-Tongji and Jiangsu Non-communicable Disease cohorts	1039 and 520	4.6 and 7.6 years	HPLC-MS	Amino acids, acylcarnitines, cholines, amines, purine and indole derivatives, B vitamins	Qiu, G., Zheng, Y., Wang, H., Sun, J., Ma, H., Xiao, Y., Li, Y., Yuan, Y., Yang, H., Li, X., Min, X., Zhang, C., Xu, C., Jiang, Y., Zhang, X., He, M., Yang, M., Hu, Z., Tang, H., … Wu, T. (2016). Plasma metabolomics identified novel metabolites associated with risk of type 2 diabetes in two prospective cohorts of Chinese adults. International Journal of Epidemiology, 45(5), 1507–1516. https://doi.org/10.1093/ije/dyw221
Razquin, 2019	Unstratified Case-cohort study nested within PREDIMED trial	PREDIMED	250	3.8 years median follow up	LC-MS	207 lipid metabolites	Razquin, C., Ruiz-Canela, M., Clish, C. B., Li, J., Toledo, E., Dennis, C., Liang, L., Salas-Huetos, A., Pierce, K. A., Guasch-Ferré, M., Corella, D., Ros, E., Estruch, R., Gómez-Gracia, E., Fitó, M., Lapetra, J., Romaguera, D., Alonso-Gómez, A., Serra-Majem, L., … Martínez-González, M. A. (2019). Lysine pathway metabolites and the risk of type 2 diabetes and cardiovascular disease in the PREDIMED study: Results from two case-cohort studies. Cardiovascular Diabetology, 18(1). https://doi.org/10.1186/s12933-019-0958-2
Rebholz, 2018	Prospective cohort	ARIC	1,126	20 years	DiscoveryHD4 platform Metabolon	Untargeted approach	Rebholz, C. M., Yu, B., Zheng, Z., Chang, P., Tin, A., Köttgen, A., Wagenknecht, L. E., Coresh, J., Boerwinkle, E., & Selvin, E. (2018). Serum metabolomic profile of incident diabetes. Diabetologia, 61(5), 1046–1054. https://doi.org/10.1007/s00125-018-4573-7
Rhee, 2011	Nested case-control	Framingham Heart Study	189	12 years	LC-MS	Triglycerides, cholesterol esters, phosphatidylcholines, lysophosphatidylethanolamines, diglycerides and sphingomyelines	Rhee, E. P., Cheng, S., Larson, M. G., Walford, G. A., Lewis, G. D., McCabe, E., Yang, E., Farrell, L., Fox, C. S., O'Donnell, C. J., Carr, S. A., Vasan, R. S., Florez, J. C., Clish, C. B., Wang, T. J., & Gerszten, R. E. (2011). Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. Journal of Clinical Investigation, 121(4), 1402–1411. https://doi.org/10.1172/JCI44442
Savolainen, 2017	Prospective cohort	Cohort of Swedish women	69	5.5 years	GC-MS	α- and δ-tocopherol, alkylresorcinols, β-alanine, CMPF, fatty acids	Savolainen, O., Fagerberg, B., Vendelbo Lind, M., Sandberg, A.-S., Ross, A. B., & Bergström, G. (2017). Biomarkers for predicting type 2 diabetes development—Can metabolomics improve on existing biomarkers? PLOS ONE, 12(7), e0177738. https://doi.org/10.1371/journal.pone.0177738

Publication	Study design	Study population	Incident T2D cases	Follow-up	Measurement technique or platform	Measured metabolite classes	References
Savolainen, 2017	Prospective cohort	Cohort of Swedish women	69	5.5 years	GC-MS	Not specified	Savolainen, O., Lind, M. V., Bergström, G., Fagerberg, B., Sandberg, A.-S., & Ross, A. (2017). Biomarkers of food intake and nutrient status are associated with glucose tolerance status and development of type 2 diabetes in older Swedish women. The American Journal of Clinical Nutrition, ajcn152850. https://doi.org/10.3945/ajcn.117.152850
Shi, 2018	Nested case-control	Västerbotten Intervention Programme cohort	503	7 years	HPLC-MS	Untargeted approach	Shi, L., Brunius, C., Lehtonen, M., Auriola, S., Bergdahl, I. A., Rolandsson, O., Hanhineva, K., & Landberg, R. (2018). Plasma metabolites associated with type 2 diabetes in a Swedish population: a case–control study nested in a prospective cohort. Diabetologia, 61(4), 849–861. https://doi.org/10.1007/s00125-017-4521-y
Stančáková, 2012	Prospective cohort	METSIM	151	4.7 years	^{1}H-NMR	Amino acids	Stancakova, A., Civelek, M., Saleem, N. K., Soininen, P., Kangas, A. J., Cederberg, H., Paananen, J., Pihlajamaki, J., Bonnycastle, L. L., Morken, M. A., Boehnke, M., Pajukanta, P., Lusis, A. J., Collins, F. S., Kuusisto, J., Ala-Korpela, M., & Laakso, M. (2012). Hyperglycemia and a Common Variant of GCKR Are Associated With the Levels of Eight Amino Acids in 9,369 Finnish Men. Diabetes, 61(7), 1895–1902. https://doi.org/10.2337/db11-1378
Sun, 2016	Prospective cohort	Nutrition and Health of Aging Pop- ulation in China study	507	6 years	LC-MS	Acylcarnitines	Sun, L., Liang, L., Gao, X., Zhang, H., Yao, P., Hu, Y., Ma, Y., Wang, F., Jin, Q., Li, H., Li, R., Liu, Y., Hu, F. B., Zeng, R., Lin, X., & Wu, J. (2016). Early Prediction of Developing Type 2 Diabetes by Plasma Acylcarnitines: A Population-Based Study. Diabetes Care, 39(9), 1563–1570. https://doi.org/10.2337/dc16-0232
Suvitaival, 2018	Prospective cohort	METSIM	107	5 years	UPLC-MS	glycerophospholipids, triacylglycerols, cholesterolesters, sphingomyelins and ceramides	Suvitaival, T., Bondia-Pons, I., Yetukuri, L., Pöhö, P., Nolan, J. J., Hyötyläinen, T., Kuusisto, J., & Orešič, M. (2018). Lipidome as a predictive tool in progression to type 2 diabetes in Finnish men. Metabolism, 78, 1–12. https://doi.org/10.1016/j.metabol.2017.08.014

Publication	Study design	Study population	Incident T2D cases	Follow-up	Measurement technique or platform	Measured metabolite classes	References
Svingen, 2016	Prospective cohort	Norwegian cohort of patients undergoing coronary angiography for stable angina pectoris	233	7.5 years	LC-MS, GC-MS	TMAO, choline, betaine, dimethylglycine and sarcosine	Svingen, G. F. T., Schartum-Hansen, H., Pedersen, E. R., Ueland, P. M., Tell, G. S., Mellgren, G., Njolstad, P. R., Seifert, R., Strand, E., Karlsson, T., & Nygard, O. (2016). Prospective Associations of Systemic and Urinary Choline Metabolites with Incident Type 2 Diabetes. Clinical Chemistry, 62(5), 755–765. https://doi.org/10.1373/clinchem.2015.250761
Tillin, 2015	Prospective cohort	SABRE	340	19 years	^{1}H-NMR	Amino acids	Tillin, T., Hughes, A. D., Wang, Q., Würtz, P., Ala-Korpela, M., Sattar, N., Forouhi, N. G., Godsland, I. F., Eastwood, S. V., McKeigue, P. M., & Chaturvedi, N. (2015). Diabetes risk and amino acid profiles: cross-sectional and prospective analyses of ethnicity, amino acids and diabetes in a South Asian and European cohort from the SABRE (Southall And Brent REvisited) Study. Diabetologia, 58(5), 968–979. https://doi.org/10.1007/s00125-015-3517-8
Wang-Sattler ,2012	Prospective cohort	Cooperative Health Research in the Region of Augsburg (KORA)	91	7 years	AbsoluteIDQ p180 Biocrates	Hexose, acylcarnitines, amino acids, biogenic amines, sphingomyelins and phosphatidylcholines	Wang-Sattler, R., Yu, Z., Herder, C., Messias, A. C., Floegel, A., He, Y., Heim, K., Campillos, M., Holzapfel, C., Thorand, B., Grallert, H., Xu, T., Bader, E., Huth, C., Mittelstrass, K., Döring, A., Meisinger, C., Gieger, C., Prehn, C., … Illig, T. (2012). Novel biomarkers for pre-diabetes identified by metabolomics. Molecular Systems Biology. https://doi.org/10.1038/msb.2012.43
Wang, 2011	Nested case-control	Framingham Heart Study	189	12 years	LC-MS	Amino acids, amines and other polar metabolites	Wang, T. J., Larson, M. G., Vasan, R. S., Cheng, S., Rhee, E. P., McCabe, E., Lewis, G. D., Fox, C. S., Jacques, P. F., Fernandez, C., O'Donnell, C. J., Carr, S. A., Mootha, V. K., Florez, J. C., Souza, A., Melander, O., Clish, C. B., & Gerszten, R. E. (2011). Metabolite profiles and the risk of developing diabetes. Nature Medicine, 17(4), 448–453. https://doi.org/10.1038/nm.2307
Wang, 2013	Nested case-control	Framingham Heart Study	188	12 years	LC-MS/MS	Organic acids, pyrimidines, purines and other metabolites	Wang, T. J., Ngo, D., Psychogios, N., Dejam, A., Larson, M. G., Vasan, R. S., Ghorbani, A., O'Sullivan, J., Cheng, S., Rhee, E. P., Sinha, S., McCabe, E., Fox, C. S., O'Donnell, C. J., Ho, J. E., Florez, J. C., Magnusson, M., Pierce, K. A., Souza, A. L., … Gerszten, R. E. (2013). 2-Aminoadipic acid is a biomarker for diabetes risk. The Journal of Clinical Investigation, 123(10), 4309–4317. https://doi.org/10.1172/JCI64801

Publication	Study design	Study population	Incident T2D cases	Follow-up	Measurement technique or platform	Measured metabolite classes	References
Wang, 2022	Nested case-cohort	China Cardiometabolic Disease and Cancer Cohort (4C) Study	1,707	5 years	UPLC-MS/MS	28 serum amino acids and 22 microbiota related metabolites	Wang, S., Li, M., Lin, H., Wang, G., Xu, Y., Zhao, X., Hu, C., Zhang, Y., Zheng, R., Hu, R., Shi, L., Du, R., Su, Q., Wang, J., Chen, Y., Yu, X., Yan, L., Wang, T., Zhao, Z., … Lu, J. (2022). Amino acids, microbiota-related metabolites, and the risk of incident diabetes among normoglycemic Chinese adults: Findings from the 4C study. *Cell Reports Medicine*, *3*(9), 100727. https://doi.org/10.1016/J.XCRM.2022.100727
Wigger, 2017	Nested case-cohort, nested case-control	D.E.S.I.R. and CoLaus	189 and 147	9 and 5 years	UPLC-MS	Ceramides and dihydroceramides	Wigger, L., Cruciani-Guglielmacci, C., Nicolas, A., Denom, J., Fernandez, N., Fumeron, F., Marques-Vidal, P., Ktorza, A., Kramer, W., Schulte, A., Le Stunff, H., Liechti, R., Xenarios, I., Vollenweider, P., Waeber, G., Uphues, I., Roussel, R., Magnan, C., Ibberson, M., & Thorens, B. (2017). Plasma Dihydroceramides Are Diabetes Susceptibility Biomarker Candidates in Mice and Humans. Cell Reports, 18(9), 2269–2279. https://doi.org/10.1016/j.celrep.2017.02.019
Wu, 2017	Prospective cohort, case-cohort, nested case-control	20 studies	4,347	3.4-21.4 years	GC, GLC	Linoleic acid and arachidonic acid	Wu, J. H. Y., Marklund, M., Imamura, F., Tintle, N., Ardisson Korat, A. V, de Goede, J., Zhou, X., Yang, W.-S., de Oliveira Otto, M. C., Kröger, J., Qureshi, W., Virtanen, J. K., Bassett, J. K., Frazier-Wood, A. C., Lankinen, M., Murphy, R. A., Rajaobelina, K., Del Gobbo, L. C., Forouhi, N. G., … Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Fatty Acids and Outcomes Research Consortium (FORCE). (2017). Omega-6 fatty acid biomarkers and incident type 2 diabetes: pooled analysis of individual-level data for 39 740 adults from 20 prospective cohort studies. The Lancet. Diabetes & Endocrinology, 5(12), 965–974. https://doi.org/10.1016/S2213-8587(17)30307-8
Yamakado, 2015	Prospective cohort	Japanese population-based cohort	174	4 years	HPLC-MS	Amino acids	Yamakado, M., Nagao, K., Imaizumi, A., Tani, M., Toda, A., Tanaka, T., Jinzu, H., Miyano, H., Yamamoto, H., Daimon, T., Horimoto, K., & Ishizaka, Y. (2015). Plasma Free Amino Acid Profiles Predict Four-Year Risk of Developing Diabetes, Metabolic Syndrome, Dyslipidemia and Hypertension in Japanese Population. Scientific Reports, 5(1), 11918. https://doi.org/10.1038/srep11918
Yengo, 2016	Case-cohort	D.E.S.I.R.	231	9 years	DiscoveryHD4 platform Metabolon	Untargeted approach	Yengo, L., Arredouani, A., Marre, M., Roussel, R., Vaxillaire, M., Falchi, M., Haoudi, A., Tichet, J., Balkau, B., Bonnefond, A., & Froguel, P. (2016). Impact of statistical models on the prediction of type 2 diabetes using non-targeted metabolomics profiling. Molecular Metabolism, 5(10), 918–925. https://doi.org/10.1016/j.molmet.2016.08.011

Publication	Study design	Study population	Incident T2D cases	Follow-up	Measurement technique or platform	Measured metabolite classes	References
Yu, 2016	Prospective cohort	Shanghai Women's and Men's Health studies	73	not reported	DiscoveryHD4 platform Metabolon	Untargeted approach	Yu, D., Moore, S. C., Matthews, C. E., Xiang, Y.-B., Zhang, X., Gao, Y.-T., Zheng, W., & Shu, X.-O. (2016). Plasma metabolomic profiles in association with type 2 diabetes risk and prevalence in Chinese adults. Metabolomics, 12(1), 3. https://doi.org/10.1007/s11306-015-0890-8
Yu, 2022	Nested case-control	MACS/WIHS Combined Cohort Study	50	8 years	LC-MS	Untargeted approach	Yu, E. A., Aleman, J. O., Hoover, D. R., Shi, Q., Verano, M., Anastos, K., Tien, P. C., Sharma, A., Kardashian, A., Cohen, M. H., Golub, E. T., Michel, K. G., Gustafson, D. R., & Glesby, M. J. (2022). Plasma metabolomic analysis indicates flavonoids and sorbic acid are associated with incident diabetes: A nested case-control study among Women's Interagency HIV Study participants. PLOS ONE, 17(7), e0271207. https://doi.org/10.1371/JOURNAL.PONE.0271207
Yun, 2020	prospective cohort	NHAPC	529	6 years	LC-MS	78 sphingolipids	Yun, H., Sun, L., Wu, Q., Zong, G., Qi, Q., Li, H., Zheng, H., Zeng, R., Liang, L., & Lin, X. (2020). Associations among circulating sphingolipids, β-cell function, and risk of developing type 2 diabetes: A population-based cohort study in China. PLoS Medicine, 17(12). https://doi.org/10.1371/journal.pmed.1003451
Zhao, 2015	Case-cohort	Strong Heart Family Study	133	5.5 years	LC-MS	Untargeted approach	Zhao, J., Zhu, Y., Hyun, N., Zeng, D., Uppal, K., Tran, V. T., Yu, T., Jones, D., He, J., Lee, E. T., & Howard, B. V. (2015). Novel Metabolic Markers for the Risk of Diabetes Development in American Indians. Diabetes Care, 38(2), 220–227. https://doi.org/10.2337/dc14-2033
Zheng, 2018	Case-cohort study	EPIC-InterAct	9,671	Various	LC-MS/MS	Plasma 25(OH)D metabolites (vitamin D metabolites): total 25(OH)D, 25(OH)D3, 25(OH)D2,, 3-epi-25(OH)D3	Zheng, J.-S., Imamura, F., Sharp, S. J., van der Schouw, Y. T., Sluijs, I., Gundersen, T. E., Ardanaz, E., Boeing, H., Bonet, C., Gómez, J. H., Dow, C., Fagherazzi, G., Franks, P. W., Jenab, M., Kühn, T., Kaaks, R., Key, T. J., Khaw, K.-T., Lasheras, C., … Wareham, N. J. (2018). Association of plasma vitamin D metabolites with incident type 2 diabetes: EPIC-InterAct case-cohort study. The Journal of Clinical Endocrinology & Metabolism. https://doi.org/10.1210/jc.2018-01522

Supplementary Table S4.3 - SNPs and their weighting in the 236 SNP T2D PGS

SNPs included in the PGS are the primary association signals from Mahajan et al (2018)[72]. Beta estimates from BMI adjusted and unadjusted analyses were used to weight alleles in the PGS

T2D_locus	rsid	chr	pos_b37	EA	OA	T2D_beta BMI_Unadjusted	T2D_beta BMI_Adjusted	SNP_in_EPICN	flag	Included in PGS
MACF1	rs3768321	1	40,035,928	T	G	0.037	0.033	Y	NA	Y
FAF1	rs58432198	1	51,256,091	C	T	0.029	0.033	Y	NA	Y
PATJ	rs12140153	1	62,579,891	G	T	0.033	0.017	Y	NA	Y
DENND2C	rs184660829	1	115,144,899	C	T	0.9	0.81	Y	NA	Y
PTGFRN	rs1127215	1	117,532,790	C	T	0.021	0.021	Y	NA	Y
NOTCH2	rs1493694	1	120,526,982	T	C	0.037	0.041	Y	NA	Y
FAM63A	rs145904381	1	151,017,991	T	C	0.079	0.079	Y	NA	Y
SEC16B	rs539515	1	177,889,025	C	A	0.025	0.0086	Y	NA	Y
DSTYK	rs12048743	1	205,114,873	G	C	0.017	0.017	Y	NA	Y
SRGAP2	rs9430095	1	206,593,900	C	G	0.013	0.013	Y	NA	Y
PROX1	rs340874	1	214,159,256	C	T	0.029	0.029	Y	NA	Y
LYPLAL1	rs2820446	1	219,748,818	C	G	0.021	0.025	Y	NA	Y
ABCB10	rs348330	1	229,672,955	G	A	0.021	0.021	Y	NA	Y
GNG4	rs291367	1	235,690,800	G	A	0.017	0.017	Y	NA	Y
TMEM18	rs62107261	2	422,144	T	C	0.045	0.013	Y	NA	Y
FAM49A	rs11680058	2	16,574,669	A	G	0.025	0.025	Y	NA	Y
DTNB	rs17802463	2	25,643,221	G	T	0.017	0.017	Y	NA	Y
GCKR	rs1260326	2	27,730,940	C	T	0.033	0.029	Y	pleiotropic	N
THADA	rs80147536	2	43,698,028	A	T	0.057	0.061	Y	NA	Y
BNIPL	rs6545714	2	59,307,725	G	A	0.017	0.0086	Y	NA	Y
BCL11A	rs243024	2	60,583,665	A	G	0.025	0.025	Y	NA	Y
CEP68	rs2249105	2	65,287,896	A	G	0.021	0.021	Y	NA	Y
TMEM127	rs79046683	2	96,913,918	T	G	0.33	0.37	N	NA	N
DDX18	rs562386202	2	118,071,061	G	A	0.51	0.52	N	NA	N
GLI2	rs11688682	2	121,347,612	G	C	0.025	0.029	Y	NA	Y
PABPC1P2	rs35999103	2	147,861,633	T	C	0.025	0.021	Y	NA	Y
CYTIP	rs13426680	2	158,339,550	A	G	0.033	0.041	Y	NA	Y
RBMS1	rs3772071	2	161,135,544	T	C	0.017	0.013	Y	NA	Y
GRB14/COBLL1	rs10195252	2	165,513,091	T	C	0.029	0.033	Y	NA	Y
CRYBA2	rs113414093	2	219,859,171	A	G	0.041	0.049	Y	NA	Y
IRS1	rs2972144	2	227,101,411	G	A	0.041	0.045	Y	NA	Y
PPARG	rs11709077	3	12,336,507	G	A	0.041	0.053	Y	NA	Y
UBE2E2	rs35352848	3	23,455,582	T	C	0.037	0.037	Y	NA	Y
KIF9	rs11926707	3	46,925,539	C	T	0.017	0.021	Y	NA	Y
RBM6	rs4688760	3	49,980,596	T	C	0.021	0.013	Y	NA	Y
RFT1	rs2581787	3	53,127,677	T	G	0.017	0.013	Y	NA	Y
CACNA2D3	rs76263492	3	54,828,827	T	G	0.041	0.045	Y	NA	Y
PSMD6	rs3774723	3	63,962,339	G	A	0.029	0.029	Y	NA	Y
ADAMTS9	rs9860730	3	64,701,146	A	G	0.025	0.029	Y	NA	Y
SHQ1	rs13085136	3	72,865,183	C	T	0.037	0.037	Y	NA	Y
ROBO2	rs2272163	3	77,671,721	C	A	0.013	0.013	Y	NA	Y
ADCY5	rs11708067	3	123,065,778	A	G	0.041	0.045	Y	NA	Y
SLC12A8	rs649961	3	124,926,637	T	C	0.017	0.017	Y	NA	Y
TMCC1	rs9828772	3	129,333,182	C	G	0.025	0.025	Y	NA	Y
TSC22D2	rs62271373	3	150,066,540	A	T	0.037	0.041	Y	NA	Y
MBNL1	rs13065698	3	152,086,533	A	G	0.017	0.013	N	NA	N
EGFEM1P	rs7629630	3	168,218,841	A	T	0.021	0.021	Y	NA	Y
SLC2A2	rs9873618	3	170,733,076	G	A	0.029	0.033	Y	NA	Y
ABCC5	rs2872246	3	183,738,460	A	C	0.017	0.013	Y	NA	Y
IGF2BP2	rs6780171	3	185,503,456	A	T	0.049	0.053	Y	NA	Y
ST6GAL1	rs3887925	3	186,665,645	T	C	0.025	0.025	Y	NA	Y
LPP	rs4686471	3	187,740,899	C	T	0.029	0.029	Y	NA	Y
PCGF3	rs1531583	4	744,972	T	G	0.053	0.057	Y	NA	Y
MAEA	rs56337234	4	1,784,403	C	T	0.025	0.033	Y	NA	Y
HTT	rs362307	4	3,241,845	T	C	0.029	0.021	Y	NA	Y
WFS1	rs10937721	4	6,306,763	C	G	0.037	0.037	Y	NA	Y
LCORL	rs12640250	4	17,792,869	C	A	0.013	0.013	Y	NA	Y
GNPDA2	rs10938398	4	45,186,139	A	G	0.017	0.0043	Y	NA	Y
USP46	rs2102278	4	52,818,664	G	A	0.017	0.013	Y	NA	Y
SCD5	rs12642790	4	83,578,271	A	G	0.021	0.021	N	NA	N
FAM13A	rs1903002	4	89,740,894	G	C	0.013	0.017	Y	NA	Y
SMARCAD1	rs6821438	4	95,091,911	A	G	0.013	0.013	Y	NA	Y
SLC9B1	rs1580278	4	104,140,848	C	A	0.017	0.017	Y	NA	Y

T2D_locus	rsid	chr	pos_b37	EA	OA	T2D_beta BMI_Unadjusted	T2D_beta BMI_Adjusted	SNP_in_EPICN	flag	Included in PGS
PABPC4L	rs1296328	4	137,083,193	A	C	0.017	0.0086	Y	NA	Y
TMEM154	rs7669833	4	153,513,369	T	A	0.021	0.025	Y	NA	Y
PDGFC	rs28819812	4	157,652,753	C	A	0.017	0.017	Y	NA	Y
ACSL1	rs58730668	4	185,717,759	T	C	0.029	0.029	Y	NA	Y
ANKH	rs146886108	5	14,751,305	C	T	0.17	0.17	Y	NA	Y
MRPS30	rs6884702	5	44,682,589	G	A	0.017	0.017	Y	NA	Y
ITGA1	rs3811978	5	52,100,489	G	A	0.021	0.025	Y	NA	Y
ARL15	rs702634	5	53,271,420	A	G	0.021	0.025	Y	NA	Y
ANKRD55	rs465002	5	55,808,475	T	C	0.033	0.033	Y	NA	Y
PIK3R1	rs4976033	5	67,714,246	G	A	0.017	0.021	Y	NA	Y
POC5	rs2307111	5	75,003,678	T	C	0.021	0.013	Y	NA	Y
ZBED3	rs4457053	5	76,424,949	G	A	0.025	0.029	Y	NA	Y
DMGDH	rs1316776	5	78,430,607	C	A	0.021	0.017	Y	NA	Y
RASA1	rs7719891	5	86,577,352	G	A	0.017	0.013	Y	NA	Y
SLCO6A1	rs138337556	5	101,232,944	G	A	0.22	0.23	Y	NA	Y
PAM	rs115505614	5	102,422,968	T	C	0.076	0.083	Y	NA	Y
PHF15	rs329122	5	133,864,599	A	G	0.017	0.025	Y	NA	Y
EBF1	rs3934712	5	157,928,196	C	T	0.021	0.021	Y	NA	Y
RREB1	rs9379084	6	7,231,843	G	A	0.049	0.053	Y	NA	Y
CDKAL1	rs7756992	6	20,679,709	G	A	0.061	0.064	Y	NA	Y
MHC	rs601945	6	32,573,415	G	A	0.033	0.037	Y	NA	Y
HMGA1	rs77136196	6	34,247,047	T	C	0.037	0.049	Y	NA	Y
LRFN2	rs34298980	6	40,409,243	T	C	0.017	0.0086	Y	NA	Y
VEGFA	rs6458354	6	43,814,190	C	T	0.021	0.021	Y	NA	Y
TFAP2B	rs3798519	6	50,788,778	C	A	0.029	0.017	Y	NA	Y
SLC25A51P1	rs555402748	6	67,387,490	T	C	0.56	0.62	Y	NA	Y
BEND3	rs4946812	6	107,431,688	G	A	0.017	0.017	Y	NA	Y
CENPW	rs11759026	6	126,792,095	G	A	0.029	0.033	Y	NA	Y
SOGA3	rs2800733	6	127,416,930	A	G	0.025	0.029	Y	NA	Y
SLC35D3	rs9494624	6	137,300,960	A	G	0.017	0.021	Y	NA	Y
MIR3668	rs2982521	6	139,835,329	A	T	0.017	0.021	Y	NA	Y
SLC22A3	rs474513	6	160,770,312	A	G	0.017	0.021	Y	NA	Y
QKI	rs4709746	6	164,133,001	C	T	0.029	0.033	Y	NA	Y
DGKB	rs10228066	7	15,063,569	T	C	0.029	0.029	Y	NA	Y
IGF2BP3	rs4279506	7	23,512,896	G	C	0.017	0.017	Y	NA	Y
JAZF1	rs1708302	7	28,198,677	C	T	0.037	0.045	Y	NA	Y
CRHR2	rs917195	7	30,728,452	C	T	0.021	0.017	Y	NA	Y
GCK	rs878521	7	44,255,643	A	G	0.025	0.025	Y	NA	Y
FBXL13	rs11496066	7	102,486,254	T	C	0.021	0.021	Y	NA	Y
RELN	rs39328	7	103,444,978	T	C	0.013	0.0086	Y	NA	Y
CTTNBP2	rs6976111	7	117,495,667	A	C	0.017	0.017	Y	NA	Y
KLF14	rs1562396	7	130,457,914	G	A	0.025	0.029	Y	NA	Y
AOC1	rs62492368	7	150,537,635	A	G	0.021	0.021	Y	NA	Y
MNX1	rs6459733	7	156,930,550	G	C	0.029	0.025	Y	NA	Y
MSRA	rs17689007	8	9,974,824	G	A	0.021	0.017	Y	NA	Y
XKR6	rs57327348	8	10,808,687	A	T	0.021	0.013	Y	NA	Y
LPL	rs10096633	8	19,830,921	C	T	0.029	0.033	Y	NA	Y
PURG	rs10954772	8	30,863,938	T	C	0.017	0.013	Y	NA	Y
ANK1	rs13262861	8	41,508,577	C	A	0.045	0.045	Y	NA	Y
TP53INP1	rs10097617	8	95,961,626	T	C	0.021	0.021	Y	NA	Y
CPQ	rs149364428	8	97,737,741	A	G	0.12	0.11	Y	NA	Y
TRHR	rs12680028	8	110,123,183	C	G	0.017	0.013	Y	NA	Y
SLC30A8	rs3802177	8	118,185,025	G	A	0.049	0.053	Y	NA	Y
CASC11	rs17772814	8	128,711,742	G	A	0.029	0.037	Y	NA	Y
PVT1	rs1561927	8	129,568,078	C	T	0.017	0.017	Y	NA	Y
BOP1	rs4977213	8	145,507,304	C	T	0.021	0.025	Y	NA	Y
GLIS3	rs10974438	9	4,291,928	C	A	0.025	0.025	Y	NA	Y
HAUS6	rs7022807	9	19,067,833	G	A	0.017	0.021	Y	NA	Y
FOCAD	rs7867635	9	20,241,069	C	T	0.017	0.013	Y	NA	Y
CDKN2A/B	rs10811660	9	22,134,068	G	A	0.072	0.076	Y	NA	Y
LINGO2	rs1412234	9	28,410,683	C	T	0.021	0.0086	Y	NA	Y
UBAP2	rs12001437	9	34,074,476	C	T	0.017	0.013	Y	NA	Y
MTND2P8	rs11137820	9	81,359,113	C	G	0.013	0.013	Y	NA	Y
TLE4	rs17791513	9	81,905,590	A	G	0.045	0.045	Y	NA	Y
TLE1	rs2796441	9	84,308,948	G	A	0.029	0.033	Y	NA	Y
ZNF169	rs55653563	9	97,001,682	A	C	0.021	0.017	Y	NA	Y
ABO	rs505922	9	136,149,229	C	T	0.017	0.017	Y	NA	Y
GPSM1	rs28505901	9	139,241,030	G	A	0.033	0.033	Y	NA	Y

T2D_locus	rsid	chr	pos_b37	EA	OA	T2D_beta BMI_Unadjusted	T2D_beta BMI_Adjusted	SNP_in_EPICN	flag	Included in PGS
CDC123/CAMK1D	rs11257655	10	12,307,894	T	C	0.037	0.041	Y	NA	Y
NEUROG3	rs2642588	10	71,466,578	G	T	0.021	0.021	Y	NA	Y
ZMIZ1	rs703972	10	80,952,826	G	C	0.033	0.033	Y	NA	Y
PTEN	rs11202627	10	89,769,340	T	C	0.025	0.025	Y	NA	Y
HHEX/IDE	rs10882101	10	94,462,427	T	C	0.049	0.053	Y	NA	Y
TCF7L2	rs7903146	10	114,758,349	T	C	0.14	0.16	N	NA	N
WDR11	rs72631105	10	122,915,345	A	G	0.021	0.025	Y	NA	Y
PLEKHA1	rs2280141	10	124,193,181	T	G	0.017	0.021	Y	NA	Y
INS/IGF2	rs4929965	11	2,197,286	A	G	0.033	0.029	Y	NA	Y
KCNQ1	rs2237895	11	2,857,194	C	A	0.045	0.049	Y	NA	Y
PDE3B	rs141521721	11	14,763,828	A	C	0.057	0.057	Y	NA	Y
KCNJ11	rs5213	11	17,408,404	C	T	0.029	0.033	Y	NA	Y
METTL15	rs4923543	11	28,534,898	A	G	0.017	0.017	Y	NA	Y
QSER1	rs145678014	11	32,927,778	G	T	0.045	0.041	Y	NA	Y
PDHX	rs2767036	11	34,982,148	C	A	0.017	0.017	Y	NA	Y
HSD17B12	rs1061810	11	43,877,934	A	C	0.021	0.013	Y	NA	Y
CRY2	rs7115753	11	45,912,013	A	G	0.017	0.017	Y	NA	Y
CELF1	rs7124681	11	47,529,947	A	C	0.017	0.0086	Y	NA	Y
MAP3K11	rs1783541	11	65,294,799	T	C	0.025	0.025	Y	NA	Y
CCND1	rs11820019	11	69,448,758	T	C	0.061	0.061	Y	NA	Y
CENTD2/ARAP1	rs77464186	11	72,460,398	A	C	0.045	0.057	Y	NA	Y
MTNR1B	rs10830963	11	92,708,710	G	C	0.045	0.045	Y	NA	Y
ETS1	rs67232546	11	128,398,938	T	C	0.025	0.025	Y	NA	Y
CCND2	rs76895963	12	4,384,844	T	G	0.2	0.23	Y	NA	Y
CDKN1B	rs2066827	12	12,871,099	G	T	0.021	0.021	Y	NA	Y
ITPR2	rs718314	12	26,453,283	G	A	0.021	0.025	Y	NA	Y
KLHDC5	rs10842994	12	27,965,150	C	T	0.033	0.033	Y	NA	Y
HMGA2	rs2258238	12	66,221,060	T	A	0.053	0.049	Y	NA	Y
TSPAN8/LGR5	rs1796330	12	71,522,953	G	C	0.021	0.021	Y	NA	Y
USP44	rs2197973	12	95,928,560	T	C	0.017	0.013	Y	NA	Y
RMST	rs77864822	12	97,848,775	A	G	0.029	0.033	Y	NA	Y
WSCD2	rs1426371	12	108,629,780	G	A	0.025	0.021	Y	NA	Y
KSR2	rs34965774	12	118,412,373	A	G	0.021	0.021	Y	NA	Y
HNF1A	rs56348580	12	121,432,117	G	C	0.025	0.021	Y	NA	Y
MPHOSPH9	rs4148856	12	123,450,765	C	G	0.021	0.025	Y	NA	Y
FBRSL1	rs12811407	12	133,069,698	A	G	0.021	0.017	Y	NA	Y
RNF6	rs34584161	13	26,776,999	A	G	0.021	0.021	Y	NA	Y
HMGB1	rs11842871	13	31,042,452	G	T	0.021	0.017	Y	NA	Y
KL	rs576674	13	33,554,302	G	A	0.021	0.021	Y	NA	Y
DLEU1	rs963740	13	51,096,095	A	T	0.017	0.021	Y	NA	Y
PCDH17	rs9537803	13	58,366,634	C	T	0.017	0.013	Y	NA	Y
SRGAP2D	rs9563615	13	59,077,406	A	T	0.021	0.017	Y	NA	Y
SPRY2	rs1359790	13	80,717,156	G	A	0.037	0.041	Y	NA	Y
IRS2	rs7987740	13	109,947,213	T	C	0.017	0.017	Y	NA	Y
SLC7A7	rs17122772	14	23,288,935	G	C	0.021	0.021	Y	NA	Y
AKAP6	rs17522122	14	33,302,882	T	G	0.017	0.0086	Y	NA	Y
CLEC14A	rs8017808	14	38,848,419	G	T	0.017	0.021	Y	NA	Y
NRXN3	rs17836088	14	79,932,041	C	G	0.029	0.017	Y	NA	Y
SMEK1	rs8010382	14	91,963,722	G	A	0.017	0.017	Y	NA	Y
MARK3	rs62007683	14	103,894,071	G	T	0.017	0.013	Y	NA	Y
RASGRP1	rs34715063	15	38,873,115	C	T	0.037	0.037	Y	NA	Y
LTK	rs11070332	15	41,809,205	A	G	0.021	0.021	Y	NA	Y
ONECUT1	rs2456530	15	53,091,553	T	C	0.029	0.025	Y	NA	Y
WDR72	rs528350911	15	53,747,228	G	C	0.097	0.12	Y	NA	Y
TCF12	rs117483894	15	57,456,802	G	A	0.041	0.037	Y	NA	Y
C2CD4A/B	rs8037894	15	62,394,264	G	C	0.021	0.021	Y	NA	Y
USP3	rs7178762	15	63,871,292	C	T	0.017	0.017	Y	NA	Y
MAP2K5	rs4776970	15	68,080,886	A	T	0.017	0.0043	Y	NA	Y
PTPN9	rs13737	15	75,932,129	G	T	0.021	0.021	Y	NA	Y
HMG20A	rs1005752	15	77,818,128	A	C	0.033	0.029	Y	NA	Y
AP3S2	rs4932265	15	90,423,293	T	C	0.029	0.033	Y	NA	Y
PRC1	rs12910825	15	91,511,260	G	A	0.025	0.025	Y	NA	Y
ITFG3	rs6600191	16	295,795	T	C	0.029	0.029	Y	NA	Y
CLUAP1	rs3751837	16	3,583,173	T	C	0.017	0.013	Y	NA	Y
ATP2A1	rs8046545	16	28,915,217	G	A	0.017	0.0043	Y	NA	Y
FAM57B	rs11642430	16	30,045,789	G	C	0.017	0.0086	Y	NA	Y
FTO	rs1421085	16	53,800,954	C	T	0.053	0.025	Y	NA	Y

T2D_locus	rsid	chr	pos_b37	EA	OA	T2D_beta BMI_Unadjusted	T2D_beta BMI_Adjusted	SNP_in_EPICN	flag	Included in PGS
NFAT5	rs862320	16	69,651,866	C	T	0.021	0.017	Y	NA	Y
BCAR1	rs72802342	16	75,234,872	C	A	0.061	0.068	Y	NA	Y
CMIP	rs2925979	16	81,534,790	T	C	0.017	0.021	Y	NA	Y
SPG7	rs12920022	16	89,564,055	A	T	0.021	0.021	Y	NA	Y
ZZEF1	rs1377807	17	4,045,440	C	G	0.025	0.021	Y	NA	Y
ATP1B2	rs1641523	17	7,549,681	C	T	0.017	0.017	Y	NA	Y
GLP2R	rs7222481	17	9,785,187	C	G	0.021	0.017	Y	NA	Y
RAI1	rs4925109	17	17,661,802	A	G	0.017	0.021	Y	NA	Y
NF1	rs71372253	17	29,413,019	C	T	0.033	0.033	Y	NA	Y
HNF1B	rs10908278	17	36,099,952	T	A	0.033	0.037	Y	NA	Y
MLX	rs34855406	17	40,731,411	C	G	0.021	0.021	Y	NA	Y
TTLL6	rs35895680	17	47,060,322	C	A	0.025	0.017	Y	NA	Y
KIF2B	rs569511541	17	52,140,805	G	A	0.99	0.98	Y	NA	Y
ACE	rs60276348	17	62,203,304	T	C	0.021	0.021	Y	NA	Y
BPTF	rs61676547	17	65,892,507	C	G	0.021	0.017	Y	NA	Y
LAMA1	rs7240767	18	7,070,642	C	T	0.013	0.017	Y	NA	Y
COMMD9	rs62080313	18	36,278,709	C	T	0.025	0.021	Y	NA	Y
TCF4	rs72926932	18	53,050,646	C	A	0.037	0.037	Y	NA	Y
WDR7	rs17684074	18	54,675,384	G	C	0.017	0.017	Y	NA	Y
GRP	rs9957145	18	56,876,228	G	A	0.021	0.013	Y	NA	Y
MC4R	rs523288	18	57,848,369	T	A	0.025	0.0086	Y	NA	Y
BCL2A	rs12454712	18	60,845,884	T	C	0.025	0.029	Y	NA	Y
UHRF1	rs7249758	19	4,948,862	A	G	0.021	0.021	Y	NA	Y
INSR	rs75253922	19	7,240,848	C	T	0.021	0.025	Y	NA	Y
MAP2K7	rs4804833	19	7,970,635	A	G	0.025	0.025	Y	NA	Y
FARSA	rs3111316	19	13,038,415	A	G	0.021	0.021	Y	NA	Y
TM6SF2	rs8107974	19	19,388,500	T	A	0.045	0.053	Y	NA	Y
PEPD	rs10406327	19	33,890,838	C	G	0.017	0.021	Y	NA	Y
TOMM40/APOE	rs429358	19	45,411,941	T	C	0.029	0.021	Y	NA	Y
GIPR	rs10406431	19	46,157,019	A	G	0.025	0.025	Y	NA	Y
ZC3H4	rs3810291	19	47,569,003	A	G	0.025	0.013	Y	NA	Y
NKX2.2	rs13041756	20	21,466,795	C	T	0.021	0.013	Y	NA	Y
RALY	rs2268078	20	32,596,704	A	G	0.017	0.013	Y	NA	Y
HNF4A	rs1800961	20	43,042,364	T	C	0.072	0.079	Y	NA	Y
EYA2	rs6063048	20	45,598,564	G	A	0.021	0.021	Y	NA	Y
CEBPB	rs11699802	20	48,832,135	C	T	0.021	0.025	Y	NA	Y
TSHZ2	rs34454109	20	51,223,594	A	T	0.021	0.013	Y	NA	Y
GNAS	rs6070625	20	57,394,628	G	C	0.021	0.021	Y	NA	Y
TCEA2	rs59944054	20	62,693,175	A	G	0.017	0.021	Y	NA	Y
MTMR3/ASCC2	rs6518681	22	30,609,554	G	A	0.037	0.037	Y	NA	Y
YWHAH	rs117001013	22	32,348,841	C	T	0.033	0.033	Y	NA	Y
EP300	rs5758223	22	41,489,920	A	G	0.017	0.013	Y	NA	Y
PNPLA3	rs738408	22	44,324,730	T	C	0.021	0.025	Y	NA	Y
PIM3	rs1801645	22	50,356,850	C	T	0.025	0.025	Y	NA	Y

Supplementary Table S4.4: SNPs and corresponding weights used in the partial PGS from Udler et al (2018)

ea-effect allele, oa – other allele, EPICN – EPIC-Norfolk

						T2D Morris 2012		Partial PGS weight							
rsid	**locus**	**chr**	**pos**	**ea**	**oa**	**T2D eaf**	**T2D pval**	**beta_ cell**	**proinsuli n**	**obesit y**	**lipodystr ophy**	**liver**	**SNP_in _EPICN**	**flag**	**includ ed_ in_P GS**
rs10146997	NRXN3	14	79945162	G	A	0.208	7.20E-03	NA	NA	1.715	NA	NA	Y	NA	Y
rs10203174	THADA	2	43690030	C	T	0.917	1.50E-06	NA	NA	NA	NA	NA	Y	NA	Y
rs10219509	CCND2	12	4300809	C	T	0.191	1.40E-02	NA	NA	NA	NA	NA	Y	NA	Y
rs10276674	DGKB	7	14922007	C	T	0.222	6.60E-07	NA	1.001	NA	NA	NA	Y	NA	Y
rs10503669	LPL	8	19847690	C	A	0.878	5.20E-03	NA	NA	NA	1.536	NA	Y	NA	Y
rs10757282	CDKN2A	9	22133984	C	T	0.442	8.80E-04	NA	NA	NA	NA	NA	Y	NA	Y
rs10758593	GLIS3	9	4292083	A	G	0.433	7.20E-04	1.094	NA	NA	NA	NA	Y	NA	Y
rs10811661	CDKN2A_B	9	22134094	T	C	0.792	1.50E-13	1.766	NA	NA	NA	NA	Y	NA	Y
rs10830961	MTNR1B	11	92694757	G	A	0.383	4.80E-09	3.806	NA	NA	NA	NA	Y	NA	Y
rs10840595	INS-IGF2	11	2231273	C	T	0.833	2.80E-03	NA	NA	NA	NA	NA	Y	NA	Y
rs10842994	KLHDC5	12	27965150	C	T	0.867	1.50E-04	NA	NA	NA	NA	NA	Y	NA	Y
rs10923931	NOTCH2	1	1.21E+08	T	G	0.117	3.10E-04	NA	NA	NA	NA	NA	Y	NA	Y
rs1106240	MPHOSPH9	12	1.24E+08	T	C	0.800	6.50E-04	NA	NA	NA	0.786	NA	Y	NA	Y
rs11063069	CCND2	12	4374373	G	A	0.209	1.50E-04	NA	0.812	NA	0.938	NA	Y	NA	Y
rs1111875	HHEX	10	94462882	C	T	0.558	1.10E-15	2.656	NA	NA	NA	NA	Y	NA	Y
rs11227234	MAP3K11	11	65365171	T	G	0.240	5.20E-04	NA	NA	NA	NA	NA	Y	NA	Y
rs11257655	CDC123/CAMK1D	10	12307894	T	C	0.275	1.30E-02	1.587	0.780	NA	NA	NA	Y	NA	Y
rs11708067	ADCY5	3	1.23E+08	A	G	0.783	4.30E-06	0.919	NA	NA	1.242	NA	Y	NA	Y
rs1182397	MNX1	7	1.57E+08	G	T	0.824	1.70E-03	NA	NA	NA	NA	NA	Y	NA	Y
rs12571751	ZMIZ1	10	80942631	A	G	0.508	1.80E-07	NA	NA	NA	NA	NA	Y	NA	Y
rs12899811	PRC1	15	91544076	G	A	0.233	3.30E-05	1.262	NA	NA	NA	NA	Y	NA	Y
rs12970134	MC4R	18	57884750	A	G	0.308	1.10E-04	NA	NA	2.783	NA	NA	Y	NA	Y
rs13389219	GRB14	2	1.66E+08	C	T	0.558	9.60E-03	NA	NA	NA	2.738	NA	Y	NA	Y
rs1359790	SPRY2	13	80717156	G	A	0.725	9.20E-07	NA	1.073	NA	NA	NA	Y	NA	Y
rs1371135	AP3S2	15	90391270	T	C	0.317	2.70E-02	1.542	NA	NA	NA	NA	Y	NA	Y
rs1496653	UBE2E2	3	23454790	A	G	0.808	1.40E-04	1.350	NA	NA	NA	NA	Y	NA	Y
rs1552224	ARAP1	11	72433098	A	C	0.875	4.90E-07	NA	4.484	NA	NA	NA	Y	NA	Y
rs15563	GIP	17	47005193	G	A	0.530	4.80E-04	0.883	NA	NA	NA	NA	Y	NA	Y

rs16996148	CILP2	19	19658472	T	G	0.083	4.00E-04	NA	NA	NA	NA	2.054	Y	NA	Y
rs17106184	FAF1	1	50909985	G	A	0.892	1.10E-03	NA	NA	NA	0.812	NA	Y	NA	Y
rs17265513	ZHX3	20	39832628	C	T	0.199	1.80E-05	1.369	NA	NA	NA	NA	Y	NA	Y
rs17676067	GLP2R	17	9791375	C	T	0.472	5.90E-04	1.216	NA	NA	NA	NA	Y	NA	Y
rs17791513	TLE4	9	81905590	A	G	0.933	1.00E-07	0.776	NA	NA	NA	NA	Y	NA	Y
rs1800961	HNF4A	20	43042364	T	C	0.040	2.50E-02	NA	NA	NA	NA	NA	Y	NA	Y
rs1801282	PPARG	3	12393125	C	G,A	0.925	5.00E-09	NA	NA	NA	2.107	NA	Y	NA	Y
rs1996546	ACSL1	4	1.86E+08	G	T	0.854	1.10E-04	1.349	NA	NA	NA	NA	Y	NA	Y
rs2184061	CDKN2A	9	22061562	A	C	0.605	2.40E-05	0.971	NA	NA	NA	NA	Y	NA	Y
rs2191349	DGKB/TMEM195	7	15064309	T	G	0.467	2.20E-04	NA	NA	NA	NA	NA	Y	NA	Y
rs2237892	KCNQ1	11	2839751	C	T	0.925	4.30E-03	NA	NA	NA	NA	NA	Y	NA	Y
rs2261181	HMGA2	12	66212318	T	C	0.092	1.00E-07	NA	NA	NA	NA	NA	Y	NA	Y
rs2276853	KIF9	3	47282303	A	G	0.607	2.40E-03	NA	NA	NA	0.956	NA	Y	NA	Y
rs2296172	MACF1	1	39835817	G	A	0.183	1.30E-02	NA	NA	NA	1.031	NA	Y	NA	Y
rs231361	KCNQ1	11	2691500	A	G	0.272	2.80E-05	NA	NA	NA	NA	NA	Y	NA	Y
rs233449	KCNQ1	11	2843803	G	A	0.733	4.40E-04	1.184	NA	NA	NA	NA	Y	NA	Y
rs2334499	DUSP8	11	1696849	T	C	0.442	7.30E-04	0.959	NA	NA	NA	NA	Y	NA	Y
rs2421016	PLEKHA1	10	1.24E+08	C	T	0.508	4.70E-06	NA	NA	NA	NA	NA	Y	NA	Y
rs243021	BCL11A	2	60584819	A	G	0.458	5.30E-06	NA	NA	NA	NA	NA	Y	NA	Y
rs2796441	TLE1	9	84308948	G	A	0.583	2.50E-04	NA	NA	NA	NA	NA	Y	NA	Y
rs2820443	LYPLAL1	1	2.2E+08	T	C	0.680	4.00E-03	NA	NA	NA	1.919	NA	Y	NA	Y
rs2866367	ZFAND6	15	80365745		NA	0.665	6.10E-05	NA	NA	NA	NA	NA	Y	NA	Y
rs2925979	CMIP	16	81534790	T	C	0.294	2.30E-03	NA	NA	NA	1.744	NA	Y	NA	Y
rs2943641	IRS1	2	2.27E+08	C	T	0.608	5.50E-05	NA	NA	NA	3.008	NA	Y	NA	Y
rs3132524	POU5F1	6	31136714	C	T	0.700	4.40E-05	NA	NA	NA	1.056	NA	Y	NA	Y
rs340874	PROX1	1	2.14E+08	C	T	0.508	1.10E-04	NA	NA	NA	NA	NA	Y	NA	Y
rs36608	MTMR3	22	30342191	A	G	0.967	1.10E-04	NA	NA	NA	NA	NA	Y	NA	Y
rs3736505	HSD17B12	11	43876435	G	A	0.298	3.00E-04	0.781	NA	1.240	NA	NA	Y	NA	Y
rs3802177	SLC30A8	8	1.18E+08	G	A	0.750	2.10E-11	2.407	NA	NA	NA	NA	Y	NA	Y
rs384037	KCNQ1	11	2912553	A	G	0.568	1.10E-03	NA	NA	NA	NA	NA	Y	NA	Y
rs3843467	ANKRD55	5	55856375	T	G	0.189	2.50E-06	NA	NA	NA	1.309	NA	Y	NA	Y
rs3996352	KLF14	7	1.3E+08	A	G	0.550	2.00E-07	NA	NA	NA	1.628	NA	Y	NA	Y
rs4402960	IGF2BP2	3	1.86E+08	T	G	0.292	2.70E-11	NA	0.969	NA	NA	NA	Y	NA	Y
rs4407733	SLC35D3	6	1.37E+08	A	G	0.534	3.10E-04	NA	NA	NA	NA	NA	Y	NA	Y
rs4430796	HNF1B	17	36098040	G	A	0.533	2.40E-06	1.038	NA	NA	NA	NA	Y	NA	Y

rs4457053	ZBED3	5	76424949	G	A	0.275	4.80E-07	NA	NA	NA	0.964	NA	Y	NA	Y
rs4458523	WFS1	4	6289986	G	T	0.733	1.90E-07	NA	NA	NA	NA	NA	Y	NA	Y
rs4502156	C2CD4A	15	62383155	T	C	0.517	1.10E-03	2.682	NA	NA	NA	NA	Y	NA	Y
rs459193	ANKRD55	5	55806751	G	A	0.783	2.10E-02	NA	NA	NA	1.799	NA	Y	NA	Y
rs4746890	NEUROG3	10	71470390	T	C	0.776	3.60E-02	NA	NA	NA	NA	NA	Y	NA	Y
rs4812829	HNF4A	20	42989267	A	G	0.175	9.10E-03	0.844	0.793	NA	NA	NA	Y	NA	Y
rs505922	ABO	9	1.36E+08	C	T	0.318	2.40E-02	1.169	NA	NA	NA	NA	Y	NA	Y
rs516946	ANK1	8	41519248	C	T	0.817	7.30E-07	1.121	NA	NA	NA	NA	Y	NA	Y
rs5215	KCNJ11	11	17408630	C	T	0.408	4.40E-06	1.191	NA	NA	NA	NA	Y	NA	Y
rs6742799	RBMS1	2	1.61E+08	A	C	0.825	2.20E-06	NA	NA	1.036	NA	NA	Y	NA	Y
rs6795735	ADAMTS9	3	64705365	C	T	0.542	2.30E-04	NA	NA	NA	0.949	NA	Y	NA	Y
rs6808574	LPP	3	1.88E+08	C	T	0.592	1.30E-04	NA	NA	NA	NA	NA	Y	NA	Y
rs6813195	TMEM154	4	1.54E+08	C	T	0.742	6.10E-05	NA	NA	NA	NA	NA	Y	NA	Y
rs702634	ARL15	5	53271420	A	G	0.725	1.80E-05	NA	NA	NA	1.286	NA	Y	NA	Y
rs7178572	HMG20A	15	77747190	G	A	0.692	1.00E-04	NA	NA	NA	NA	NA	Y	NA	Y
rs738409	PNPLA3	22	44324727	G	C	0.233	5.00E-03	NA	NA	NA	NA	0.841	Y	NA	Y
rs7756992	CDKAL1	6	20679709	G	A	0.250	1.30E-22	3.049	NA	NA	NA	NA	Y	NA	Y
rs780094	GCKR	2	27741237	C	T	0.617	2.50E-02	NA	NA	NA	NA	5.340	Y	pleiotropic	N
rs7845219	TP53INP1	8	95937502	T	C	0.517	1.40E-05	NA	NA	NA	NA	NA	Y	NA	Y
rs7903146	TCF7L2	10	1.15E+08	T	C	0.250	5.50E-65	2.468	NA	NA	NA	NA	Y	NA	Y
rs7957197	HNF1A	12	1.21E+08	T	A	0.850	1.10E-06	1.555	NA	NA	NA	NA	Y	NA	Y
rs7961581	TSPAN8/LGR5	12	71663102	C	T	0.233	3.10E-05	NA	NA	NA	NA	0.832	Y	NA	Y
rs8068804	ZZEF1	17	3985864	A	G	0.319	4.90E-03	NA	NA	NA	NA	NA	Y	NA	Y
rs8108269	GIPR	19	46158513	G	T	0.283	3.10E-03	1.255	NA	NA	NA	NA	Y	NA	Y
rs864745	JAZF1	7	28180556	T	C	0.517	7.30E-11	NA	NA	NA	NA	NA	Y	NA	Y
rs9271775	HLA-DQA1	6	32594328	T	C	0.761	2.80E-04	NA	NA	NA	NA	0.976	Y	NA	Y
rs9502570	RREB1	6	7258617	C	T	0.717	6.10E-04	NA	NA	NA	NA	NA	Y	NA	Y
rs9891146	C17orf58	17	65988049	T	C	0.257	1.30E-02	NA	NA	NA	1.078	NA	Y	NA	Y
rs9921586	CTRB2	16	75245003	G	T	0.858	2.10E-05	1.005	NA	NA	NA	NA	Y	NA	Y
rs9939609	FTO	16	53820527	A	T	0.450	2.00E-09	NA	NA	4.963	NA	NA	Y	NA	Y
rs9948462	LAMA1	18	7076836	T	C	0.325	1.50E-03	NA	NA	NA	NA	NA	Y	NA	Y

Supplementary Table S4.5- SNPs and the corresponding weights used in the construction of the 93 SNP T2D PGS

log(OR) from Morris et al (2012) was used for weighting the SNPs included in the 93 SNP T2D PGS

rsid	EFFECT_ALLELE	OTHER_ALLELE	OR	log_OR	SNP_in_EPICN	flag	Included in PGS?
rs2296172	G	A	1.06	0.058269	Y	NA	Y
rs17106184	G	A	1.1	0.09531	Y	NA	Y
rs10923931	T	G	1.1	0.09531	Y	NA	Y
rs340874	C	T	1.08	0.076961	Y	NA	Y
rs2820443	T	C	1.06	0.058269	Y	NA	Y
rs780094	C	T	1.04	0.039221	Y	Pleiotropic locus	N
rs10203174	C	T	1.15	0.139762	Y	NA	Y
rs243021	A	G	1.09	0.086178	Y	NA	Y
rs6742799	A	C	1.11	0.10436	Y	NA	Y
rs13389219	C	T	1.05	0.04879	Y	NA	Y
rs2943641	C	T	1.08	0.076961	Y	NA	Y
rs1801282	C	G	1.16	0.14842	Y	NA	Y
rs1496653	A	G	1.08	0.076961	Y	NA	Y
rs2276853	A	G	1.06	0.058269	Y	NA	Y
rs6795735	C	T	1.07	0.067659	Y	NA	Y
rs11708067	A	G	1.1	0.09531	Y	NA	Y
rs4402960	T	G	1.13	0.122218	Y	NA	Y
rs6808574	C	T	1.08	0.076961	Y	NA	Y
rs4458523	G	T	1.09	0.086178	Y	NA	Y
rs6813195	C	T	1.08	0.076961	Y	NA	Y
rs1996546	G	T	1.1	0.09531	Y	NA	Y
rs702634	A	G	1.08	0.076961	Y	NA	Y
rs459193	G	A	1.05	0.04879	Y	NA	Y
rs3843467	T	G	1.1	0.09531	Y	NA	Y
rs4457053	G	A	1.13	0.122218	Y	NA	Y
rs9502570	C	T	1.08	0.076961	Y	NA	Y
rs7756992	G	A	1.2	0.182322	Y	NA	Y
rs3132524	C	T	1.08	0.076961	Y	NA	Y
rs9271775	T	C	1.09	0.086178	Y	NA	Y
rs4407733	A	G	1.07	0.067659	Y	NA	Y
rs10276674	C	T	1.13	0.122218	Y	NA	Y
rs2191349	T	G	1.07	0.067659	Y	NA	Y
rs864745	T	C	1.12	0.113329	Y	NA	Y
rs3996352	A	G	1.1	0.09531	Y	NA	Y
rs1182397	G	T	1.08	0.076961	Y	NA	Y
rs10503669	C	A	1.08	0.076961	Y	NA	Y
rs516946	C	T	1.1	0.09531	Y	NA	Y
rs7845219	T	C	1.08	0.076961	Y	NA	Y
rs3802177	G	A	1.16	0.14842	Y	NA	Y
rs10758593	A	G	1.06	0.058269	Y	NA	Y
rs2184061	A	C	1.08	0.076961	Y	NA	Y
rs10757282	C	T	1.07	0.067659	Y	NA	Y
rs10811661	T	C	1.18	0.165514	Y	NA	Y
rs17791513	A	G	1.21	0.19062	Y	NA	Y
rs2796441	G	A	1.07	0.067659	Y	NA	Y
rs505922	C	T	1.04	0.039221	Y	NA	Y
rs11257655	T	C	1.06	0.058269	Y	NA	Y
rs4746890	T	C	1.05	0.04879	Y	NA	Y

rs12571751	A	G	1.09	0.086178	Y	NA	Y
rs1111875	C	T	1.15	0.139762	Y	NA	Y
rs7903146	T	C	1.4	0.336472	Y	NA	Y
rs2421016	C	T	1.08	0.076961	Y	NA	Y
rs2334499	T	C	1.07	0.067659	Y	NA	Y
rs10840595	C	T	1.1	0.09531	Y	NA	Y
rs231361	A	G	1.1	0.09531	Y	NA	Y
rs2237892	C	T	1.12	0.113329	Y	NA	Y
rs233449	G	A	1.08	0.076961	Y	NA	Y
rs384037	A	G	1.07	0.067659	Y	NA	Y
rs5215	C	T	1.08	0.076961	Y	NA	Y
rs3736505	G	A	1.07	0.067659	Y	NA	Y
rs11227234	T	G	1.07	0.067659	Y	NA	Y
rs1552224	A	C	1.13	0.122218	Y	NA	Y
rs10830961	G	A	1.12	0.113329	Y	NA	Y
rs10219509	C	T	1.05	0.04879	Y	NA	Y
rs11063069	G	A	1.1	0.09531	Y	NA	Y
rs10842994	C	T	1.09	0.086178	Y	NA	Y
rs2261181	T	C	1.16	0.14842	Y	NA	Y
rs7961581	C	T	1.08	0.076961	Y	NA	Y
rs7957197	T	A	1.12	0.113329	Y	NA	Y
rs1106240	T	C	1.07	0.067659	Y	NA	Y
rs1359790	G	A	1.1	0.09531	Y	NA	Y
rs10146997	G	A	1.06	0.058269	Y	NA	Y
rs4502156	T	C	1.06	0.058269	Y	NA	Y
rs7178572	G	A	1.08	0.076961	Y	NA	Y
rs2866367	G	T	1.08	0.076961	Y	NA	Y
rs1371135	T	C	1.05	0.04879	Y	NA	Y
rs12899811	G	A	1.09	0.086178	Y	NA	Y
rs9939609	A	T	1.11	0.10436	Y	NA	Y
rs9921586	G	T	1.15	0.139762	Y	NA	Y
rs2925979	T	C	1.06	0.058269	Y	NA	Y
rs8068804	A	G	1.06	0.058269	Y	NA	Y
rs17676067	C	T	1.06	0.058269	Y	NA	Y
rs4430796	G	A	1.13	0.122218	Y	NA	Y
rs15563	G	A	1.06	0.058269	Y	NA	Y
rs9891146	T	C	1.05	0.04879	Y	NA	Y
rs9948462	T	C	1.07	0.067659	Y	NA	Y
rs12970134	A	G	1.08	0.076961	Y	NA	Y
rs16996148	T	G	1.11	0.10436	Y	NA	Y
rs8108269	G	T	1.06	0.058269	Y	NA	Y
rs17265513	C	T	1.11	0.10436	Y	NA	Y
rs4812829	A	G	1.07	0.067659	Y	NA	Y
rs1800961	T	C	1.13	0.122218	Y	NA	Y
rs36608	A	G	1.13	0.122218	Y	NA	Y
rs738409	G	C	1.06	0.058269	Y	NA	Y

Supplementary Table S4.6: Metabolite associations with the T2D BMI unadjusted Genetic Risk Score in EPIC-Norfolk Quasi-random subcohort (FDR adjusted P (q) < 0.05) in PGS analyses

PGS: Metabolite associations with the T2D BMI unadjusted Genetic Risk Score in EPIC-Norfolk Quasi-random subcohort. Fixed effect metanalysis of two measurement batches in the EPIC-Norfolk quasi-random sub-cohort

Observational: metabolites associated with Incident T2D in the EPIC-Norfolk T2D Case-Cohort. Base Model - Incident T2D ~ metabolite + age + sex

	T2D PGS				Observational							Metabolite		
ID	beta	SE	Pvalue	P.FDR	log_HR	HR	SE	Pvalue	P.FDR	N_total	N_T2D_case	BIOCHEMICAL	SUB_PATHWAY	SUPER_PATHWAY
M48153	0.359	0.039	5.56E-20	5.51E-17	1.298	3.663	0.043	1.54E-25	4.20E-23	1496	670	mannose	Fructose, Mannose and Galactose Metabolism	Carbohydrate
M48152	0.307	0.039	6.39E-15	2.11E-12	0.902	2.464	0.032	2.71E-15	7.39E-14	1500	671	glucose	Glycolysis, Gluconeogenesis, and Pyruvate Metabolism	Carbohydrate
M52478	-0.296	0.038	5.27E-15	2.11E-12	-0.406	0.667	0.039	1.88E-09	1.79E-08	1498	670	1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*	Phospholipid Metabolism	Lipid
M52281	0.281	0.039	7.56E-13	1.88E-10	NA	NA	NA	NA	NA	NA	NA	2-hydroxybutyrate/2-hydroxyisobutyrate	Methionine, Cysteine, SAM and Taurine Metabolism	Amino Acid
M52713	-0.265	0.038	3.21E-12	6.37E-10	NA	NA	NA	NA	NA	NA	NA	1-(1-enyl-palmitoyl)-2-palmitoleoyl-GPC (P-16:0/16:1)*	Plasmalogen	Lipid
M52612	-0.251	0.039	8.84E-11	1.46E-08	NA	NA	NA	NA	NA	NA	NA	1-(1-enyl-stearoyl)-2-oleoyl-GPC (P-18:0/18:1)	Plasmalogen	Lipid
M52437	-0.214	0.036	4.53E-09	6.42E-07	-0.524	0.592	0.050	3.81E-10	4.10E-09	1501	672	sphingomyelin (d18:2/24:1, d18:1/24:2)*	Sphingolipid Metabolism	Lipid
M52712	-0.227	0.039	5.56E-09	6.89E-07	NA	NA	NA	NA	NA	NA	NA	1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2)*	Plasmalogen	Lipid
M20675	-0.207	0.038	6.08E-08	6.70E-06	-0.774	0.461	0.030	2.37E-13	4.51E-12	1501	672	1,5-anhydroglucitol (1,5-AG)	Glycolysis, Gluconeogenesis, and Pyruvate Metabolism	Carbohydrate
M52633	0.212	0.039	8.27E-08	8.20E-06	NA	NA	NA	NA	NA	NA	NA	palmitoyl-linoleoyl-glycerol (16:0/18:2) [1]*	Diacylglycerol	Lipid
M52701	-0.200	0.039	3.52E-07	3.17E-05	NA	NA	NA	NA	NA	NA	NA	1-(1-enyl-oleoyl)-GPC (P-18:1)*	Lysoplasmalogen	Lipid
M52682	-0.196	0.039	4.13E-07	3.41E-05	NA	NA	NA	NA	NA	NA	NA	1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)*	Plasmalogen	Lipid
M06146	0.205	0.041	6.77E-07	5.17E-05	0.462	1.587	0.045	2.73E-11	3.73E-10	1356	614	2-aminoadipate	Lysine Metabolism	Amino Acid
M44560	-0.192	0.039	1.02E-06	6.72E-05	-0.556	0.573	0.045	5.83E-14	1.33E-12	1453	651	1-eicosenoyl-GPC (20:1)*	Lysolipid	Lipid
M46588	0.190	0.039	1.08E-06	6.72E-05	0.675	1.964	0.045	3.38E-20	2.31E-18	1296	619	X - 12063	Unknown	Unknown
M52629	0.192	0.039	1.01E-06	6.72E-05	NA	NA	NA	NA	NA	NA	NA	1-stearoyl-2-dihomo-linolenoyl-GPC (18:0/20:3n3 or 6)*	Phospholipid Metabolism	Lipid
M48491	-0.178	0.037	1.43E-06	7.91E-05	-0.314	0.730	0.045	4.01E-06	2.18E-05	1501	672	sphingomyelin (d18:1/20:1, d18:2/20:0)*	Sphingolipid Metabolism	Lipid
M52704	-0.186	0.038	1.36E-06	7.91E-05	NA	NA	NA	NA	NA	NA	NA	1-(1-enyl-stearoyl)-2-docosahexaenoyl-GPC (P-18:0/22:6)*	Plasmalogen	Lipid
M52613	-0.186	0.040	2.43E-06	1.27E-04	NA	NA	NA	NA	NA	NA	NA	1-(1-enyl-stearoyl)-2-arachidonoyl-GPC (P-18:0/20:4)	Plasmalogen	Lipid
M43829	0.172	0.037	2.71E-06	1.34E-04	0.549	1.731	0.053	8.58E-11	1.02E-09	1424	644	gamma-glutamylvaline	Gamma-glutamyl Amino Acid	Peptide
M15676	0.175	0.038	5.13E-06	2.42E-04	0.446	1.563	0.044	1.25E-10	1.42E-09	1501	672	3-methyl-2-oxovalerate	Leucine, Isoleucine and Valine Metabolism	Amino Acid
M34456	0.166	0.037	5.80E-06	2.62E-04	0.287	1.332	0.047	4.67E-05	2.02E-04	1116	543	gamma-glutamylisoleucine*	Gamma-glutamyl Amino Acid	Peptide
M22116	0.169	0.038	7.10E-06	2.89E-04	0.378	1.460	0.045	2.26E-08	1.81E-07	1501	672	4-methyl-2-oxopentanoate	Leucine, Isoleucine and Valine Metabolism	Amino Acid

	T2D PGS				Observational							Metabolite		
ID	beta	SE	Pvalue	P.FDR	log_HR	HR	SE	Pvalue	P.FDR	N_total	N_T2D_case	BIOCHEMICAL	SUB_PATHWAY	SUPER_PATHWAY
M22137	0.171	0.038	7.29E-06	2.89E-04	0.319	1.375	0.042	1.95E-07	1.34E-06	1501	672	homoarginine	Urea cycle; Arginine and Proline Metabolism	Amino Acid
M44526	0.174	0.039	7.03E-06	2.89E-04	0.602	1.826	0.044	2.84E-16	8.62E-15	1501	672	3-methyl-2-oxobutyrate	Leucine, Isoleucine and Valine Metabolism	Amino Acid
M35107	-0.199	0.045	1.04E-05	3.96E-04	0.044	1.045	0.046	5.08E-01	6.31E-01	1236	538	isovalerylglycine	Leucine, Isoleucine and Valine Metabolism	Amino Acid
M33953	0.169	0.039	1.42E-05	5.20E-04	NA	NA	NA	NA	NA	NA	NA	N-acetylarginine	Urea cycle; Arginine and Proline Metabolism	Amino Acid
M52716	-0.171	0.040	1.58E-05	5.59E-04	NA	NA	NA	NA	NA	NA	NA	1-(1-enyl-palmitoyl)-2-palmitoyl-GPC (P-16:0/16:0)*	Plasmalogen	Lipid
M52474	-0.166	0.039	2.34E-05	7.75E-04	-0.577	0.561	0.037	2.18E-18	8.49E-17	1481	658	1-(1-enyl-palmitoyl)-GPC (P-16:0)*	Phospholipid Metabolism	Lipid
M52634	0.166	0.039	2.32E-05	7.75E-04	NA	NA	NA	NA	NA	NA	NA	palmitoyl-linoleoyl-glycerol (16:0/18:2) [2]*	Diacylglycerol	Lipid
M36738	0.163	0.039	2.60E-05	8.23E-04	0.367	1.443	0.045	9.13E-08	6.67E-07	1295	584	gamma-glutamylglutamate	Gamma-glutamyl Amino Acid	Peptide
M46225	-0.129	0.031	2.65E-05	8.23E-04	-0.352	0.704	0.050	4.75E-06	2.56E-05	1501	672	pyroglutamine*	Glutamate Metabolism	Amino Acid
M00057	0.159	0.038	3.13E-05	9.41E-04	0.734	2.083	0.041	5.57E-25	9.12E-23	1501	672	glutamate	Glutamate Metabolism	Amino Acid
M35625	0.164	0.040	3.24E-05	9.46E-04	0.434	1.543	0.039	2.32E-13	4.51E-12	1490	670	1-myristoylglycerol (14:0)	Monoacylglycerol	Lipid
M01649	0.157	0.038	3.58E-05	9.88E-04	0.413	1.512	0.039	8.33E-12	1.24E-10	1501	672	valine	Leucine, Isoleucine and Valine Metabolism	Amino Acid
M19258	0.162	0.039	3.52E-05	9.88E-04	0.402	1.494	0.041	8.79E-11	1.03E-09	1501	672	1-myristoyl-2-palmitoyl-GPC (14:0/16:0)	Phospholipid Metabolism	Lipid
M52470	0.158	0.039	4.62E-05	1.24E-03	0.696	2.006	0.041	1.59E-21	1.57E-19	1501	672	1-palmitoyl-2-palmitoleoyl-GPC (16:0/16:1)*	Phospholipid Metabolism	Lipid
M27710	-0.153	0.038	5.59E-05	1.46E-03	-0.854	0.426	0.052	3.11E-24	4.24E-22	1501	672	N-acetylglycine	Glycine, Serine and Threonine Metabolism	Amino Acid
M15705	0.166	0.041	6.09E-05	1.51E-03	0.322	1.380	0.044	1.38E-06	8.32E-06	1309	602	cystathionine	Methionine, Cysteine, SAM and Taurine Metabolism	Amino Acid
M46799	0.156	0.039	6.03E-05	1.51E-03	NA	NA	NA	NA	NA	NA	NA	oleoyl-linoleoyl-glycerol (18:1/18:2) [2]	Diacylglycerol	Lipid
M42374	0.156	0.039	7.44E-05	1.80E-03	0.061	1.062	0.040	2.73E-01	3.88E-01	1500	671	2-aminobutyrate	Methionine, Cysteine, SAM and Taurine Metabolism	Amino Acid
M44630	0.155	0.040	1.00E-04	2.36E-03	0.094	1.098	0.041	1.02E-01	1.72E-01	1442	648	1-eicosatrienoyl-GPE (20:3)*	Lysolipid	Lipid
M46966	-0.179	0.046	1.17E-04	2.71E-03	NA	NA	NA	NA	NA	NA	NA	X - 12705	Unknown	Unknown
M37506	-0.146	0.038	1.23E-04	2.75E-03	-0.417	0.659	0.043	2.01E-09	1.89E-08	1501	672	palmitoyl sphingomyelin (d18:1/16:0)	Sphingolipid Metabolism	Lipid
M46611	-0.175	0.045	1.25E-04	2.75E-03	NA	NA	NA	NA	NA	NA	NA	X - 12013	Unknown	Unknown
M53261	-0.149	0.039	1.30E-04	2.79E-03	NA	NA	NA	NA	NA	NA	NA	arachidonoylcholine	Phospholipid Metabolism	Lipid
M00058	-0.143	0.038	1.46E-04	3.07E-03	-0.720	0.487	0.043	2.99E-26	1.23E-23	1501	672	glycine	Glycine, Serine and Threonine Metabolism	Amino Acid
M34445	-0.150	0.040	1.49E-04	3.09E-03	0.032	1.033	0.038	5.45E-01	6.68E-01	1500	671	sphingosine 1-phosphate	Sphingolipid Metabolism	Lipid
M39271	-0.149	0.039	1.64E-04	3.31E-03	-0.291	0.747	0.041	3.56E-06	1.96E-05	1501	672	1-(1-enyl-stearoyl)-GPE (P-18:0)*	Lysolipid	Lipid
M52707	0.146	0.039	1.71E-04	3.39E-03	NA	NA	NA	NA	NA	NA	NA	1-myristoyl-2-linoleoyl-GPC (14:0/18:2)*	Phospholipid Metabolism	Lipid
M47087	-0.148	0.040	1.81E-04	3.43E-03	NA	NA	NA	NA	NA	NA	NA	1-nonadecanoyl-GPC (19:0)	Lysolipid	Lipid
M48259	-0.147	0.039	1.83E-04	3.43E-03	-0.242	0.785	0.040	4.42E-05	1.94E-04	1501	672	2-oleoyl-GPC (18:1)*	Lysolipid	Lipid
M52689	-0.148	0.039	1.79E-04	3.43E-03	NA	NA	NA	NA	NA	NA	NA	1-(1-enyl-palmitoyl)-2-arachidonoyl-GPC (P-16:0/20:4)*	Plasmalogen	Lipid
M44682	-0.145	0.039	2.13E-04	3.91E-03	-0.252	0.778	0.038	5.42E-06	2.86E-05	1499	672	1-margaroyl-GPC (17:0)	Lysolipid	Lipid

	T2D PGS				Observational							Metabolite		
ID	beta	SE	Pvalue	P.FDR	log_HR	HR	SE	Pvalue	P.FDR	N_total	N_T2D_case	BIOCHEMICAL	SUB_PATHWAY	SUPER_PATHWAY
M45456	-0.147	0.040	2.22E-04	4.00E-03	-0.506	0.603	0.055	1.15E-09	1.16E-08	921	371	1-arachidoyl-GPC (20:0)	Lysolipid	Lipid
M42459	-0.134	0.036	2.27E-04	4.02E-03	-0.142	0.868	0.049	6.28E-02	1.15E-01	1501	672	sphingomyelin (d18:2/16:0, d18:1/16:1)*	Sphingolipid Metabolism	Lipid
M47929	-0.144	0.039	2.34E-04	4.07E-03	NA	NA	NA	NA	NA	NA	NA	X - 12707	Unknown	Unknown
M46798	0.143	0.039	2.42E-04	4.13E-03	NA	NA	NA	NA	NA	NA	NA	oleoyl-linoleoyl-glycerol (18:1/18:2) [1]	Diacylglycerol	Lipid
M46633	0.139	0.038	2.94E-04	4.95E-03	0.160	1.173	0.043	1.01E-02	2.40E-02	1311	597	X - 12844	Unknown	Unknown
M52715	-0.147	0.041	3.34E-04	5.52E-03	NA	NA	NA	NA	NA	NA	NA	1-(1-enyl-palmitoyl)-2-myristoyl-GPC (P-16:0/14:0)*	Plasmalogen	Lipid
M32412	0.140	0.039	3.53E-04	5.57E-03	0.351	1.420	0.038	1.60E-09	1.56E-08	1325	631	butyrylcarnitine	Fatty Acid Metabolism (also BCAA Metabolism)	Lipid
M52135	-0.142	0.040	3.52E-04	5.57E-03	NA	NA	NA	NA	NA	NA	NA	X - 24106	Unknown	Unknown
M52604	0.147	0.041	3.54E-04	5.57E-03	NA	NA	NA	NA	NA	NA	NA	N-palmitoyl-sphinganine (d18:0/16:0)	Sphingolipid Metabolism	Lipid
M46613	-0.140	0.039	3.67E-04	5.63E-03	-0.257	0.773	0.038	1.99E-05	9.27E-05	1439	638	X - 12216	Unknown	Unknown
M52944	-0.139	0.039	3.75E-04	5.63E-03	NA	NA	NA	NA	NA	NA	NA	palmitoylcholine	Fatty Acid Metabolism (Acyl Choline)	Lipid
M53260	-0.139	0.039	3.75E-04	5.63E-03	NA	NA	NA	NA	NA	NA	NA	oleoylcholine	Phospholipid Metabolism	Lipid
M15990	-0.139	0.039	4.02E-04	5.96E-03	-0.402	0.669	0.036	1.62E-11	2.29E-10	1501	672	glycerophosphorylcholine (GPC)	Phospholipid Metabolism	Lipid
M01299	0.138	0.039	4.08E-04	5.96E-03	0.265	1.303	0.041	1.34E-05	6.59E-05	1501	672	tyrosine	Phenylalanine and Tyrosine Metabolism	Amino Acid
M20458	-0.138	0.039	4.34E-04	6.24E-03	-0.441	0.643	0.050	9.40E-09	8.19E-08	986	415	1-palmityl-GPC (O-16:0)	Lysolipid	Lipid
M47347	0.146	0.042	4.62E-04	6.55E-03	NA	NA	NA	NA	NA	NA	NA	X - 18889	Unknown	Unknown
M53263	-0.134	0.039	6.05E-04	8.45E-03	NA	NA	NA	NA	NA	NA	NA	docosahexaenoylcholine	Phospholipid Metabolism	Lipid
M22185	-0.130	0.038	6.20E-04	8.55E-03	-0.309	0.734	0.039	4.03E-07	2.64E-06	1500	671	N-acetylaspartate (NAA)	Alanine and Aspartate Metabolism	Amino Acid
M52446	0.134	0.040	6.85E-04	9.31E-03	0.484	1.623	0.042	1.60E-13	3.36E-12	1501	672	1-stearoyl-2-linoleoyl-GPE (18:0/18:2)*	Phospholipid Metabolism	Lipid
M01589	-0.129	0.039	8.18E-04	1.02E-02	-0.602	0.548	0.037	8.13E-10	8.43E-09	1492	665	N-acetylmethionine	Methionine, Cysteine, SAM and Taurine Metabolism	Amino Acid
M18369	0.123	0.037	8.45E-04	1.02E-02	0.134	1.144	0.042	2.25E-02	4.83E-02	1501	672	gamma-glutamylleucine	Gamma-glutamyl Amino Acid	Peptide
M19263	0.135	0.040	8.06E-04	1.02E-02	0.719	2.053	0.046	5.47E-19	2.49E-17	1427	656	1-palmitoyl-2-oleoyl-GPE (16:0/18:1)	Phospholipid Metabolism	Lipid
M46643	0.127	0.038	8.02E-04	1.02E-02	NA	NA	NA	NA	NA	NA	NA	X - 13529	Unknown	Unknown
M46662	0.120	0.036	7.85E-04	1.02E-02	0.387	1.472	0.047	1.20E-07	8.63E-07	1144	554	X - 15492	Unknown	Unknown
M47653	0.125	0.037	8.04E-04	1.02E-02	NA	NA	NA	NA	NA	NA	NA	X - 11593	Unknown	Unknown
M52431	0.132	0.039	8.22E-04	1.02E-02	0.605	1.831	0.043	1.13E-18	4.89E-17	1497	671	1-palmitoleoylglycerol (16:1)*	Monoacylglycerol	Lipid
M52466	0.128	0.038	8.29E-04	1.02E-02	0.714	2.042	0.046	1.63E-17	5.81E-16	1494	671	1-stearoyl-2-docosahexaenoyl-GPE (18:0/22:6)*	Phospholipid Metabolism	Lipid
M52630	-0.130	0.039	8.46E-04	1.02E-02	NA	NA	NA	NA	NA	NA	NA	1-(1-enyl-palmitoyl)-2-docosahexaenoyl-GPC (P-16:0/22:6)*	Plasmalogen	Lipid
M53157	0.126	0.038	9.13E-04	1.09E-02	NA	NA	NA	NA	NA	NA	NA	X - 24728	Unknown	Unknown
M52467	0.131	0.040	9.40E-04	1.11E-02	0.416	1.516	0.043	2.82E-10	3.16E-09	1303	598	1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4)*	Phospholipid Metabolism	Lipid
M53010	-0.129	0.039	1.02E-03	1.19E-02	NA	NA	NA	NA	NA	NA	NA	lactosyl-N-palmitoyl-sphingosine	Sphingolipid Metabolism	Lipid
M33132	0.122	0.037	1.10E-03	1.26E-02	0.092	1.096	0.040	1.09E-01	1.82E-01	1501	672	X - 11787	Unknown	Unknown
M00587	0.122	0.038	1.34E-03	1.51E-02	0.522	1.686	0.038	2.92E-10	3.23E-09	1481	663	gluconate	Food Component/Plant	Xenobiotics
M01549	0.127	0.040	1.35E-03	1.51E-02	0.328	1.388	0.038	2.45E-09	2.28E-08	1500	671	3-hydroxyisobutyrate	Leucine, Isoleucine and Valine Metabolism	Amino Acid

	T2D PGS				Observational							Metabolite		
ID	beta	SE	Pvalue	P.FDR	log_HR	HR	SE	Pvalue	P.FDR	N_total	N_T2D_case	BIOCHEMICAL	SUB_PATHWAY	SUPER_PATHWAY
M39600	0.126	0.039	1.40E-03	1.51E-02	-0.169	0.845	0.039	2.34E-03	6.66E-03	1493	669	3-hydroxyhippurate	Benzoate Metabolism	Xenobiotics
M47348	0.134	0.042	1.39E-03	1.51E-02	NA	NA	NA	NA	NA	NA	NA	X - 22102	Unknown	Unknown
M48001	0.133	0.041	1.38E-03	1.51E-02	0.041	1.042	0.049	5.49E-01	6.70E-01	879	406	X - 17351	Unknown	Unknown
M48258	-0.125	0.039	1.39E-03	1.51E-02	-0.232	0.793	0.042	2.80E-04	1.00E-03	1501	672	1-oleoyl-GPC (18:1)	Lysolipid	Lipid
M12017	-0.122	0.038	1.44E-03	1.54E-02	-0.208	0.812	0.041	4.18E-04	1.45E-03	1394	619	3-methoxytyrosine	Phenylalanine and Tyrosine Metabolism	Amino Acid
M52703	-0.126	0.040	1.52E-03	1.60E-02	NA	NA	NA	NA	NA	NA	NA	1-stearyl-GPC (O-18:0)*	Lyso-phospho-ether	Lipid
M46677	0.139	0.044	1.69E-03	1.77E-02	NA	NA	NA	NA	NA	NA	NA	X - 16654	Unknown	Unknown
M33949	-0.120	0.038	1.78E-03	1.84E-02	-0.645	0.524	0.035	2.89E-18	1.08E-16	1492	665	gamma-glutamylglycine	Gamma-glutamyl Amino Acid	Peptide
M42370	0.128	0.041	1.81E-03	1.85E-02	NA	NA	NA	NA	NA	NA	NA	S-1-pyrroline-5-carboxylate	Glutamate Metabolism	Amino Acid
M52697	-0.118	0.038	1.90E-03	1.93E-02	NA	NA	NA	NA	NA	NA	NA	1-oleoyl-2-docosahexaenoyl-GPC (18:1/22:6)*	Phospholipid Metabolism	Lipid
M42449	0.122	0.039	1.99E-03	1.99E-02	0.488	1.629	0.042	1.11E-13	2.45E-12	1501	672	1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	Phospholipid Metabolism	Lipid
M27731	0.119	0.039	2.48E-03	2.46E-02	NA	NA	NA	NA	NA	NA	NA	ribonate	Pentose Metabolism	Carbohydrate
M37455	-0.118	0.039	2.53E-03	2.48E-02	-0.178	0.837	0.039	1.40E-03	4.26E-03	1498	670	glycerophosphoethanolamine	Phospholipid Metabolism	Lipid
M35257	-0.118	0.039	2.61E-03	2.54E-02	-0.467	0.627	0.040	1.78E-13	3.64E-12	1494	668	2-linoleoyl-GPC (18:2)*	Lysolipid	Lipid
M33441	-0.119	0.040	2.65E-03	2.55E-02	-0.291	0.748	0.041	7.47E-07	4.67E-06	1466	655	isobutyrylcarnitine	Leucine, Isoleucine and Valine Metabolism	Amino Acid
M52234	-0.120	0.040	2.77E-03	2.64E-02	NA	NA	NA	NA	NA	NA	NA	glycosyl-N-stearoyl-sphingosine	Sphingolipid Metabolism	Lipid
M49461	-0.124	0.042	2.83E-03	2.68E-02	NA	NA	NA	NA	NA	NA	NA	X - 23585	Unknown	Unknown
M02734	0.114	0.039	3.12E-03	2.92E-02	0.088	1.092	0.044	1.59E-01	2.49E-01	1446	648	gamma-glutamyltyrosine	Gamma-glutamyl Amino Acid	Peptide
M12129	0.110	0.038	3.37E-03	3.09E-02	0.255	1.291	0.043	8.04E-05	3.15E-04	1501	672	beta-hydroxyisovalerate	Leucine, Isoleucine and Valine Metabolism	Amino Acid
M48434	0.139	0.047	3.34E-03	3.09E-02	0.241	1.273	0.049	2.97E-03	8.25E-03	765	355	N-acetylcitrulline	Urea cycle; Arginine and Proline Metabolism	Amino Acid
M01125	0.112	0.038	3.42E-03	3.11E-02	0.289	1.335	0.039	3.48E-07	2.32E-06	1501	672	isoleucine	Leucine, Isoleucine and Valine Metabolism	Amino Acid
M01519	-0.123	0.042	3.46E-03	3.12E-02	0.120	1.127	0.041	3.62E-02	7.27E-02	1363	621	sucrose	Disaccharides and Oligosaccharides	Carbohydrate
M00443	0.115	0.039	3.59E-03	3.20E-02	0.461	1.585	0.043	8.12E-12	1.24E-10	1501	672	aspartate	Alanine and Aspartate Metabolism	Amino Acid
M46515	0.113	0.039	3.70E-03	3.28E-02	-0.031	0.970	0.057	7.30E-01	8.15E-01	782	356	X - 21470	Unknown	Unknown
M33959	0.116	0.040	3.82E-03	3.36E-02	0.270	1.309	0.044	8.07E-05	3.15E-04	1332	610	N-acetyltryptophan	Tryptophan Metabolism	Amino Acid
M48493	-0.109	0.038	4.13E-03	3.60E-02	-0.205	0.814	0.039	2.69E-04	9.67E-04	1496	669	sphingomyelin (d18:1/22:1, d18:2/22:0, d16:1/24:1)*	Sphingolipid Metabolism	Lipid
M01638	0.113	0.040	4.21E-03	3.61E-02	-0.088	0.916	0.039	1.10E-01	1.82E-01	1501	672	arginine	Urea cycle; Arginine and Proline Metabolism	Amino Acid
M46599	0.100	0.035	4.22E-03	3.61E-02	0.170	1.186	0.047	1.18E-02	2.80E-02	1473	661	X - 11440	Unknown	Unknown
M48187	-0.109	0.038	4.42E-03	3.72E-02	-0.121	0.886	0.040	3.27E-02	6.69E-02	1448	648	N-acetyltaurine	Methionine, Cysteine, SAM and Taurine Metabolism	Amino Acid
M52476	-0.111	0.039	4.40E-03	3.72E-02	-0.155	0.857	0.041	9.82E-03	2.35E-02	1500	672	1-(1-enyl-stearoyl)-2-docosahexaenoyl-GPE (P-18:0/22:6)*	Phospholipid Metabolism	Lipid

	T2D PGS				Observational							Metabolite		
ID	beta	SE	Pvalue	P.FDR	log_HR	HR	SE	Pvalue	P.FDR	N_total	N_T2D_case	BIOCHEMICAL	SUB_PATHWAY	SUPER_PATHWAY
M15765	0.111	0.039	4.56E-03	3.72E-02	0.286	1.331	0.034	2.32E-08	1.85E-07	1500	671	ethylmalonate	Leucine, Isoleucine and Valine Metabolism	Amino Acid
M32452	0.108	0.038	4.52E-03	3.72E-02	0.240	1.272	0.041	2.83E-05	1.28E-04	1501	672	propionylcarnitine	Fatty Acid Metabolism (also BCAA Metabolism)	Lipid
M38399	-0.128	0.045	4.49E-03	3.72E-02	NA	NA	NA	NA	NA	NA	NA	9,10-DiHOME	Fatty Acid, Dihydroxy	Lipid
M47708	0.128	0.045	4.57E-03	3.72E-02	NA	NA	NA	NA	NA	NA	NA	X - 12283	Unknown	Unknown
M32390	0.116	0.041	4.70E-03	3.79E-02	0.312	1.366	0.047	1.99E-05	9.27E-05	1221	565	N-acetyltyrosine	Phenylalanine and Tyrosine Metabolism	Amino Acid
M53224	-0.117	0.041	4.81E-03	3.85E-02	NA	NA	NA	NA	NA	NA	NA	pimeloylcarnitine/3-methyladipoylcarnitine	Fatty Acid Metabolism(Acyl Carnitine)	Lipid
M46409	0.103	0.037	5.03E-03	3.99E-02	0.193	1.213	0.046	3.30E-03	9.03E-03	1164	525	X - 21364	Unknown	Unknown
M00060	0.106	0.038	5.23E-03	4.12E-02	0.260	1.297	0.039	3.91E-06	2.13E-05	1501	672	leucine	Leucine, Isoleucine and Valine Metabolism	Amino Acid
M46384	0.108	0.039	5.75E-03	4.49E-02	0.183	1.201	0.042	2.19E-03	6.26E-03	1303	587	X - 21339	Unknown	Unknown
M35126	-0.105	0.038	5.91E-03	4.51E-02	-0.044	0.957	0.041	4.60E-01	5.82E-01	1501	672	phenylacetylglutamine	Phenylalanine and Tyrosine Metabolism	Amino Acid
M47784	0.158	0.057	5.83E-03	4.51E-02	NA	NA	NA	NA	NA	NA	NA	X - 22520	Unknown	Unknown
M52433	-0.104	0.038	5.86E-03	4.51E-02	-0.260	0.771	0.041	1.38E-05	6.73E-05	1501	672	sphingomyelin (d18:1/15:0, d16:1/17:0)*	Sphingolipid Metabolism	Lipid
M47872	0.104	0.038	6.26E-03	4.74E-02	NA	NA	NA	NA	NA	NA	NA	X - 17340	Unknown	Unknown
M35437	-0.125	0.046	6.33E-03	4.75E-02	0.259	1.295	0.049	4.12E-04	1.44E-03	1250	555	isobutyrylglycine	Leucine, Isoleucine and Valine Metabolism	Amino Acid
M37202	0.096	0.035	6.52E-03	4.86E-02	0.225	1.253	0.048	1.14E-03	3.52E-03	1496	667	4-androsten-3beta,17beta-diol disulfate (1)	Steroid	Lipid
M27718	0.094	0.035	6.71E-03	4.97E-02	0.037	1.038	0.041	5.12E-01	6.35E-01	1501	672	creatine	Creatine Metabolism	Amino Acid

Supplementary Table S4.7- Metabolite associations with the partial T2D genetic risk scores defined by Udler et al (2018) in EPIC-Norfolk quasi-random subcohort (FDR < 0.05)

Fixed effect meta-analysis of two measurement batches in the EPIC-Norfolk quasi-random sub-cohort

het_pvalue denotes heterogeneity p value for fixed effect metanalysis across both measurement batches

pPGS	ID	BIOCHEMICAL	SUPER_PATHWAY	SUB_PATHWAY	Beta (95% CI)	pvalue	pvalue.adjFDR	het_pvalue
Lipodystrophy	M52713	1-(1-enyl-palmitoyl)-2-palmitoleoyl-GPC (P-16:0/16:1)*	Plasmalogen	Lipid	-0.08 (-0.08,-0.08)	2.13E-18	1.06E-14	0.490
Lipodystrophy	M52478	1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1)*	Phospholipid Metabolism	Lipid	-0.078 (-0.078,-0.078)	2.12E-17	5.25E-14	0.941
Lipodystrophy	M52633	palmitoyl-linoleoyl-glycerol (16:0/18:2) [1]*	Diacylglycerol	Lipid	0.072 (0.072,0.072)	3.12E-14	5.16E-11	0.307
Lipodystrophy	M52634	palmitoyl-linoleoyl-glycerol (16:0/18:2) [2]*	Diacylglycerol	Lipid	0.067 (0.067,0.067)	1.42E-12	1.76E-09	0.633
Lipodystrophy	M46798	oleoyl-linoleoyl-glycerol (18:1/18:2) [1]	Diacylglycerol	Lipid	0.065 (0.065,0.065)	4.23E-12	4.19E-09	0.770
Lipodystrophy	M52612	1-(1-enyl-stearoyl)-2-oleoyl-GPC (P-18:0/18:1)	Plasmalogen	Lipid	-0.063 (-0.063,-0.063)	1.55E-11	1.28E-08	0.371
Lipodystrophy	M52437	sphingomyelin (d18:2/24:1, d18:1/24:2)*	Sphingolipid Metabolism	Lipid	-0.056 (-0.056,-0.056)	2.67E-10	1.89E-07	0.763
Beta Cell	M48153	mannose	Fructose, Mannose and Galactose Metabolism	Carbohydrate	0.06 (0.06,0.06)	4.60E-10	2.85E-07	0.945
Lipodystrophy	M46799	oleoyl-linoleoyl-glycerol (18:1/18:2) [2]	Diacylglycerol	Lipid	0.057 (0.057,0.057)	1.15E-09	6.34E-07	1.000
Lipodystrophy	M52701	1-(1-enyl-oleoyl)-GPC (P-18:1)*	Lysoplasmalogen	Lipid	-0.054 (-0.054,-0.054)	1.19E-08	5.88E-06	0.819
Lipodystrophy	M52712	1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2)*	Plasmalogen	Lipid	-0.052 (-0.052,-0.052)	2.69E-08	1.22E-05	0.566
Lipodystrophy	M22116	4-methyl-2-oxopentanoate	Leucine, Isoleucine and Valine Metabolism	Amino Acid	0.05 (0.05,0.05)	3.49E-08	1.44E-05	0.552
Lipodystrophy	M52446	1-stearoyl-2-linoleoyl-GPE (18:0/18:2)*	Phospholipid Metabolism	Lipid	0.052 (0.052,0.052)	5.84E-08	2.23E-05	0.117
Beta Cell	M48152	glucose	Glycolysis, Gluconeogenesis, and Pyruvate Metabolism	Carbohydrate	0.051 (0.051,0.052)	9.08E-08	3.22E-05	0.351
Lipodystrophy	M15676	3-methyl-2-oxovalerate	Leucine, Isoleucine and Valine Metabolism	Amino Acid	0.049 (0.049,0.049)	1.51E-07	4.99E-05	0.576
Lipodystrophy	M52604	N-palmitoyl-sphinganine (d18:0/16:0)	Sphingolipid Metabolism	Lipid	0.051 (0.051,0.051)	3.62E-07	0.000112	0.237
Liver	M18467	eicosapentaenoate (EPA; 20:5n3)	Polyunsaturated Fatty Acid (n3 and n6)	Lipid	-0.049 (-0.049,-0.048)	4.21E-07	0.000123	0.612
Lipodystrophy	M52682	1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)*	Plasmalogen	Lipid	-0.047 (-0.047,-0.046)	6.00E-07	0.000165	0.477
Lipodystrophy	M48491	sphingomyelin (d18:1/20:1, d18:2/20:0)*	Sphingolipid Metabolism	Lipid	-0.044 (-0.045,-0.044)	7.54E-07	0.000197	0.913
Liver	M52463	1-palmitoyl-2-eicosapentaenoyl-GPC (16:0/20:5)*	Phospholipid Metabolism	Lipid	-0.047 (-0.048,-0.047)	8.43E-07	0.000209	0.730
Liver	M44563	1-eicosapentaenoyl-GPC (20:5)*	Lysolipid	Lipid	-0.047 (-0.047,-0.046)	1.21E-06	0.000286	0.648
Lipodystrophy	M52613	1-(1-enyl-stearoyl)-2-arachidonoyl-GPC (P-18:0/20:4)	Plasmalogen	Lipid	-0.045 (-0.046,-0.044)	2.51E-06	0.000565	0.931

pPGS	ID	BIOCHEMICAL	SUPER_PATHWAY	SUB_PATHWAY	Beta (95% CI)	pvalue	pvalue.adjFDR	het_pvalue
Lipodystrophy	M48009	X - 17357	Unknown	Unknown	0.046 (0.045,0.048)	3.79E-06	0.000818	0.558
Lipodystrophy	M53242	5-bromotryptophan	Phenylalanine and Tyrosine Metabolism	Amino Acid	0.044 (0.042,0.045)	4.18E-06	0.000863	0.569
Lipodystrophy	M52629	1-stearoyl-2-dihomo-linolenoyl-GPC (18:0/20:3n3 or 6)*	Phospholipid Metabolism	Lipid	0.043 (0.042,0.045)	4.68E-06	0.000928	0.166
Lipodystrophy	M42448	1-stearoyl-2-oleoyl-GPE (18:0/18:1)	Phospholipid Metabolism	Lipid	0.044 (0.042,0.046)	4.88E-06	0.00093	0.008
Lipodystrophy	M46588	X - 12063	Unknown	Unknown	0.043 (0.04,0.045)	6.69E-06	0.00123	0.719
Lipodystrophy	M27710	N-acetylglycine	Glycine, Serine and Threonine Metabolism	Amino Acid	-0.041 (-0.044,-0.039)	6.97E-06	0.00123	0.451
Lipodystrophy	M46968	X - 21792	Unknown	Unknown	-0.045 (-0.048,-0.042)	8.18E-06	0.00137	0.452
Lipodystrophy	M52689	1-(1-enyl-palmitoyl)-2-arachidonoyl-GPC (P-16:0/20:4)*	Plasmalogen	Lipid	-0.043 (-0.045,-0.04)	8.35E-06	0.00137	0.551
Obesity	M01604	urate	Purine Metabolism, (Hypo)Xanthine/Inosine containing	Nucleotide	0.038 (0.035,0.04)	8.57E-06	0.00137	0.748
Beta Cell	M01126	alanine	Alanine and Aspartate Metabolism	Amino Acid	0.043 (0.04,0.045)	9.11E-06	0.00138	0.428
Lipodystrophy	M35625	1-myristoylglycerol (14:0)	Monoacylglycerol	Lipid	0.042 (0.04,0.045)	9.24E-06	0.00138	0.668
Lipodystrophy	M44560	1-eicosenoyl-GPC (20:1)*	Lysolipid	Lipid	-0.042 (-0.045,-0.039)	9.46E-06	0.00138	0.078
Beta Cell	M20675	1,5-anhydroglucitol (1,5-AG)	Glycolysis, Gluconeogenesis, and Pyruvate Metabolism	Carbohydrate	-0.041 (-0.044,-0.038)	1.01E-05	0.00142	0.301
Lipodystrophy	M52716	1-(1-enyl-palmitoyl)-2-palmitoyl-GPC (P-16:0/16:0)*	Plasmalogen	Lipid	-0.042 (-0.045,-0.039)	1.15E-05	0.00158	0.397
Liver	M53195	1-myristoyl-2-arachidonoyl-GPC (14:0/20:4)*	Phospholipid Metabolism	Lipid	-0.042 (-0.045,-0.038)	1.38E-05	0.00184	0.513
Lipodystrophy	M52704	1-(1-enyl-stearoyl)-2-docosahexaenoyl-GPC (P-18:0/22:6)*	Plasmalogen	Lipid	-0.04 (-0.044,-0.037)	1.44E-05	0.00188	0.310
Lipodystrophy	M42459	sphingomyelin (d18:2/16:0, d18:1/16:1)*	Sphingolipid Metabolism	Lipid	-0.038 (-0.042,-0.033)	1.73E-05	0.0022	0.960
Liver	M37231	1-docosapentaenoyl-GPC (22:5n3)*	Lysolipid	Lipid	-0.041 (-0.046,-0.037)	1.81E-05	0.00225	0.946
Lipodystrophy	M44526	3-methyl-2-oxobutyrate	Leucine, Isoleucine and Valine Metabolism	Amino Acid	0.04 (0.036,0.045)	1.90E-05	0.0023	0.875
Lipodystrophy	M46633	X - 12844	Unknown	Unknown	0.039 (0.034,0.045)	2.39E-05	0.00277	0.794
Lipodystrophy	M46643	X - 13529	Unknown	Unknown	0.039 (0.033,0.044)	2.41E-05	0.00277	0.114
Lipodystrophy	M52474	1-(1-enyl-palmitoyl)-GPC (P-16:0)*	Phospholipid Metabolism	Lipid	-0.04 (-0.046,-0.034)	2.80E-05	0.00316	0.678
Obesity	M53157	X - 24728	Unknown	Unknown	0.038 (0.031,0.045)	3.09E-05	0.00341	0.768
Obesity	M47695	X - 12117	Unknown	Unknown	0.036 (0.029,0.044)	3.41E-05	0.00368	0.295
Obesity	M52834	X - 24513	Unknown	Unknown	0.034 (0.026,0.042)	3.74E-05	0.00395	0.849
Obesity	M52839	X - 24518	Unknown	Unknown	0.045 (0.037,0.053)	3.93E-05	0.00406	0.125

pPGS	ID	BIOCHEMICAL	SUPER_PATHWAY	SUB_PATHWAY	Beta (95% CI)	pvalue	pvalue.adjFDR	het_pvalue
Lipodystrophy	M48493	sphingomyelin (d18:1/22:1, d18:2/22:0, d16:1/24:1)*	Sphingolipid Metabolism	Lipid	-0.037 (-0.047,-0.028)	4.93E-05	0.00499	0.424
Lipodystrophy	M00058	glycine	Glycine, Serine and Threonine Metabolism	Amino Acid	-0.037 (-0.047,-0.027)	5.21E-05	0.00517	0.292
Lipodystrophy	M01648	serine	Glycine, Serine and Threonine Metabolism	Amino Acid	-0.037 (-0.048,-0.026)	5.68E-05	0.00553	0.537
Lipodystrophy	M34456	gamma-glutamylisoleucine*	Gamma-glutamyl Amino Acid	Peptide	0.035 (0.024,0.047)	6.27E-05	0.00598	0.564
Lipodystrophy	M42449	1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	Phospholipid Metabolism	Lipid	0.038 (0.026,0.05)	6.68E-05	0.00625	0.474
Liver	M52697	1-oleoyl-2-docosahexaenoyl-GPC (18:1/22:6)*	Phospholipid Metabolism	Lipid	-0.037 (-0.051,-0.022)	7.94E-05	0.00729	0.930
Lipodystrophy	M44877	N-palmitoyl-sphingosine	Sphingolipid Metabolism	Lipid	0.037 (0.022,0.051)	8.16E-05	0.00736	0.387
Lipodystrophy	M33949	gamma-glutamylglycine	Gamma-glutamyl Amino Acid	Peptide	-0.036 (-0.051,-0.022)	8.62E-05	0.00763	0.740
Lipodystrophy	M31932	propionylglycine	Fatty Acid Metabolism (also BCAA Metabolism)	Lipid	-0.04 (-0.056,-0.024)	9.32E-05	0.00811	0.095
Obesity	M46588	X - 12063	Unknown	Unknown	0.036 (0.015,0.058)	0.000129	0.011	0.280
Lipodystrophy	M47154	sphingomyelin (d18:2/14:0, d18:1/14:1)*	Sphingolipid Metabolism	Lipid	-0.03 (-0.053,-0.008)	0.000137	0.0115	0.082
Lipodystrophy	M52707	1-myristoyl-2-linoleoyl-GPC (14:0/18:2)*	Phospholipid Metabolism	Lipid	0.036 (0.011,0.06)	0.000152	0.0124	0.380
Obesity	M15650	N1-methyladenosine	Purine Metabolism, Adenine containing	Nucleotide	0.034 (0.01,0.058)	0.000152	0.0124	0.281
Lipodystrophy	M19258	1-myristoyl-2-palmitoyl-GPC (14:0/16:0)	Phospholipid Metabolism	Lipid	0.035 (0.005,0.065)	0.000193	0.0154	0.035
Lipodystrophy	M19263	1-palmitoyl-2-oleoyl-GPE (16:0/18:1)	Phospholipid Metabolism	Lipid	0.036 (0.006,0.067)	0.000197	0.0155	0.038
Liver	M47888	1-eicosapentaenoyl-GPE (20:5)*	Lysolipid	Lipid	-0.036 (-0.068,-0.004)	0.000211	0.0164	0.904
Lipodystrophy	M36738	gamma-glutamylglutamate	Gamma-glutamyl Amino Acid	Peptide	0.034 (-0.01,0.078)	0.000297	0.0226	0.981
Lipodystrophy	M47872	X - 17340	Unknown	Unknown	0.033 (-0.011,0.078)	0.000302	0.0227	0.319
Liver	M01110	arachidonate (20:4n6)	Polyunsaturated Fatty Acid (n3 and n6)	Lipid	-0.035 (-0.08,0.011)	0.000315	0.0233	0.575
Obesity	M46662	X - 15492	Unknown	Unknown	0.031 (-0.017,0.079)	0.000333	0.0243	0.944
Lipodystrophy	M22137	homoarginine	Urea cycle; Arginine and Proline Metabolism	Amino Acid	0.033 (-0.018,0.084)	0.000365	0.026	0.467
Liver	M33822	1-docosahexaenoyl-GPC (22:6)*	Lysolipid	Lipid	-0.034 (-0.085,0.017)	0.000367	0.026	0.328
Beta Cell	M48491	sphingomyelin (d18:1/20:1, d18:2/20:0)*	Sphingolipid Metabolism	Lipid	-0.032 (-0.084,0.02)	0.000386	0.0266	0.118
Lipodystrophy	M52630	1-(1-enyl-palmitoyl)-2-docosahexaenoyl-GPC (P-16:0/22:6)*	Plasmalogen	Lipid	-0.033 (-0.086,0.019)	0.000386	0.0266	0.219
Beta Cell	M42374	2-aminobutyrate	Methionine, Cysteine, SAM and Taurine Metabolism	Amino Acid	0.034 (-0.02,0.088)	0.000422	0.0275	0.617
Lipodystrophy	M44621	1-(1-enyl-oleoyl)-GPE (P-18:1)*	Lysolipid	Lipid	-0.034 (-0.088,0.02)	0.000411	0.0275	0.720

pPGS	ID	BIOCHEMICAL	SUPER_PATHWAY	SUB_PATHWAY	Beta (95% CI)	pvalue	pvalue.adjFDR	het_pvalue
Lipodystrophy	M52466	1-stearoyl-2-docosahexaenoyl-GPE (18:0/22:6)*	Phospholipid Metabolism	Lipid	0.033 (-0.021,0.087)	0.00042	0.0275	0.446
Obesity	M46260	X - 11564	Unknown	Unknown	0.03 (-0.024,0.084)	0.000419	0.0275	0.162
Beta Cell	M34396	choline phosphate	Phospholipid Metabolism	Lipid	-0.033 (-0.093,0.028)	4.79E-04	3.08E-02	0.952
Lipodystrophy	M52281	2-hydroxybutyrate/2-hydroxyisobutyrate	Methionine, Cysteine, SAM and Taurine Metabolism	Amino Acid	0.033 (-0.027,0.093)	0.000489	0.0308	0.104
Lipodystrophy	M46473	X - 15497	Unknown	Unknown	0.032 (-0.029,0.092)	0.000495	0.0308	0.104
Lipodystrophy	M52447	1-stearoyl-2-arachidonoyl-GPE (18:0/20:4)*	Phospholipid Metabolism	Lipid	0.033 (-0.027,0.093)	0.000496	0.0308	0.384
Lipodystrophy	M47964	X - 13722	Unknown	Unknown	-0.037 (-0.1,0.026)	0.000524	0.0321	0.375
Lipodystrophy	M52467	1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4)*	Phospholipid Metabolism	Lipid	0.033 (-0.03,0.096)	0.000545	0.0321	0.257
Lipodystrophy	M48258	1-oleoyl-GPC (18:1)	Lysolipid	Lipid	-0.033 (-0.096,0.03)	0.000548	0.0321	0.025
Liver	M45453	1-myristoyl-GPC (14:0)	Lysolipid	Lipid	-0.033 (-0.096,0.03)	0.00055	0.0321	0.621
Obesity	M35137	N2,N2-dimethylguanosine	Purine Metabolism, Guanine containing	Nucleotide	0.03 (-0.033,0.093)	0.000546	0.0321	0.214
Liver	M34534	laurylcarnitine	Fatty Acid Metabolism(Acyl Carnitine)	Lipid	-0.033 (-0.098,0.032)	0.000578	0.0333	0.133
Lipodystrophy	M52715	1-(1-enyl-palmitoyl)-2-myristoyl-GPC (P-16:0/14:0)*	Plasmalogen	Lipid	-0.034 (-0.101,0.032)	6.04E-04	0.0339	0.359
Liver	M45951	1-linolenoyl-GPC (18:3)*	Lysolipid	Lipid	-0.033 (-0.1,0.033)	0.000607	0.0339	0.175
Liver	M33969	stearidonate (18:4n3)	Polyunsaturated Fatty Acid (n3 and n6)	Lipid	-0.033 (-0.1,0.033)	0.000612	0.0339	0.837
Liver	M33230	1-palmitoleoyl-GPC (16:1)*	Lysolipid	Lipid	-0.033 (-0.099,0.034)	0.000616	0.0339	0.598
Liver	M03155	3-ureidopropionate	Pyrimidine Metabolism, Uracil containing	Nucleotide	0.032 (-0.039,0.103)	0.000665	0.0362	0.505
Proinsulin	M46624	X - 12798	Unknown	Unknown	-0.033 (-0.105,0.04)	0.000684	0.0369	0.397
Obesity	M35157	N6-carbamoylthreonyladenosine	Purine Metabolism, Adenine containing	Nucleotide	0.03 (-0.044,0.105)	0.000714	0.0381	0.635
Obesity	M43488	N-acetylcarnosine	Dipeptide Derivative	Peptide	0.027 (-0.05,0.103)	0.000738	0.0389	0.251
Lipodystrophy	M47348	X - 22102	Unknown	Unknown	0.034 (-0.044,0.112)	0.000761	0.0397	0.916
Obesity	M46799	oleoyl-linoleoyl-glycerol (18:1/18:2) [2]	Diacylglycerol	Lipid	0.032 (-0.047,0.11)	0.000778	0.0402	0.277
Obesity	M49521	X - 23644	Unknown	Unknown	0.033 (-0.05,0.116)	0.00083	0.0424	0.062
Lipodystrophy	M43330	2-hydroxyibuprofen	Drug	Xenobiotics	0.159 (0.072,0.246)	0.000897	0.0446	0.202
Lipodystrophy	M01649	valine	Leucine, Isoleucine and Valine Metabolism	Amino Acid	0.031 (-0.057,0.118)	0.000899	0.0446	0.117
Obesity	M39600	3-hydroxyhippurate	Benzoate Metabolism	Xenobiotics	0.032 (-0.056,0.119)	0.000895	0.0446	0.298

Supplementary Material: Chapter 5

Supplementary Table S5.1: Proteomics associations with MC3R variants of interest in the Fenland Study with IGF-related proteins P<0.05

Somalogic SeqId	Target	UniProt	rsid	chr	pos	OA	EA	EAF	Effect	SE	Pvalue	N
2771_35	IGFBP-1	P08833	rs3827103	20	54824029	G	A	0.0838	0.0864	0.0231	0.0001896	10708
8469_41	IGFBP-2	P18065	rs3827103	20	54824029	G	A	0.0837	0.0506	0.0244	0.0381	10708
16057_6	IGF-II receptor	P11717	rs61735259	20	54824557	C	A	0.0023	0.2855	0.1432	0.04621	10708

Supplementary Table S5.2: Metabolomics associations with MC3R variants of interest in the EPIC-Norfolk Study
P<0.05

Variant	Metabolite	SUPER_PATHWAY	SUB_PATHWAY	Beta	SE	Pvalue	N
rs143321797	pipecolate	Amino Acid	Lysine Metabolism	1.099234	0.33317	0.000969	9707
rs143321797	4-hydroxyphenylpyruvate	Amino Acid	Tyrosine Metabolism	0.960454	0.317789	0.002509	9711
rs143321797	cortisol	Lipid	Corticosteroids	-0.92738	0.31364	0.003108	9696
rs143321797	cysteine	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism	0.860093	0.328023	0.00874	9711
rs143321797	glucuronate	Carbohydrate	Aminosugar Metabolism	-0.8269	0.317797	0.009269	9645
rs143321797	S-methylcysteine	Amino Acid	Methionine, Cysteine, SAM and Taurine Metabolism	0.857839	0.333021	0.009997	9710
rs143321797	myristoleate (14:1n5)	Lipid	Long Chain Fatty Acid	-0.78935	0.312366	0.011504	9712
rs143321797	arabonate/xylonate	Carbohydrate	Pentose Metabolism	0.723225	0.319036	0.023396	9703
rs143321797	pseudouridine	Nucleotide	Pyrimidine Metabolism, Uracil containing	-0.6681	0.305604	0.028804	9712
rs143321797	indolepropionate	Amino Acid	Tryptophan Metabolism	0.713721	0.329842	0.030478	9648
rs143321797	1-stearoyl-2-oleoyl-GPE (18:0/18:1)	Lipid	Phosphatidylethanolamine (PE)	0.719377	0.339703	0.034203	9682
rs143321797	asparagine	Amino Acid	Alanine and Aspartate Metabolism	0.686494	0.327768	0.03622	9712
rs143321797	glycerol 3-phosphate	Lipid	Glycerolipid Metabolism	0.691107	0.332757	0.03781	9712
rs143321797	1-oleoyl-GPE (18:1)	Lipid	Lysophospholipid	0.603968	0.293732	0.039764	9710
rs143321797	phenylacetate	Amino Acid	Phenylalanine Metabolism	-0.76894	0.374307	0.039947	6292
rs143321797	N-acetyltyrosine	Amino Acid	Tyrosine Metabolism	0.819295	0.398902	0.039988	8630
rs143321797	propionylcarnitine (C3)	Lipid	Fatty Acid Metabolism (also BCAA Metabolism)	0.646528	0.316567	0.041121	9708
rs143321797	palmitoleate (16:1n7)	Lipid	Long Chain Fatty Acid	-0.63272	0.311461	0.042208	9711

rs143321797	prolylglycine	Peptide	Dipeptide	0.665739	0.33608	0.047603	7826
rs3827103	N-acetylcitrulline	Amino Acid	Urea cycle; Arginine and Proline Metabolism	-0.08072	0.031688	0.01086	6682
rs3827103	5alpha-androstan-3alpha,17beta-diol monosulfate (2)	Lipid	Androgenic Steroids	-0.09391	0.037956	0.013354	4207
rs3827103	2-aminoheptanoate	Lipid	Fatty Acid, Amino	-0.06624	0.027345	0.015418	8767
rs3827103	isoursodeoxycholate	Lipid	Secondary Bile Acid Metabolism	0.063969	0.026554	0.015996	9610
rs3827103	alpha-hydroxyisovalerate	Amino Acid	Leucine, Isoleucine and Valine Metabolism	0.059308	0.025002	0.017688	9712
rs3827103	N2,N2-dimethylguanosine	Nucleotide	Purine Metabolism, Guanine containing	-0.05592	0.024098	0.020305	9705
rs3827103	succinylcarnitine (C4-DC)	Energy	TCA Cycle	-0.05571	0.024522	0.023099	9692
rs3827103	glutamine	Amino Acid	Glutamate Metabolism	-0.05736	0.026154	0.028307	9712
rs3827103	maltose	Carbohydrate	Glycogen Metabolism	-0.06328	0.028959	0.028865	7688
rs3827103	N-acetylarginine	Amino Acid	Urea cycle; Arginine and Proline Metabolism	-0.05731	0.026264	0.029111	9671
rs3827103	N-acetylglutamine	Amino Acid	Glutamate Metabolism	-0.05638	0.026849	0.035737	9002
rs3827103	N1-methyladenosine	Nucleotide	Purine Metabolism, Adenine containing	-0.05149	0.024833	0.038119	9692
rs3827103	N-acetylaspartate (NAA)	Amino Acid	Alanine and Aspartate Metabolism	-0.05273	0.025496	0.038645	9444
rs3827103	1-(1-enyl-palmitoyl)-2-palmitoleoyl-GPC (P-16:0/16:1)*	Lipid	Plasmalogen	0.052581	0.025686	0.040654	9563
rs3827103	N-acetyl-3-methylhistidine*	Amino Acid	Histidine Metabolism	-0.06934	0.034523	0.044596	5516
rs3827103	gamma-glutamylglutamate	Peptide	Gamma-glutamyl Amino Acid	-0.05123	0.025922	0.048137	9639
rs61735259	N-palmitoyl-sphinganine (d18:0/16:0)	Lipid	Sphingolipid Metabolism	0.464424	0.152988	0.0024	8846

rs61735259	7-ketodeoxycholate	Lipid	Secondary Bile Acid Metabolism	0.743372	0.284821	0.009055	3082
rs61735259	1-linoleoyl-2-linolenoyl-GPC (18:2/18:3)*	Lipid	Phosphatidylcholine (PC)	0.416668	0.162	0.010111	9576
rs61735259	etiocholanolone glucuronide	Lipid	Androgenic Steroids	0.384609	0.152875	0.011875	9435
rs61735259	glycocholenate sulfate*	Lipid	Secondary Bile Acid Metabolism	0.399282	0.158727	0.011886	9711
rs61735259	1-linolenoylglycerol (18:3)	Lipid	Monoacylglycerol	0.389817	0.160154	0.014933	9336
rs61735259	citraconate/glutaconate	Energy	TCA Cycle	0.361703	0.159414	0.023271	9710
rs61735259	mannitol/sorbitol	Carbohydrate	Fructose, Mannose and Galactose Metabolism	0.361818	0.159498	0.0233	9710
rs61735259	1-palmitoleoylglycerol (16:1)*	Lipid	Monoacylglycerol	0.359334	0.160164	0.024862	9532
rs61735259	linolenate [alpha or gamma; (18:3n3 or 6)]	Lipid	Polyunsaturated Fatty Acid (n3 and n6)	0.342182	0.153905	0.026193	9711
rs61735259	cysteinylglycine	Amino Acid	Glutathione Metabolism	0.362542	0.163082	0.026211	8897
rs61735259	maleate	Lipid	Fatty Acid, Dicarboxylate	0.358728	0.16335	0.028087	8803
rs61735259	fructose	Carbohydrate	Fructose, Mannose and Galactose Metabolism	-0.3412	0.159039	0.031919	9712
rs61735259	indolepropionate	Amino Acid	Tryptophan Metabolism	-0.34538	0.161239	0.032189	9648
rs61735259	imidazole propionate	Amino Acid	Histidine Metabolism	0.333378	0.15734	0.034104	8867
rs61735259	2-palmitoleoylglycerol (16:1)*	Lipid	Monoacylglycerol	0.419203	0.198064	0.034302	6649
rs61735259	1-stearoyl-GPE (18:0)	Lipid	Lysophospholipid	0.329782	0.1574	0.036155	9712
rs61735259	gamma-glutamylglycine	Peptide	Gamma-glutamyl Amino Acid	-0.32763	0.157188	0.037129	9707
rs61735259	1-myristoylglycerol (14:0)	Lipid	Monoacylglycerol	0.331785	0.159346	0.037327	9461
rs61735259	1-linolenoyl-GPC (18:3)*	Lipid	Lysophospholipid	0.331726	0.159362	0.03738	9697
rs61735259	deoxycholate	Lipid	Secondary Bile Acid Metabolism	0.377292	0.182209	0.038391	8146

rs61735259	1-stearoyl-2-docosapentaenoyl-GPC (18:0/22:5n3)*	Lipid	Phosphatidylcholine (PC)	0.331849	0.161548	0.039957	9711
rs61735259	palmitoyl-linoleoyl-glycerol (16:0/18:2) [2]*	Lipid	Diacylglycerol	0.322776	0.157609	0.040564	9553
rs61735259	spermidine	Amino Acid	Polyamine Metabolism	0.381149	0.186875	0.041391	6262
rs61735259	1-palmitoyl-2-arachidonoyl-GPI (16:0/20:4)*	Lipid	Phosphatidylinositol (PI)	0.32222	0.159425	0.043266	9609
rs61735259	cholesterol	Lipid	Sterol	0.321197	0.159032	0.043415	9707
rs61735259	1-stearoyl-2-dihomo-linolenoyl-GPC (18:0/20:3n3 or 6)*	Lipid	Phosphatidylcholine (PC)	0.325956	0.161596	0.043684	9711
rs61735259	glycosyl-N-stearoyl-sphingosine (d18:1/18:0)	Lipid	Ceramides	0.338978	0.16851	0.04426	8856
rs61735259	N-palmitoyl-sphingosine (d18:1/16:0)	Lipid	Ceramides	0.309231	0.156743	0.048513	9709
rs61735259	2,3-dihydroxy-2-methylbutyrate	Amino Acid	Leucine, Isoleucine and Valine Metabolism	0.322019	0.163874	0.049409	8969

Analysis Plans

Analysis Plan below was used in the running of GWAS for Insulin Fold Change and Modified Stumvoll ISI and is as shared with study-level data analysts.

Information has been restricted to that for ISI and IFC only, and information with regards to data upload/contacts have been removed.

Refined Insulin Traits Imputation and Analysis Plan
Association Analysis Stage
Updated November 2019

Overview

There are four major components to this analysis plan:
Genome-wide genotypes must be on the correct build (37/hg19) and correct strand (forward).
Imputation of genotypes is performed either using the Haplotype Reference Consortium (for samples with European-ancestry) or 1KG phase 3 (for samples with non-European ancestry)
Association analysis is conducted using specific software tools that will provide all necessary summary statistics for flexible central meta-analysis.
Results are uploaded to the designated University of Michigan sftp account

We anticipate that most participating studies have performed Steps 1-2 and will be ready for Step 3, association analysis. Therefore, we only briefly discuss Steps 1-2 below. If alignment to the correct build or imputation has not yet been performed on your data, please see the accompanying document, “Refined Insulin – Data Cleaning and Imputation.docx”. Please note that we are requesting data for the X Chromosome; instructions on imputing ChrX is also the accompanying document.

1. Genotypes & Pre-Imputation QC

All studies must have some version of a genome-wide array, for example with >200,000 genome-wide common variants. Genotypes must be on GRCh37/hg19.

QC should be done separately for: individual studies, major continental ancestries within a study, and samples of the same study that were genotyped on different arrays. Additional QC will be conducted centrally at the meta-analysis stage.

2. Imputation

We recommend that European cohorts impute using HRC, while non-European cohorts use 1KG. We also recommend you use an imputation server, such as the Michigan Imputation Server, if possible. We intend to include the X chromosome in this analysis; if this has not yet been performed for your study, please see detailed instructions for more information on X-chromosome-specific QC and imputation.

3. Phenotype definition, covariate adjustment, and trait transformation

We are requesting analysis of four outcomes, listed below. For each model, we are requesting analyses be run both with and without BMI adjustment (in addition to other covariates), for a total of eight models. Details of all models are described in this section.

Modified Stumvoll Insulin Sensitivity Index (ISI), inverse normally transformed
Insulin fold change (pmol/l), inverse normally transformed

Covariate adjustment prior to regression analysis
We request that you do NOT impute for missing traits or covariates**.**
Please perform modeling separated by race/ethnicity/ancestry
We strongly recommend that association analyses adjust for population structure where possible. The methods used will vary from cohort to cohort, and include (but are not limited to): estimation of principal components and including these as covariates, adjustment for covariates that reflect birth origin, or by estimation of a kinship matrix, which is included in the final regression analysis model.

To control type I error rate of low frequency variants and to fully remove trait-covariate correlations, covariate adjustment will be performed in two steps. Failure to include the second stage of covariate adjustment can lead to undesired reintroduction of correlation between trait and covariates.

Model trait ~ all covariates in model, then inverse normal transform (INT) residuals
Model INT(residuals from step 1) ~ all covariates in model

The residuals from step 2 will be used in the final regression analysis.

Regression Analysis
Please use an additive genetic model for all analyses. Test each polymorphic variant for association with the traits where available. Do not adjust analysis for genomic control. Do not exclude variants based on imputation quality. Further QC will be performed centrally.
Analysis Details by Phenotypes:

Modified Stumvoll Insulin Sensitivity Index (ISI) 2 models tested
Trait is modified Stumvoll Insulin Sensitivity Index, controlled for the covariates listed below and modeling on inverse normal transformed residuals.

Exclude all individuals with Diabetes (T2D, T1D) from data sets ('diagnosed', on diabetes treatment (oral and insulin), and/or fasting glucose ≥7 mmol/L, 2hr ≥11.1mmol/L, HbA1c ≥6.5%)
Studies should apply their 'usual' adjustments (eg. age2, cohort/site), even if not listed below.
Model Modified Stumvoll Insulin Sensitivity Index (ISI):
Stumvoll ISI = 0.156 - (0.0000459 x $PI_{120[pmol/L]}$)
- (0.000321 x $PI_{0[pmol/L]}$)
- (0.0054 x $PG_{120[mmol/L]}$)
Adjusted for age, sex, population structure
Adjusted for age, sex, population structure, and BMI

Insulin Fold Change (IFC) 2 models tested (updated)
Trait natural log (nlog) transformed fold change in insulin, controlled for the covariates listed below and modeling on inverse normal transformed residuals.
Plasma insulin is in pmol/L (Check units: (1pmol/L=6 mg/dL))
Exclude all individuals with Diabetes (T2D, T1D) from data sets ('diagnosed', on diabetes treatment (oral and insulin), and/or fasting glucose ≥7 mmol/L, 2hr ≥11.1mmol/L, HbA1c ≥6.5%)
Studies should apply their 'usual' adjustments (eg. age2, cohort/site), even if not listed below.
Insulin fold change (IFC) = ln($PI_{120[pmol/L]}$ / $PI_{0[pmol/L]}$)
I. Adjusted for age, sex, population structure
II. Adjusted for age, sex, population structure, and BMI

4. Sharing Results

Cohort information spreadsheet
Please complete the following summary spreadsheet:

Results Formatting:
The minimum level of information required for each SNP is: SNP name, chromosome number, position, effect and non-effect allele, effect allele frequency, effect allele count, number of samples, if the SNP was imputed or not, imputation quality, HWE p-value for genotyped SNPs, genotyping call rate, and regression results (beta, SE, and p-value). See the bottom of this document for an example table listing all requested fields, data formatting, and field naming.
Please submit one compressed result file (eg. gz)containing the results for all SNPs.
Results should be reported in a tab delimited text file, with one row per SNP. The first row is a header. Please also provide a file that summarizes column naming convention for your study.
Missing or unavailable data should be denoted using the period character (“.”).
Please do not filter variants on the basis of minor allele frequency or imputation quality. Additional QC will be performed centrally, prior to meta-analysis.

For SNP names and alleles, please use the name and allele directly from the imputation program. Please do not recode these names and alleles. We will work with them centrally for the meta-analysis.

File naming convention:

Please name your results file as follows:

TRAIT_STUDY_ANCESTRY_SOFTWARE__IMPUTATION_DATE_INITIAL.txt

In this filename:

TRAIT is the trait being analyzed. Please add "_adjBMI" if the model has adjusted for BMI, and "_noBMI" if the model was not adjusted for BMI.
Modified Stumvoll Insulin Sensitivity Index (ISI): ISI_adjBMI, ISI_noBMI
Insulin fold change: IFC_adjBMI and IFC_noBMI
STUDY is replaced by a unique cohort identifier for your cohort.
ANCESTRY is replaced using the following abbreviations, based on the 1000 Genomes superpopulations: EUR (European), AFR (African), AMR (Ad Mixed American), SAS (South Asian), EAS (East Asian)
SOFTWARE should be replaced by the analysis software used for association testing.
IMPUTATION should be replaced by the software used for imputation.
DATE is replaced by the date of file generation (YYYYMMDD format).
INITIAL is replaced by the initials of the analyst submitting the results file.

Example file name: PROI_adjBMI_METSIM_EUR_EPACTS_MINIMAC_20190402_KAB.txt

Descriptions of Required Results Fields:

Column Header	Description	Format	Examples
SNPID	SNP ID by chromosome number and position	As provided by analysis software	16:50345708:G_T
RSID	dbSNP rsID label (if easily available)	rs number from dbSNP, or set to missing if no dbSNP number exists	rs693
CHR	Chromosome on which SNP resides	Numeric for chromosomes 1-22, X for ChrX	1
POS	Position of SNP on chromosome	Base pairs on human genome build used	34000345
EFFECT_ALLELE	Allele for which the effect (BETA) is reported	As provided by analysis software	A or eg. I for indel
NON_EFFECT_ALLELE	Other allele at this site	As provided by analysis software	G or eg. D for indel
N	Total number of samples analyzed at this site	Numeric, integer	1243
EAF	Allele frequency of the EFFECT_ALLELE	Frequency (0,1)	0.3708
HWE_P	Exact HWE p-value for the samples analyzed, **only for genotyped SNPs**	Scientific E notation (set to missing if imputed)	1.1234E-02
CALL_RATE	Genotyping call rate for the SNP (N_successful_genotypes / N_attempted_genotypes)	Frequency (0,1]; (set to missing if imputed)	0.9936
BETA	Estimate of the effect size	Numeric float	0.2036 .
SE	Estimated standard error on the estimate of the effect size, uncorrected for genomic control	Numeric float	0.5611 .
P_VAL	Significance of the variant association, uncorrected for genomic control	Scientific E notation	3.244E-04
IMPUTED	Is the SNP imputed?	0 = Genotyped 1 = Imputed	1
INFO_TYPE	Type of information provided in the INFO column	0 = SNP is genotyped, 1 = "r2_Hat", e.g. RSQR column in MACH2DAT output 2 = "info" from SNPTEST 3 = "INFO" from PLINK 4 = Information from other imputation software	1
INFO	Measure of information content for the imputed SNP result (range 0-1)	Numeric float	0.4832