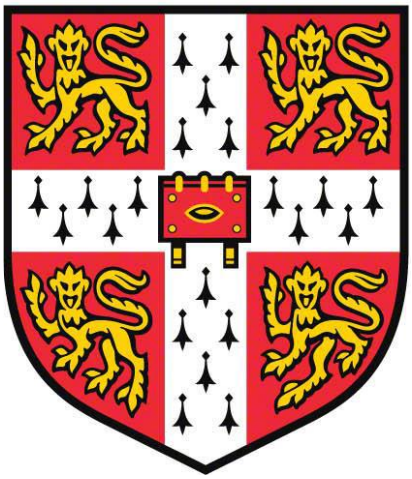
Meat consumption and type 2 diabetes: investigating heterogeneity and potential causal mechanisms



Chunxiao Li

Downing College

MRC Epidemiology Unit

University of Cambridge

November 2022

This thesis is submitted for the degree of Doctor of Philosophy

Preface

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 at the University of Cambridge or any other universities and research institutions. As per requirement by the School of Clinical Medicine and Clinical Veterinary Medicine Degree Committee, this thesis does not exceed the prescribed word limit of 60,000 words, exclusive of figures, tables, supplementary materials and references.

The projects contained in this thesis were under the supervision of Professor Nicholas J. Wareham and Dr Fumiaki Imamura and were conducted at the MRC Epidemiology Unit at the University of Cambridge. Professor Nicholas J. Wareham provided guidance on the direction and oversight throughout the work. I proposed analysis plans in collaboration with my supervisors and conducted data processing and data analyses independently with support from my collaborators. My supervisors provided advice for these studies, contributed to interpreting results, revising manuscripts and providing feedbacks.

Others contributed to this work and their contributions are mentioned below.

The EPIC-Norfolk study is coordinated by the MRC Epidemiology Unit led by Professor Nicholas J. Wareham. Abigail Britten and Nicolas Kimber managed and provided the EPIC-Norfolk data (Chapter 2, 3 and 4). Nicola Kerrison supported the data application process. Angela Mulligan contributed to the management of the EPIC-Norfolk dietary data. Metabolomics profiling in the EPIC-Norfolk study was performed by Metabolon under the responsibility of Dr Claudia Langenberg and Professor Nicholas J. Wareham (Chapters 3 and 4). Dr Isobel D. Stewart processed the raw data and conducted initial quality control for the metabolites in the EPIC-Norfolk study. The laboratory analyses of the intervention study in Chapter 3 were conducted by Drs Roland Wedekind and Augustin Scalbert at the International Agency for Research on Cancer.

The EPIC-InterAct project was coordinated by Professor Nicholas J. Wareham, Dr Claudia Langenberg, Professor Nita G. Forouhi and Stephen Sharp (Chapters 5 and 6). Primary data management and cleaning, including quality control of phenotypic and genetic data, genotyping data processing and imputation were performed by Nicola Kerrison and Jian'an Luan. UK Biobank data were managed by Nicola Kerrison under the responsibility of Dr Claudia Langenberg and Professor Nicholas J. Wareham. The InterConenct project is led by Professors Nicholas J. Wareham and Nita G. Forouhi.

Dr Isobel D. Stewart and Dr Maik Pietzner guided the analysis for metabolomics data in the EPIC-Norfolk study (Chapter 3). I performed the genome-wide association study in EPIC-Norfolk guided by Jian'an Luan and Dr Isobel D. Stewart (Chapter 4). Ascertainment of diabetes cases in UK Biobank (UKBB) was conducted by Nicola Kerrison and Dr Rupal Shah (Chapters 5 and 6). Dr Eleanor Wheeler constructed genetic risk scores (GRSs) in UKBB (Chapter 5). I constructed GRSs in the EPIC-InterAct study guided by Dr Eleanor Wheeler and Jian'an Luan (Chapter 5). Dr Fumiaki Imamura and Stephen Sharp provided statistical advice on calculating the absolute risk of diabetes in the EPIC-InterAct study (Chapter 5). Tom Bishop provided technical support in server set-up, variable harmonisation and data analyses using DataSHIELD in the InterConnect project (Chapter 6). Dr Jean-Philippe Drouin-Chartier performed association analysis of meat consumption with T2D risk in the Health Professionals Follow-Up Study, Nurses' Health Study (NHS) I and NHS II following the analysis plan; Drs Maria Kakkoura and Huaidong Du performed the analysis in the China Kadoorie Biobank study. They provided summary statistics for me to perform meta-analysis in Chapter 6.

November 2022

Chunxiao Li

Summary

Meat consumption and type 2 diabetes: investigating heterogeneity and potential causal mechanisms

Chunxiao Li

Type 2 diabetes (T2D) is a complex metabolic disease which affects more than 500 million people worldwide, imposing enormous burdens on affected individuals and their families, healthcare systems and societies. Healthy diets play a crucial role in preventing T2D and meat has been reported to be associated with an increased risk of T2D. However, it is unclear whether different types of meat are all associated with increased risk and whether associations are the same in all individuals and in all populations. Finally, there are uncertainties about the causal nature of the association and the mechanisms that may underlie it. In my PhD, I aimed to investigate these uncertainties in analyses of epidemiological studies.

The initial elements of my work focused on refining measures of the exposure (meat intake) and the outcome (incident T2D) in the EPIC-Norfolk study, a population-based cohort study of over 25,000 participants. I worked on improving case ascertainment of T2D as the primary outcome in my analyses. I updated T2D case ascertainment in EPIC-Norfolk by linkage of multiple external data sources, including diabetic eye screening data and clinical biochemistry data. I identified over 2,000 additional incident diabetes cases. I then reported the association of self-reported intake of different types of meat with T2D in EPIC-Norfolk.

Dietary biomarkers can provide complementary information about diet-disease associations. I used untargeted metabolomics profiling to derive metabolite scores to quantify the consumption of red meat, processed meat and poultry based on 781 circulating metabolites and 7-day diet diary data in 11,432 participants in EPIC-Norfolk. The best performing score was for red meat, comprising 139 metabolites and accounting for 17% of the explained variance of red meat consumption. Eleven top-ranking metabolites that were included in the

red meat score were validated in a trial conducted by collaborators in Lyon, France. These metabolites were mainly classified into groups of lipids, amino acids, and xenobiotics, such as plasmalogens, trimethylamine N-oxide, and stearoylcarnitine. I then showed that this red meat metabolite score was strongly associated with T2D incidence in EPIC-Norfolk.

I then investigated the potential causal roles of these eleven red meat-related metabolites in T2D incidence by conducting Mendelian randomisation (MR) analyses. I observed weak evidence of possible causal associations between meat-related metabolites and incident T2D, possibly due to limited power and weak genetic instruments.

In an analysis in two large studies (EPIC-InterAct and UK Biobank), I evaluated whether the association between meat consumption and T2D incidence differed in sub-populations with varying genetic and clinical baseline risks. I found that meat intake was associated with incident T2D independently of genetic and clinical predisposition to T2D. This suggests that there are benefits of reducing meat intake on T2D burden in the entire population.

Finally, I examined associations between types of meat intake (red meat, processed meat and poultry) and T2D risk based on a federated platform in the InterConnect, which enabled harmonised data analysis of 1.5 million individuals from 23 studies across the world. This meta-analysis of individual participant data provided unique evidence of meat-T2D associations in previously understudied populations, such as those in East Asia, and East Mediterranean. I included over 60,000 new-onset T2D cases with a median of 13 years of follow-up showing that consumption of red meat, processed meat and poultry were each individually associated with increased risk of T2D.

In summary, my work provides strong evidence on the consistency of the association of meat consumption with T2D risk in sub-groups within European populations and also across heterogeneous populations worldwide. This has implications for public health approaches to T2D prevention.

Acknowledgements

Foremost, I would like to express my deep gratitude to my primary supervisor, Professor Nick Wareham, for his invaluable expertise and support. This thesis would not have been possible without his generous guidance and help. He was always patient with me, gave me the confidence to explore and offered me great opportunities over the years. I give my special thanks to my supervisor Dr Fumiaki Imamura, for his patient supervision throughout my PhD. He guided me to develop critical thinking and improve my research skills which are important to my career development.

I would like to thank all the members of the Diabetes Aetiology team, Claudia, Maik, Isobel, Ellie, Jian'an, Nicola, Luca, Erin, Mesut, Lina, Nick, Vicky, Chen, Yaling and Heba, members of the Nutrition Epidemiology group, Nita, Daniel, Solomen, Yanaina, Jusheng, Jun, Eirini, Bridget, Gabi and Marinka, and members of the Genetics Forum group, Felix, John and Ken, for being my inspirations in research. I am so grateful for their help to address all of my questions and provide insightful feedbacks to my work. I am also grateful to the following people for providing their unique expertise and help for my PhD, Tom, Matthew, Soumya, Stephen, Angela, Linda and Albert. Thank you also to Meriel and Ligia for their tremendous support and to all the colleagues in the Unit who have supported me throughout my PhD.

I also owe special thanks to Nita, for her continuous support throughout my PhD. She always patiently guided me on research communications and provide thorough and constructive advice to my queries, all of which inspired me to become an independent and passionate researcher. I would also like to thank my PhD fellows within the unit who shared the journey with me, especially Dudan, Zhe, Yajie, Alice, Julia, Mine, Jakub, Camille, Dolly, Tuckseng, Rebecca, Yuru, Raina, Divya, Shayan, Lorenzo and Anwesha for being a supportive community that I am lucky to be a part of.

Finally, I would like to thank my family for shaping the person I am today, and to my husband, Junqing, for your encouragement and support as always. I am truly grateful to go through this PhD journey with your company.

Publications and presentations

Publications related to this thesis

Li C, Imamura F, Wedekind R, Stewart ID, Pietzner M, Wheeler E, Forouhi NG, Langenberg C, Scalbert A, Wareham NJ. Development and validation of a metabolite score for red meat intake: an observational cohort study and randomised controlled dietary intervention. Am J Clin Nutr. 2022 Aug 4;116(2):511-522. doi: 10.1093/ajcn/nqac094. PMID: 35754192; PMCID: PMC9348983.

Li C, Imamura F, [other co-authors to be determined], Wareham NJ. Association between meat consumption and incident type 2 diabetes ascertained by multiple data sources: food diary vs. FFQ measurements in EPIC-Norfolk. (*Manuscript in preparation, work in Chapter 2*)

Li C, Imamura F, [other co-authors to be determined], Wareham NJ. Differential impacts of meat consumption on type 2 diabetes risk in population subgroups: in EPIC-InterAct and UK Biobank. (*Manuscript in preparation, work in Chapter 5*)

Li C, Imamura F, [other co-authors to be determined], Wareham NJ. Associations of meat consumption with incident type 2 diabetes across populations worldwide: a federated meta-analysis in InterConnect. (*Manuscript in preparation, work in Chapter 6*)

Other publications

Strain T, [6 authors], **Li C**, [4 authors]. Quantifying the relationship between physical activity energy expenditure and incident Type 2 Diabetes: a prospective cohort study of device-measured activity in 90,096 adults. (*Under review by the Journal*)

Presentations

Ascertainment of diabetes cases using survey data and linked electronic health records in the EPIC-Norfolk study (*The EPIC-Norfolk Participant Advisory Panel meeting, March 2022*)

Investigating the relationship between meat intake and type 2 diabetes *(EPIC Scientific Day, Oct 2022)*

Table of Contents

Preface 2

Summary 4

Acknowledgements 6

Publications and presentations 7

Table of Contents 8

Index of Figures 14

Index of Tables 17

Abbreviations 19

Chapter 1 22

Introduction 22

1.1 General introduction 23

1.2 Diabetes: definition, epidemiology, aetiology, and prevention 25

1.2.1 Definition and diagnosis of diabetes 25

1.2.2 Epidemiology of diabetes 26

1.2.3 Aetiology of diabetes 27

1.2.4 Type 2 diabetes prevention 28

1.3 Role of meat in T2D incidence 29

1.3.1 Meat consumption trends 29

1.3.2 Associations between types of meat and T2D 30

1.3.3 Meat and T2D association in different subgroups and across populations 31

1.3.4 Meat consumption and T2D risk factors in RCTs 35

1.3.5 Biological mechanisms insights 35

1.4 Main Studies involved in this thesis 38

1.4.1 EPIC Norfolk Study 38

1.4.2 EPIC-InterAct Study 39

1.4.3 UK Biobank 41

1.4.4 InterConnect 43

1.5 Main methods 45

1.5.1 Metabolomics 45

1.5.2 Mendelian randomisation analysis 45

1.5.3 Federated meta-analysis in InterConnect 47

1.6 Thesis objectives and outlines 49

Chapter 2 50

Association between meat consumption and incident type 2 diabetes ascertained by multiple data sources: food diary vs. FFQ measurements in EPIC-Norfolk 50

Abstract 51

2.1 Introduction 52

2.2 Methods 54

2.2.1 Study population 54

2.2.2 Dietary assessment 54

2.2.3 Ascertainment of T2D cases 54

2.2.4 Assessment of covariates 56

2.2.5 Statistical analysis 56

2.3 Results 58

2.3.1 Baseline characteristics 58

2.3.2 Association of red meat consumption with incident T2D 59

2.4 Discussion 61

Chapter 3 64

Development and validation of a metabolite score for red meat intake: an observational cohort study and a randomised controlled dietary intervention 64

Abstract 66

3.1 Introduction 67

3.2 Methods 68

3.2.1 Data source and study design 68

3.2.2 Observational data for the derivation and validation of the metabolite scores: the EPIC-Norfolk study 69

3.2.3 Randomised controlled trial of meat consumption 70

3.2.4 Prospective cohort analysis of the association of the red meat metabolite score with incident disease outcomes in the EPIC-Norfolk study 71

3.3 Results 74

3.3.1 Baseline characteristics and meat consumption of study participants in the EPIC-Norfolk study 74

3.3.2 Development and validation of metabolite scores for meat consumption 76

3.3.3 Associations of metabolites in the red meat metabolite score with meat intake in an RCT 79

3.3.4 Association of the red meat metabolite score with T2D 82

3.3.5 Association of the red meat metabolite score with other health outcomes 84

3.4. Discussion....86

Chapter 4 90

Investigation of causal associations between meat-related metabolites and type 2 diabetes 90

Abstract....91

4.1 Introduction....92

4.2 Methods....93

4.2.1 Observational association analyses....93

4.2.2 Mendelian randomisation analyses....93

4.3 Results....96

4.3.1 Observational associations....96

4.3.2 Identification of genetic instruments for metabolites....97

4.3.3 MR associations....100

4.4 Discussion....102

Chapter 5 106

Differential impacts of meat consumption on type 2 diabetes risk in population subgroups: in EPIC-InterAct and UK Biobank 106

Abstract....107

5.1 Introduction....108

5.2 Methods....109

5.2.1 Study populations....109

5.2.2 Dietary assessment....109

5.2.3 Genotyping and polygenic risk scores....110

5.2.4 HbA1c and Cambridge diabetes risk score 110

5.2.5 Outcome ascertainment 111

5.2.6 Assessment of covariates 111

5.2.7 Statistical analysis 111

5.3 Results 113

5.3.1 Associations with incident T2D 115

5.3.2 Absolute risk of meat intake with T2D by groups 117

5.3.3 Joint associations of meat intake and genetic or clinical risks with incident T2D 119

5.4 Discussion 121

Chapter 6 124

Associations of meat consumption with incident type 2 diabetes in 1.5 million participants from 23 studies: a federated meta-analysis in InterConnect 124

Abstract 125

6.1 Introduction 126

6.2 Methods 127

6.2.1 Study design and participants 127

6.2.2 Dietary assessment 129

6.2.3 Incident T2D ascertainment 129

6.2.4 Covariates 130

6.2.5 Statistical analysis 130

6.3 Results 132

6.4 Discussion 142

Chapter 7 146

General Discussion 146

7.1 Summary of findings 147

7.2 Strengths 148

7.3 Limitations 149

7.3.1 Measurement of dietary information 149

7.3.2 Metabolomics in observational vs. interventional studies 150

7.3.3 Causal inference of the metabolite-T2D association 151

7.3.4 Generalisability of findings 151

7.4 Implications and future perspectives 152

7.4.1 Utilisation of different dietary assessment approaches to facilitate nutritional research 152

7.4.2 EHRs linkage for the ascertainment of T2D 153

7.4.3 Potential mechanisms of the meat-metabolite-T2D association 154

7.4.4 Dietary recommendations for the general population 154

Conclusions 155

References 156

Supplementary Information 178

Supplementary Figures 184

Supplementary Tables 200

Index of Figures

Figure 1.1 Modified diagnostic criteria for diabetes by the International Diabetes Federation (IDF). **P25**

Figure 1.2 Age-adjusted comparative prevalence of diabetes in adults aged 20-79 years in 2021, reported by International Diabetes Federation. **P27**

Figure 1.3 Trends of meat consumption in different regions in million metric tons from the 1960s to 2010s. **P29**

Figure 1.4 Red and processed meat consumption across regions, reported by GBD 2017 Diet Collaborators. **P30**

Figure 1.5 Possible mechanisms linking red and processed meat metabolites to the aetiology of type 2 diabetes. **P37**

Figure 1.6 Existing approaches to perform cross-cohort analyses and their limitations. **P44**

Figure 1.7 The framework of how federated meta-analysis was performed using the DataSHIELD. **P47**

Figure 3.1 Flow chart for the overall analytic approach for development and validation of the meat metabolomics score. **P68**

Figure 3.2 Coefficients of metabolites with self-reported red and processed meat and poultry intake: the EPIC-Norfolk study (n=11,432). **P77**

Figure 3.3 Volcano plot of candidate metabolites for red meat intake (n=139) with self-reported red meat intake and comparison of the red meat metabolite score across different categories of meat consumer groups: the EPIC Norfolk study (n=11,432). **P78**

Figure 3.4 The associations of the red meat metabolite score and self-reported red meat intake with incident type 2 diabetes in a nested case-cohort study and exploratory analyses of multiple other health outcomes in the EPIC-Norfolk study. **P84**

Figure 5.1 Relative risk (adjusted hazard ratios) for T2D in subpopulations defined by meat

intake and other risk factors in the EPIC-InterAct (n=20,628) and the UK Biobank (n=316,222) studies. **P116**

Figure 5.2 Adjusted absolute risk differences in the EPIC-InterAct (n=20,628) and the UK Biobank (n=316,222) studies. **P118**

Figure 5.3 Joint effects of meat intake and genetic or clinical risk on the risk of incident T2D in the EPIC-InterAct (n=20,628) and the UK Biobank (n=316,222) studies and in meta-analysis. **P120**

Figure 6.1. Map of the distribution of included studies (n=53). **P128**

Figure 6.2 Hazard ratios and 95% confidence intervals (CIs) for association between red meat consumption (per 100 g/d) and incident type 2 diabetes (primary outcome) in the InterConnect project. **P138**

Figure 6.3 Hazard ratios and 95% confidence intervals (CIs) for association between processed meat consumption (per 50 g/d) and incident type 2 diabetes (primary outcome) in the InterConnect project. **P139**

Figure 6.4 Hazard ratios and 95% confidence intervals (CIs) for association between poultry consumption (per 100 g/d) and incident type 2 diabetes (primary outcome) in the InterConnect project. **P140**

Supplementary Figure 2.1 The time line of multiple data sources used for T2D case ascertainment in the EPIC-Norfolk study. **P184**

Supplementary Figure 2.2 Flow chart of inclusion of participants through the study. **P185**

Supplementary Figure 2.3 The intersection of different data sources to ascertain incident type 2 diabetes cases in EPIC-Norfolk after exclusion of prevalent diabetes. **P186**

Supplementary Figure 3.1 Flowchart for identification of metabolites that make up the red meat metabolite score in the trial. **P187**

Supplementary Figure 3.2 The correlations between meat scores and 7-day diet diary (7dDD) measured meat intake. **P188**

Supplementary Figure 3.3 Plasma levels of selected metabolites after consumption of pork and tofu in the randomised cross-over trial. **P189**

Supplementary Figure 3.4 Chromatographic tracing of selected metabolites after consumption of pork vs. tofu in the intervention study. **P190**

Supplementary Figure 3.5 Heatmap of correlations between types of meat consumption and top-ranked metabolites (n=11) in the red meat metabolite score that validated in the intervention study: EPIC-Norfolk study (n=11,432). **P192**

Supplementary Figure 4.1 Plots of Mendelian randomisation (MR) sensitivity analyses of deoxycarnitine on T2D. **P193**

Supplementary Figure 4.2 Funnel plot of Mendelian randomisation (MR) sensitivity analysis of 1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4) on T2D. **P194**

Supplementary Figure 5.1 Unadjusted cumulative incidence of T2D over 10 years in subgroups by meat intake and genetic or clinical risk factors in the EPIC-InterAct (n=20,628) study. **P195**

Supplementary Figure 5.2 Unadjusted cumulative incidence of T2D over 10 years in subgroups by meat intake and genetic or clinical risk factors in the UK Biobank (n=316,222) study. **P196**

Supplementary Figure 6.1 Hazard ratios and 95% confidence intervals (CIs) for association between red meat consumption (per 100 g/d) and incident type 2 diabetes (secondary outcome) in the InterConnect project. **P197**

Supplementary Figure 6.2 Hazard ratios and 95% confidence intervals (CIs) for association between processed meat consumption (per 50 g/d) and incident type 2 diabetes (secondary outcome) in the InterConnect project. **P198**

Supplementary Figure 6.3 Hazard ratios and 95% confidence intervals (CIs) for association between poultry consumption (per 100 g/d) and incident type 2 diabetes (secondary outcome) in the InterConnect project. **P199**

Index of Tables

Table 1.1 Summarised information about the published meta-analysis studies with their involved individual cohorts. **P33**

Table 2.1 Baseline characteristics of the study participants by tertiles of red meat consumption and in total in the EPIC-Norfolk study (n=24,464). **P59**

Table 2.2 Association between types of meat consumption and incident T2D measured by 7-day diet diary (7dDD) or food frequency questionnaire (FFQ): the EPIC-Norfolk Study. **P60**

Table 3.1 Baseline characteristics of the study participants for development and validation of meat metabolite scores in the EPIC-Norfolk study. **P75**

Table 3.2 Metabolites from the red meat metabolomics score that were positively associated with red meat consumption in both the EPIC-Norfolk and the randomised cross-over trial. **P80**

Table 3.3 Characteristics of the study participants from baseline of the T2D case-cohort in the EPIC-Norfolk cohort. **P83**

Table 4.1 The associations of the red meat related metabolites with incident type 2 diabetes in the EPIC-Norfolk study (n=11,432). **P97**

Table 4.2 Summary statistics for the genetic instrumental variants of the meat related metabolites and their association with T2D. **P99**

Table 4.3 Causal effects of mea-related metabolites on T2D risk. **P101**

Table 5.1 Characteristics of participants at baseline in total and by tertiles of meat intake in the EPIC-InterAct (n=20,628) and the UK Biobank (n=316,222) studies. **P113**

Table 6.1. Participant characteristics in the cohorts participating in the InterConnect project on the association between meat consumption and type 2 diabetes (n=1,501,177). **P132**

Table 6.2. The consumption types of meat intake in 23 cohorts participating the InterConnect project on the association between meat consumption and incident type 2 diabetes (n=1,501,177). **P135**

Supplementry table 3.1 The definition of non-communicable diseases outcomes in the exploratory analyses for the association between red meat metabolite score and health outcomes. **P201**

Supplementary table 5.1 SNPs that made up for GRSs and components for Cambridge diabetes risk score. **P204**

Supplementary table 5.2 Associations of meat intake, genetic and clinical risk indexes with T2D in the EPIC-InterAct (n=20,628) and the UK Biobank (n=316,222) studies. **P217**

Supplementary table 5.3 Association between meat consumption and incident T2D in subpopulations in meta-analysis of InterAct and UKBB. **P219**

Supplementary table 5.4 Cumulative incidence rates by subgroups defined by meat intake and other risk factors in the EPIC-InterAct (n=20,628) and the UK Biobank (n=316,222) studies. **P220**

Supplementary table 5.5 Joint effects of meat intake, genetics and clinical risks with T2D incidence in the EPIC-InterAct (n=20,628) and the UK Biobank (n=316,222) studies after meta-analysis. **P221**

Supplementary Table 6.1 Characteristics of 23 cohorts to study the association between meat consumption and incident type 2 diabetes in InterConnect. **P222**

Supplementary Table 6.2 Details of exposure variables used to study the association between meat consumption and incident type 2 diabetes in InterConnect. **P223**

Supplementary Table 6.3 Portion sizes used to study the association between meat consumption and incident type 2 diabetes in InterConnect. **P224**

Supplementary Table 6.4 Covariate variables used to study the association between meat consumption and incident type 2 diabetes in InterConnect. **P225**

Abbreviations

7dDD	7-day Diet Diary
ARI	Absolute Risk Increase
BMI	Body Mass Index
BNF	British National Formulary
Chr	Chromosome
CI	Confidence Intervals
DESP	Diabetic Eye Screening Programme
DIAGRAM	Diabetes Genetics Replication And Meta-Analysis
EAF	Effect Allele Frequency
EDTA	Ethylenediaminetetraacetic Acid
EHR	Electronic Health Record
EPIC	European Prospective Investigation into Cancer and Nutrition
FFQ	Food Frequency Questionnaire
GPC	Glycerophosphocholine
GPE	Glycerophosphoethanolamine
GRS	Generic Risk Score
GWAS	Genome-Wide Association Study
g/d	grams per day
HbA1c	Glycated Haemoglobin
HES	Hospital Episode Statistics
HPLC	High-performance Liquid Chromatography

HR	Hazard Ratio
HRC	Haplotype Reference Consortium
IARC	International Agency for Research into Cancer
IDF	International Diabetes Federation
IQR	Interquartile Range
IVW	Inverse Variance Weighted
LD	Linkage Disequilibrium
MENA	Middle East and North Africa
MI	Myocardial Infarction
MR	Mendelian randomisation
NNUH	Norfolk and Norwich University Hospital
NutriRECS	Nutritional Recommendations
OR	Odds Ratio
PH	Proportional Hazard
PPAR	Peroxisome Proliferator-Activated Receptor
PUFA	Polyunsaturated Fatty Acid
RCT	Randomised Controlled Trial
SACN	Scientific Advisory Committee on Nutrition
SD	Standard Deviation
SE	Standard Error
SFA	Saturated Fatty Acid
SNP	Single Nucleotide Polymorphism
SSB	Sugar-sweetened Beverages

T2D Type 2 Diabetes

TEI Total Energy Intake

TMAO Trimethylamine N-Oxide

UKBB UK Biobank

UPLC MS/MS Ultra-Performance Liquid Chromatography-Tandem Mass Spectrometry

USDA United States Department of Agriculture

WCRF World Cancer Research Fund

WHO World Health Organization

Chapter 1

Introduction

1.1 General introduction

Diabetes mellitus is a complex metabolic disorder and one of the fastest-growing global health problems of the 21st century. One in ten adults aged 20-79 are living with diabetes worldwide, and more than 90% of the cases are type 2 diabetes (T2D). Diabetes and its related complications have imposed substantial diseases and economic burdens on affected individuals, their families, healthcare systems and societies.

The causes of T2D are not entirely understood. Although unmodifiable factors such as genetic susceptibility, family history, ethnicity and ageing contribute to the development of T2D, modifiable factors such as obesity, diet and physical activity also have a role in the onset of T2D. Evidence also shows that T2D can be prevented by interventions on modifiable factors.

Diet is a key modifiable factor in preventing T2D, and meat has received particular attention given its widespread consumption. Meat is an important source of energy and a wide range of nutrients, including protein, B-vitamins, zinc, and iron[1]. Despite its cultural and physiological importance in the diet of many populations, the consumption of meat products may have negative impacts on human health[2]. Moreover, livestock production, as a resource-intensive industry, requires substantial grasslands and water, generates greenhouse gas emissions, and consequently affects our planet's health[3,4]. Therefore, comprehending the impacts of meat consumption is crucial in guiding the transformation of health diets and sustainable food systems, and will contribute to the health of both humans and the planet[5].

Epidemiological research has indicated that habitual intake of red and processed meat might elevate the incidence of T2D. However, it is unclear whether other types of meat (e.g., poultry) are also associated with increased risk. Moreover, currently available research about meat-T2D associations was mainly performed in populations in Europe and North America. Whether associations are the same in individuals from other continents whose culture of diet is significantly different between countries remained unclear. Finally, the causal nature of the association and the mechanisms that may underlie it are not fully understood.

Therefore, the broad objective of this PhD is to provide evidence on meat intake and T2D incidence, with a particular focus on the effect heterogeneity and potential causal mechanisms underlying the meat-T2D associations. This goal was achieved using epidemiological approaches across large-scale population-based cohorts. In the following sections of Chapter 1, I document the background of this research, highlight the research gaps to be filled, and introduce core data resources and methods used in this thesis. Lastly, I outline the overall objectives and chapter scope of this thesis.

1.2 Diabetes: definition, epidemiology, aetiology, and prevention

1.2.1 Definition and diagnosis of diabetes

Diabetes mellitus, widely called diabetes, is a chronic metabolic disease characterised by high levels of blood glucose (hyperglycaemia), resulting from diminished response to the hormone insulin (insulin resistance) and inadequate insulin secretion[6]. T2D is the most common (over 90%) type of diabetes.

Diagnosis of diabetes is mainly based on the examination of plasma glucose or glycated haemoglobin (HbA1c). The cut-off values of diabetes diagnosis are displayed in ***Figure 1.1***, which are recommended by the International Diabetes Federation (IDF) and the World Health Organization (WHO)[6,7]. In clinical practice, two abnormal tests are required to diagnose diabetes if symptoms of hyperglycaemia are absent. In epidemiological studies of a large population, one abnormal test is accepted for the diabetes diagnosis.

Figure 1.1 Modified diagnostic criteria for diabetes by the International Diabetes Federation (IDF)[7]. Fasting is defined as no caloric intake for at least 8 hours. The HbA1c test should be

performed in a laboratory using an NGSP-certified method and standardised to the Diabetes Control and Complications Trial assay. The 2-hour postprandial plasma glucose test should be performed using a glucose load containing the equivalent of 75g anhydrous glucose dissolved in water.

1.2.2 Epidemiology of diabetes

Diabetes prevalence is rising in all regions worldwide. In 2021, over half a billion adults aged 20-79 years were affected by diabetes, which is predicted to reach 783 million by 2045[7].

Figure 1.2 shows the global picture of the estimated prevalence of diabetes among adults in 2021 with adjustment for age. The Middle East and North Africa (MENA) Region has the highest prevalence of diabetes (18.1%), followed by the North America and Caribbean Region (11.9%) and the South-East Asia Region (10.0%). The estimate in the MENA Region will reach 20.4% by 2045. The age-adjusted comparative prevalence in Europe is relatively lower (7.0%), and the Africa Region has the lowest estimate (5.3%). However, the number of people with diabetes is growing sharply in Africa. By 2045, it is expected to increase by 129%, the highest predicted increase across global regions.

The development of T2D is slow and it often exists without apparent symptoms, making a timely diagnosis of diabetes cases extremely challenging in clinical practice. Almost one in two adults with diabetes are undiagnosed, and the percentage of undiagnosed diabetes is exceptionally high in Africa and Western Pacific areas[7].

Diabetes is one of the leading factors that cause death. The West Pacific Region is estimated with the highest number (~2.3 million) of diabetes-related deaths among adults aged 20-79 years old, followed by the Europe Region, with approximately 1.1 million deaths related to diabetes.

Figure removed for copy right reseasons. Copyright belongs to IDF Committee (10th Edition).

Figure 1.2 Age-adjusted comparative prevalence of diabetes in adults aged 20-79 years in 2021, reported by International Diabetes Federation[7].

1.2.3 Aetiology of diabetes

The mechanisms underlying the development of T2D mainly involve insulin resistance in peripheral organs (e.g., liver, muscle, adipose and kidney) and impaired insulin secretion by the pancreatic beta-cells[8,9]. Insulin resistance is the earliest detectable abnormality in the development of T2D. Multiple factors contribute to insulin resistance, including genetic predisposition, ageing, obesity, sedentary lifestyle, smoking and unhealthy dietary factors[10–14]. These risk factors lead to impaired insulin receptor signalling, declined glucose uptake from blood after food intake, excessive hepatic glucose production in the liver, increased glucose reabsorption in the kidney, increased lipolysis, decreased incretin effect, decreased insulin-mediated vasodilation, all of which induce hyperglycaemia. The impaired glucose homeostasis further contributes to dysfunction of the pancreatic beta-cells and a progressive decrease in insulin secretion leads to T2D over the long-term[15].

Hyperglycaemia is associated with the risk of micro- and macro-vascular complications which are disabling and even life-threatening, such as eye disease (retinopathy), kidney injury

(nephropathy), nerve damage (neuropathy), lower-limb amputation, as well as myocardial infarction (MI), stroke and other diseases[16]. If people with diabetes can be diagnosed early with appropriate management, these severe complications can be prevented or delayed, improving quality of life and avoiding death.

1.2.4 Type 2 diabetes prevention

Over the past two decades, randomised controlled trials (RCTs) have demonstrated that T2D can be prevented or at very least delayed through various non-pharmaceutical interventions[17–19]. The lifestyle or behaviours interventions targeting diet, exercise and weight loss have consistently shown beneficial effects on T2D prevention, which were sustainable over time, were practical across age, sex and ethnicity groups and, importantly, were more effective than pharmaceutical therapy such as Metformin[20–24]. However, translating the efficacy evidence from idealised trials into the real-world effectiveness of different prevention programmes has proved challenging.

1.3 Role of meat in T2D incidence

Healthy eating is a public health priority, and meat consumption is the central focus because of its potential adverse impacts on a wide range of disease outcomes, including the onset of T2D.

1.3.1 Meat consumption trends

Consumption of red and processed meat varies between different areas, with a marked increase globally over time. It is above optimal intake levels in many regions, as depicted in ***Figure 1.1***[2,25]. The red and processed meat consumption levels are high in Australia, North America, Western Europe, Latin America, and East Asia, according to GBD 2017 Diet Collaborators[25]. Countries with the highest poultry consumption include Israel, Malaysia, Peru, and United States, according to data from the Organisation for Economic Co-operation and Development[26].

Figure removed for copy right reseasons. Copyright belongs to Godfray et al.[2]

Figure 1.3 Trends of meat consumption in different regions in million metric tons from the 1960s to 2010s[2].

Figure removed for copy right reseasons. Copyright belongs to Godfray et al.[2]

Figure 1.4 Red and processed meat consumption across regions, reported by GBD 2017 Diet Collaborators; solid line, optimal level of intake according to the midpoint of the optimal range of intake; dotted line, global average intake in 2017[25].

1.3.2 Associations between types of meat and T2D

Evidence from epidemiological studies suggests that meat consumption might negatively impact human health by elevating the risk of non-communicable diseases, such as T2D[27–29]. Red meat refers to muscle meat obtained from mammals characterised by its red color, which is attributed to the presence of a protein called myoglobin. This includes products such as beef, veal, pork, lamb, and game. Processed meat is defined as meat that is preserved by curing, salting, smoking, chemical preservatives, or fermentation, such as bacon, sausage, ham, and canned meat products[30]. In the past decade, several meta-analyses of prospective cohort studies have been conducted, demonstrating positive associations between red and processed meat consumption and the risk of T2D[31–40]**.** While consumption of red meat and processed meat has been of interest, the impact of poultry intake on T2D risk were only investigated by a limited number of studies, and results are inconsistent[33,41–45]. Poultry

consumption is regarded as one of the alternatives to red and processed meat consumption[46,47], and is the primary meat type consumed in specific populations, such as those in East Mediterranean and South Asia. Studies about poultry intake and T2D risk are low quality and have potential publication biases, hindering the evaluation of its health impacts[40,48]. Therefore, the evidence for poultry on the risk of T2D needs to be updated with more studies and standardised methods.

1.3.3 Meat and T2D association in different subgroups and across populations

Dietary guidelines in many countries have suggested that adults should limit the consumption of red and processed meat[49]. For instance, the Scientific Advisory Committee on Nutrition (SACN) in the UK advises people to eat less than 70 grams/day (g/d) of red and processed meat [50], and WHO and the World Cancer Research Fund (WCRF) recommend consuming no more than 500 g/week of red meat and very little if any processed meat[51]. However, the Nutritional Recommendations (NutriRECS) Consortium has recently recommended that individuals do not need to change their meat consumption due to the uncertainty of increased risk associated with higher consumption[52]. The authors of the NutriRECS demonstrated that the conclusion was mainly due to concerns regarding the potential risk of bias in the available evidence. It is worth to note that NutriRECS's report has been criticised to have potential bias because NutriRECS has ties to Agriculture and life Sciences Research programme, an arm of Texas A&M University, partialy founded by the beef industry[53,54]. Given this conflicting information, a comprehensive evaluation of the meat-T2D association with reduced bias is needed.

Apart from challenges in delivering dietary guidelines to real-world settings, one more commonly asked question is whether this one-fit-for-all advice can benefit the entire population or might be more effective within subgroups with specific characteristics. The impact of a dietary intervention for each individual in a population may vary due to societal influences, differences in adherence to an intervention, social-economics status and biological

dissimilarities in individuals' responses to food intake. The investigation of meat-T2D associations within population subgroups may improve our understanding of the interplay between meat and other factors in the development of T2D and help optimise dietary prevention strategies for public health.

Habits of meat consumption have a considerable variation across populations. Therefore, data in different geographic areas can help better characterise the association of meat consumption with T2D incidence, especially in non-Western countries where the T2D prevalence is growing more rapidly. The available evidence from literature-based meta-analyses includes studies predominantly from North America and Europe, with a few from Asia and Australia and rare or none from other areas (***Table 1.1***). Even for these available data, results showed a high degree of heterogeneity of associations between meat consumption and T2D incidence[31,36,38], and sources for the heterogeneity remain unclear. The heterogeneity in these meta-analyses may be due to the lack of standardisation or harmonisation for methods and variables, including the differences in serving sizes, meat types, and different degrees of adjustment for potential confounders, all of which could contribute to the heterogeneity.

Table 1.1 Summarised information about the published meta-analysis studies with their involved individual cohorts

Related publication of the individual study	Country	Cohort name	Anne et al., 2009[38]	Micha et al., 2010[37]	Pan et al., 2011[55]	Micha et al., 2012[35]	InterAct et al., 2013[33]	Feskens et al., 2013[36]	Schwingshackl et al., 2017[31]	Tian et al., 2017[32]	Yang et al., 2020[40]
North America											
Meyer, 2001[56]	USA	IWHS		√				√			
Van Dam, 2002[57]	USA	HPFS	√	√						√	
Schulze,2003[58]	USA	NHS II	√	√						√	
Lee, 2004[59]	USA	IWHS	√						√	√	
Fung, 2004[60]	USA	NHS I	√	√						√	
Song,2004[61]	USA	WHS	√	√	√	√		√	√	√	√
Vang, 2008[62]	USA	AHS	√					√		√	√
Steinbrecher, 2011[44]	USA	MEC			√	√		√	√		√
Pan, 2011[55]	USA	HPFS			√	√		√	√		√
Pan, 2011	USA	NHS II			√	√		√	√		√
Pan, 2011	USA	NHS I			√	√		√			√
Fretts, 2012[63]	USA	SHFS						√	√		√
Isanejad, 2017[64]	USA	WHI									√
Europe											
Montonen, 2005[41]	Finland	FMC	√		√	√		√	√	√	√
Schulze, 2007[65]	Germany	EPIC-Potsdam	√	√	√	√				√	
Simmons, 2007[66]	UK	EPIC-Norfolk	√							√	
Mannisto, 2010[43]	Finland	ATBC			√	√		√	√		√
EPIC-InterAct, 2013[33]	Europe	EPIC-InterAct					√	√	√		
Van Woudenbergh, 2012[45]	Netherland	Rotterdam Study						√	√		√
Lajous, 2012[67]	French	E3N						√			
Ericson, 2015[68]	Sweden	MDC									√

Following the last page

Related publication of the individual study	Country	Cohort name	Anne et al., 2009[38]	Micha et al., 2010[37]	Pan et al., 2011[55]	Micha et al., 2012[35]	InterAct et al., 2013[33]	Feskens et al., 2013[36]	Schwingshackl et al., 2017[31]	Tian et al., 2017[32]	Yang et al., 2020[40]
Mari-Sanchis, 2016[69]	Spain	SUN							√	√	
Virtanen, 2017[70]	Finland	KIHD								√	
Australia											
Hodge, 2007[42]	Australia	MCC	√					√		√	
East Asia											
Hirayama,1992[71]	Japan	SPCS	√					√			
Villegas, 2006[72]	China	SHWS	√	√	√	√		√	√	√	√
Kurotani, 2013[73]	Japan	JPHC							√		√
Talaei, 2017[74]	Singapore	SCHS									√
RR for red meat: each 100 g/d increase			1.2 (1.04, 1.38) ⊓	1.16 (0.92, 1.46)	1.19 (1.04,1.37)	1.19 (1.04,1.37)	1.1 (1.04, 1.15) ǂ	1.15 (0.99,1.33)	1.17 (1.08, 1.26)	-	1.31 (1.19, 1.45)
RR for processed meat: each 50 g/d increase			1.57 (1.28, 1.93)	1.19 (1.11, 1.27)	1.51 (1.25,1.83)	1.51 (1.25,1.83)	1.13 (1.04, 1.22)	1.04 (0.82,1.32)	1.37 (1.22, 1.55)	-	1.46 § (1.26, 1.69)
RR for red meat (high vs. low)			1.21 (1.07, 1.38)	-	-	-	1.2 (1.07, 1.35)	-	1.21 (1.13, 1.30)	1.22 (1.10, 1.36)	1.22 (1.16, 1.28)
RR for processed meat (high vs. low)			1.41 (1.25, 1.60)	-	-	-	1.16 (1.04, 1.31)	-	1.27 (1.20, 1.35)	1.39 (1.29, 1.49)	1.25 (1.13, 1.37)

RR, relative risk; ⊓ reported RR for each 120 g/d increase of meat intake; ǂ reported RR for each 50 g/d increase of meat intake; § reported RR for each 100 g/d increase of meat intake.

1.3.4 Meat consumption and T2D risk factors in RCTs

Although observational studies have suggested a link between meat intake and T2D risk, results from trials have not clearly supported associations between red meat intake and risk factors for T2D, such as HbA1c, fasting glucose, fasting or postprandial insulin, HOMA-IR[75–78]. Most of these interventional studies focused on indexes of insulin sensitivity and relatively few evaluated T2D risk factors on other pathways of glucose homeostasis, such as pancreatic beta-cell function, incretin responses to food intake, hepatic glucose production, and adipose tissue[79]. It is unclear what molecules may be involved in the potential causal pathway between meat consumption and the development of T2D. Further investigations on other markers are needed to recognise the potential causal role of meat in the progression of T2D risk and identify possible pathways that underlie the meat-T2D association.

1.3.5 Biological mechanisms insights

The underlying mechanisms between meat intake and T2D development are not understood clearly. Obesity is a major risk factor for diabetes, and it might be a confounder or a modifier of the association between meat and T2D risk. Some cohort studies reported attenuations in associations between meat and T2D after adjustment for BMI[33,80], suggesting that the T2D risk associated with meat may be due to higher adiposity. But the attenuation were not complete and associations persisted after adjustment, suggesting that meat could affect T2D risk via other pathways independent of BMI, such as insulin sensitivity and pancreatic beta-cell function.

Meat is a source of various nutrients, vitamins and minerals, such as protein, B-vitamins and haem iron[1]. It remains unclear which components of meat contribute to the risk of T2D. RCTs and experimental animal studies have explored potential metabolic pathways that link red meat consumption with the development of T2D (***Figure 1.5***)[81]. For example, trimethylamine N-oxide (TMAO) is a gut microbiota-dependent metabolite generated during the digestion of

choline and L-carnitine, which are abundant in red meat. Increased plasma TMAO is associated with an increased risk of atherosclerotic heart disease[82–84]. Although TMAO may be a biomarker for meat consumption, its role in the pathogenesis of T2D has yet to be established[85]. Meat contains saturated fatty acids (SFA) and cholesterol which might affect insulin resistance. However, observational and interventional studies have had mixed results about the role of SFA in linking meat to the risk of T2D or insulin resistance[86–90]. The high iron content of red meat might also contribute to the meat-T2D association based on observational population-based studies. Questionnaire-measured heme iron intake and genetic instrumented heme iron levels were associated with a higher risk of T2D[91], and the association was attenuated but remained significant after adjusting for red meat consumption[92]. The red meat-associated diabetes risk was partly or wholly attenuated after additional adjustment for heme iron in different studies[92,93]. Advanced glycation end products (AGEs) are compounds that are formed primarily by the reaction of proteins with sugars. AGEs can be generated when cooking meat products at high temperatures, and AGEs in meat may contribute to the association between meat intake and T2D risk due to oxidative stress and promotion of inflammation, which can lead to insulin resistance and impaired pancreatic beta-cell function[81,94]. Evidence on the association of between dietary AGEs intake and the development of T2D is limited and additional high-quality studies with larger sample size and long-term follow-up are needed to understand the impacts of AGEs in meat on the risk of T2D[95–97].

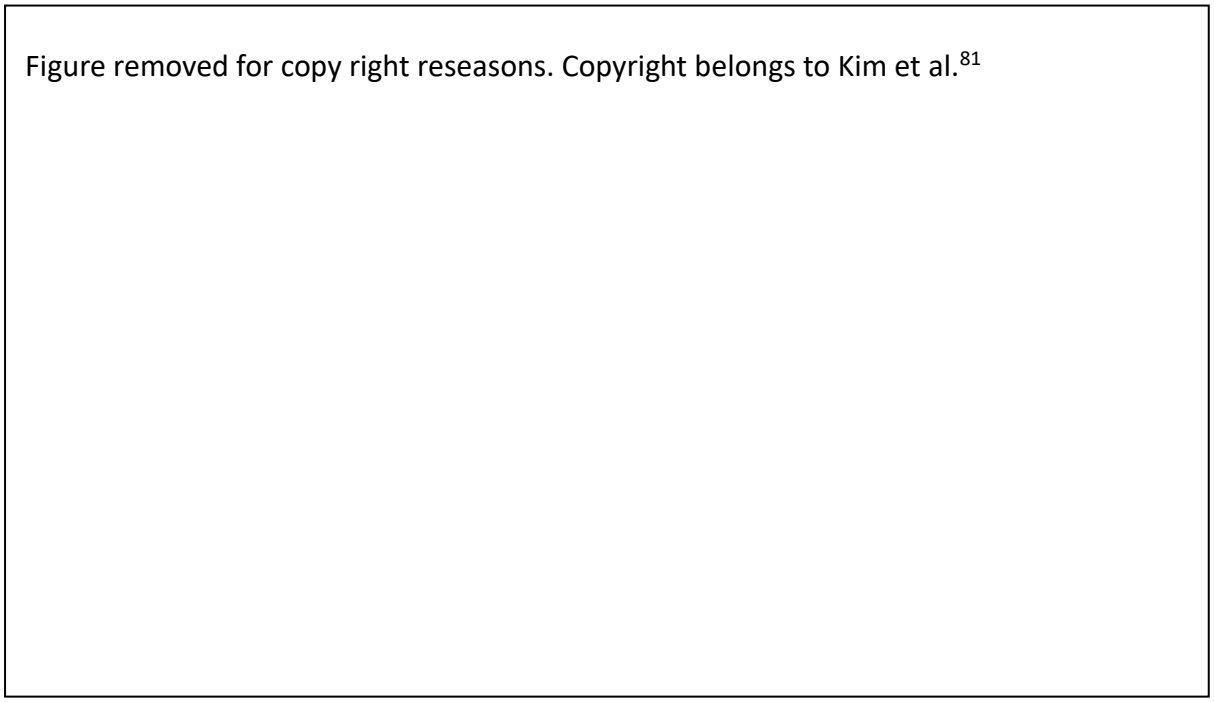


Figure 1.5 Possible mechanisms linking red and processed meat metabolites to the aetiology of type 2 diabetes[81]. AGEs, advanced glycation end products; ATP, adenosine triphosphate; BCAAs, branched amino acids; CRP, C-reactive protein; IRS-1, insulin receptor substrate 1; FFA, free fatty acid; PI3K, phosphatidylinositol 3-kinase; SFA, saturated fatty acid; STMA, trimethylamine; TMAO, trimethylamine N-oxide; TNF-α, tumour necrosis factor-alpha; VCAM-1, vascular cell adhesion 1.

1.4 Main Studies involved in this thesis

1.4.1 EPIC Norfolk Study

The European Prospective Investigation into Cancer and Nutrition in Norfolk (EPIC-Norfolk) study is part of the Europe-wide EPIC study. The EPIC study is an extensive study of diet and health, involving over half a million people in ten countries (Denmark, France, Germany, Greece, Italy, Norway, Spain, Sweden, the Netherlands, and the UK). Details of the recruitment procedures and data collection in EPIC-Norfolk have been described previously [98]. Briefly, the EPIC-Norfolk study is a prospective cohort of 25,636 men and women aged 40 to 79 years, recruited between 1993 and 1998 in Norfolk, England. Baseline characteristics for all participants were collected, including socio-demographic factors, behavioural information, medical history, and anthropometric measurements. Blood samples were collected at baseline and stored in liquid nitrogen at -175^{o}C. The Norwich Local Ethics Committee approved the EPIC-Norfolk study. All participants gave their informed written consent before entering the study.

Dietary assessment

Dietary information in EPIC-Norfolk was assessed by two approaches: the food frequency questionnaire (FFQ) and the 7-day diet diary (7dDD).

FFQ: The FFQ approach has a relatively low burden for participants and can capture a habitual diet. In EPIC-Norfolk, a 130-item semi-quantitative FFQ was used to assess the habitual food intake of a participant during the previous year. The EPIC-Norfolk FFQ was initially developed in 1988 and its food list and portion sizes represent the adult population following a traditional UK diet. More information about the FFQ can be found on the EPIC-Norfolk website (https://www.epic-norfolk.org.uk/for-researchers/ffq/). There were nine frequency categories from "never or less than once/month" to "6 times per day" and specified serving sizes. All the data from the questionnaire were processed with FETA software[99].

7dDD: At recruitment, participants were requested to record prospectively everything they had eaten (food types, amounts, brands, recipes, and cooking methods) in a 45-page booklet for seven consecutive days. The dietary information was processed into nutrient and food group data by several programmes and databases (DINER and CAFÉ) [100–102]. The estimates of diet intake from 7dDDs were considered more accurate than that using FFQs, and 7dDDs allow the disaggregation of food items[100,103–106]. For example, savoury pie or pasta consumption was a single entry in FFQ, but in 7dDD, the item can be disaggregated into food items, including different types of meat and cereal products, if consumed as ingredients. A copy of the 7-day food diary and the portion sizes of the photographs displayed in the diary are available for download from the EPIC-Norfolk website (https://www.epic-norfolk.org.uk/for-researchers/7ddd/).

Both dietary assessment approaches in EPIC-Norfolk have been validated by comparison with weighed records and biomarkers in urine (nitrogen, potassium, and sodium) and blood (carotenoids and vitamin C). The 7dDD and FFQ showed similar correlations with plasma vitamin C but the 7dDD had higher correlations with other biomarkers or weighed records than the FFQ[107–112].

1.4.2 EPIC-InterAct Study

The EPIC-InterAct study is a large prospective case-cohort of T2D nested in the EPIC study, facilitating T2D research involving genetic and lifestyle factors among European populations. A total of 26 research centres in eight European countries (Denmark, France, Germany, Italy, Spain, Sweden, the Netherlands, and the UK) were included in EPIC-InterAct. The detailed description of the study design was published previously[113,114]. In brief, 340,234 eligible participants with information on diabetes status and stored blood samples from the EPIC study were included. During 4 million person-years follow-up, 12,403 incident cases of T2D were ascertained and verified in each participating research centre through multiple sources, including self-reported history of T2D and diabetes medication use, linkage to primary care

registers, secondary care registers, pharmacy registers), hospital admissions data, mortality data, local and national diabetes and pharmaceutical registers. A sub-cohort of 16,835 participants was randomly selected from eligible participants as the control group representative of all EPIC participants. Individuals with prevalent (n=548), unknown (n=129) and post-censoring diabetes status (n=4) were excluded, with a total of 16,154 individuals remaining in the EPIC-InterAct sub-cohort.

As part of the EPIC study, comprehensive information was obtained at recruitment, including socio-demographic information, lifestyle and behavioural factors, and medical history. Quantitative anthropometric measures and blood samples were collected by trained nurses[115]. Blood samples were stored in liquid nitrogen at the International Agency for Research into Cancer (IARC) in Lyon, France, or in local biorepositories. Blood samples in Umeå were stored in -80 °C freezers. Stored blood samples were used for biochemical and genotyping assessments. The details of genetics data, including genotyping, imputation, and quality control were described in **Chapter 5**.

The local ethics committee approved the EPIC-InterAct study in the participating countries and the Internal Review Board of the IARC also provided approval. All participants gave written informed consent. The Medical Research Council Epidemiology Unit coordinated the study at the University of Cambridge.

Dietary assessment

Dietary information was assessed using country-specific dietary questionnaires (e.g., FFQs, dietary history questionnaires, or a modified dietary history) at baseline, which included up to 260 food items. In addition, 24-hour dietary recall data were collected from a representative subgroup of individuals (n=36,900). The dietary questionnaires were validated within each country and country-specific dietary data were standardised using a 24-hour dietary recall program (EPIC-SOFT) to provide comparable dietary measurements across participating countries[114–118].

1.4.3 UK Biobank

The UK Biobank (UKBB) study is a large-scale population-based prospective cohort of 503, 325 individuals (participation rate 5.45%) aged 40-69 years who were recruited from England, Scotland and Wales between 2006-2010[119,120]. Participants provided extensive data on socio-demographic status, lifestyle and behavioural exposure, medical history, and anthropometric and physical measures (e.g., blood pressure, anthropometry, and spirometry) [119,121,122]. Biological samples were collected and stored centrally by UKBB for biochemical assessing (e.g., candidate biomarkers, metabolomics), genotyping, whole-exome sequencing, and recently released whole-genome sequencing[123]. The genotyping of UKBB was described in detail in **Chapter 5**. All participants have been followed up by linkage to electronic health records (EHRs) for various chronic diseases such as diabetes and cardiovascular conditions. Primary care data has been made available for approximately 260,000 participants (45% of the cohort). Additionally, web-based questionnaires were periodically sent to all participants by email to update particular exposures (e.g., diet, occupation) and disorders that are not easily ascertained through linkage to medical records. The study was approved by the North West Multi-Centre Research Ethics Committee (11/NW/0382). All participants provided written informed consent.

Dietary assessment

Dietary intake information was assessed at recruitment for all participants using a touchscreen questionnaire which contained 29 questions about diet. These questions asked the average frequency of consumption of main foods and food groups over the past year. In addition, approximately 200k UKBB participants were followed up to complete at least one online 24-hour dietary recall questionnaire, in which the actual amount of diet consumption was self-reported. The mean daily intakes of foods were estimated by multiplying the frequency of consumption of each food item with each standard portion size. The performance of the touchscreen dietary questionnaire was evaluated by a previous study showing a good agreement with the 24-hour recall data regarding the ability to rank participants' food

consumption[122]. The foods and nutrients assessed by the online 24-hour dietary recall showed good agreements with those measured by an interviewer-administered 24-hour recall completed on the same day (r=0.5-0.9)[124].

1.4.4 InterConnect

InterConnect is an international diabetes and obesity research initiative using global data. InterConnect aims to optimise existing data to enable cross-cohort analyses for associations of genetic and environmental factors with disease risk between populations. Existing methods to address this type of question use results from published literature, require research groups to share results following an agreed analysis plan, or ask collaborative groups to physically bring data together in one place (***Figure 1.6***). However, each approach entails some limitations. For example, a literature-based meta-analysis is liable to potential publication bias. Effect estimates reported in publications can reflect heterogeneous statistical approaches and, therefore, can increase the heterogeneity of published results. Alternatively, some researchers conduct cross-cohort analyses based on study-level data. This approach can alleviate methodological variations because participating studies follow a fixed analysis plan. However, this approach will impose a high burden on the collaborators to perform analysis, especially when researchers need to modify analysis or investigate some points more deeply. To reduce the burden, some researchers adopt pooled meta-analysis, gathering and evaluating individual participant data. This approach involves ethical and legal constraints for data sharing across institutions. Additionally, it is challenging to move and process large datasets all in one place in the central coordinating centre. One key feature of the InterConnect project is that it provides a solution to perform analyses across multiple, distributed datasets without physical pooling of data by taking the analysis to the data***.*** The approaches that InterConnect used are described in **Chapter 1.5.4*.***

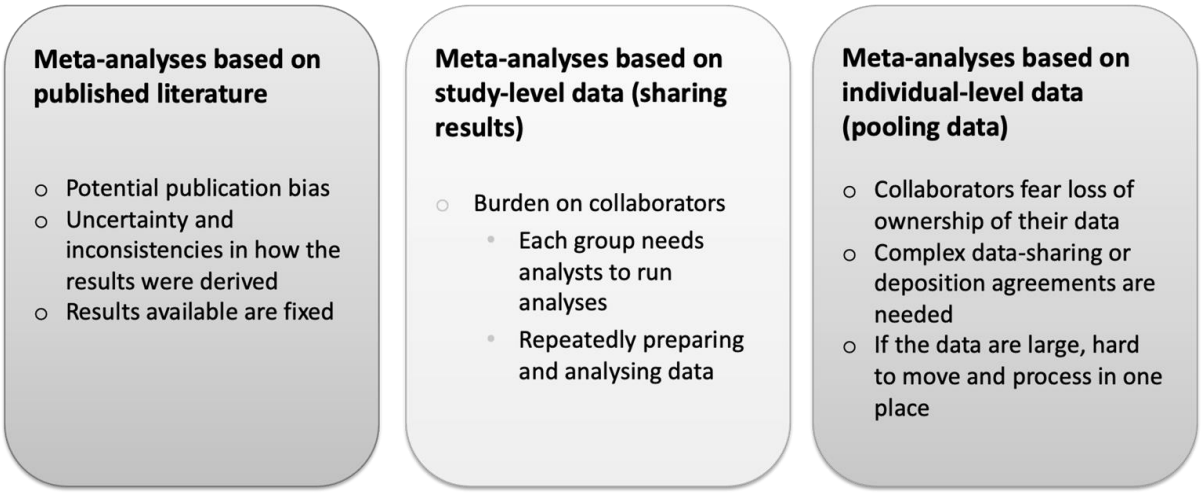


Figure 1.6 Existing approaches to perform cross-cohort analyses and their limitations

1.5 Main methods

1.5.1 Metabolomics

Metabolomics refers to the comprehensive analysis of small molecules (commonly known as metabolites) in a biological specimen, such as blood, urine and saliva. It has the potential to enable systematic characterisation of metabolic phenotypes and simultaneous identification of pathways across diverse diseases[125,126].

Metabolomics plays a crucial role in the field of nutrition. Nutritional epidemiological studies usually rely on self-reported approaches, such as FFQs and dietary records, that are prone to recall bias and measurement errors[127]. Objective biomarkers for nutrients and foods are still limited[128–130]. Metabolomics can be a complementary method to objectively identify new biomarkers for dietary exposures with distinctive strengths. For example, metabolomics profiles are more time sensitive than other omics (genomics, proteomics) and can represent the current biological status of an individual[128,129,131–133]. Metabolomics can account for the intrinsic variability in metabolism by capturing downstream components as well as metabolic products of foods. Metabolomic profiling obtained through untargeted assessment of hundreds of metabolites can reflect the joint impacts of genes as well as behavioural and environmental factors. This provides a unique opportunity to identify the downstream effects of dietary intake in the molecular level and indicate the biological mechanisms that mediate its impact on health consequences[130,134–136].

1.5.2 Mendelian randomisation analysis

Mendelian randomisation (MR) has become a powerful epidemiological approach for assessing the likely causality between an exposure and an outcome within the framework of observational study design[137–141]. MR uses independent genetic variants as instrumental variables to examine how a genetically predicted increase in an exposure affects an outcome.

The causal inference by MR is permissible mainly for two reasons: (1) the genetic variants are free from traditional confounders because the segregation of alleles is randomised during the meiosis of genome. In MR analysis, allocating individuals according to the possession of genetic variants is conceptually similar to the random grouping of participants in RCTs; (2) the design of MR can avoid reverse causality due to the non-modifiable nature of the genome. A sound MR relies on the choice of valid genetic instruments that satisfy with three assumptions. First, the instrument is associated with the exposure of interest. Second, the instrument is not associated with the confounders for the exposure-outcome association. Third, the instrument is not associated with the outcome other than the exposure-outcome pathway[139,142,143].

MR studies adopt individual-level data or summary data. Individual-level data are often unavailable owing to restrictions in data-sharing on a large scale, particularly for data consisting of genetics variables. There has been growing interest in the use of summary data extracted from genome-wide association studies (GWASs), which estimate the association of genetic variants with traits (the exposures or outcomes). The recent development of GWASs has increased the statistical power of MR analysis and prompted rapid increase in MR application in medical research[142].

Two-sample MR method has broadened the application of MR in practice. Genetic risk scores for the exposure can be generated from an independent large-scale GWAS as an alternative measure of the exposure or outcome of interest. This approach is applied in **Chapter 4** of this thesis. An important assumption for two-sample MR is that the two samples for the exposure GWAS and the outcome GWAS are from the same underlying population. The primary effect estimation method for MR is the inverse variance weighted (IVW) regression with random effects estimation[144]. The overall causal effect is estimated by regressing the instrument-outcome effects on the instrument-exposure effects. Other methods have been developed to allow MR estimation under different plausible assumptions, such as MR-Egger, Weighted median and Weighted mode [143,145–150]. These methods are usually used as sensitivity analyses. The details of these approaches are described in **Chapter 4**.

1.5.3 Federated meta-analysis in InterConnect

The InterConnect project is changing how global data is used in research on diabetes and obesity between populations. It supports decentralised variable harmonisation, data analysis, and results synthesis using a secure, scalable and sustainable way.

To achieve this goal, a 'federated meta-analysis' approach is taken using software called DataSHIELD. The working process of the federated meta-analysis is shown in ***Figure 1.7***. This approach enables researchers to send analytical commands remotely, and analysis is executed locally so that all data remain where they originally reside. Only results that do not disclose the identity of any study participants are returned to the analyst. Results generated from each study can be combined to give overall results that are mathematically equivalent to having all the data pooled together in one place in a meta-analysis of harmonised individual-level data.

Figure removed for copy right reseasons. Copyright belongs to DataSHIELD.[151]

Figure 1.7 The framework of how federated meta-analysis was performed using the DataSHIELD[151]. This approach is appropriate for analysing harmonised individual participant data stored at multiple institutions. Each institution installs the server-side DataSHIELD infrastructure that holds a snapshot of the harmonised data to be co-analysed. One of the locations also installs and manages the DataSHIELD client portal, the mechanism by which users are authenticated to send analysis commands within the DataSHIELD infrastructure.

While the InterConnect approach may seem the same as the conventional results-sharing method, there are significant advantages. Because the data stays at the source, it remains

secure behind the firewalls under the control of the originating study institution. The DataSHIELD prevents any viewing or analysing of data for an individual participant. Furthermore, the security and privacy of data held on each server are maintained using standard web security methods. Only users with appropriate permissions can perform analyses, and participating institutions can remove access anytime. This can avoid the challenges of existing methods for analysing data from multiple studies, such as ethical and legal constraints which limit researchers' ability to bring data together physically. What's more, this approach have a principle advantage to reduce the burden for local reasearchers on data analyses and results sharing. The analyst in the central coordinating centre has the flexibility to refine and rerun analyses quickly, ensuring that the analysis plan is executed correctly and that there is no need for someone to analyse each participating institution data.

1.6 Thesis objectives and outlines

In this Chapter, I have summarised the uncertainties about the relationship between meat consumption and T2D. The research performed in this thesis aimed to investigate associations between types of meat consumption and the risk of incident T2D in sub-groups within European populations and across heterogeneous populations worldwide and to explore the causal nature of the association and the mechanisms that may underlie it in analyses of epidemiological studies. Specific objectives and outlines of this thesis are summarised as follows:

Chapter 2: To refine measures of the exposure (meat intake) and the outcome (incident T2D) in the EPIC-Norfolk study and describe the association of self-reported intake of different types of meat (red meat, processed meat and poultry) with T2D.

Chapter 3: To develop and validate a metabolite score as a combined biomarker for meat consumption using observational and trial data and assess the association between the derived score with incident T2D.

Chapter 4: To investigate the potential causal roles of meat-related metabolites (from Chapter 3) on T2D incidence.

Chapter 5: To evaluate whether the association between meat consumption and T2D incidence differed in sub-populations with varying genetic and clinical baseline risks within European populations.

Chapter 6: To investigate associations between types of meat intake and T2D risk across heterogeneous populations worldwide based on a federated platform in InterConnect.

Finally, Chapter 7 discusses the strengths and limitations of the approaches used in this thesis and demonstrates how my PhD work's findings can be translated into practice in the prevention of T2D and public health.

Chapter 2

Association between meat consumption and incident type 2 diabetes ascertained by multiple data sources: food diary vs. FFQ measurements in EPIC-Norfolk

Abstract

Objectives This study aimed to investigate the association of red meat, processed meat and poultry intake with incident type 2 diabetes (T2D), using a 7-day diet diary (7dDD) and a food frequency questionnaire (FFQ).

Methods We used data from 25,636 men and women aged 40-79 years in the European Prospective Investigation into Cancer, Norfolk (EPIC-Norfolk) study. Participants were recruited from 1993-1998 and followed up until 2020. Diet was assessed at recruitment using a prospective 7dDD and a retrospective FFQ. T2D cases were ascertained through linkage to multiple external data sources. Multivariable-adjusted Cox models were used to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) for each type of meat consumption.

Results During a median follow-up of 23 years, 3,019 incident T2D cases were ascertained. After adjustment for potential risk factors, red meat consumption measured by 7dDDs was associated with a higher risk of incident T2D (HR for one SD per day increment 1.06, 95% CI 1.03 to 1.10). After multivariate adjustment, processed meat consumption measured by FFQs was associated with an increased risk of T2D (1.08, 1.05 to 1.12). Poultry intake was not associated with incident T2D by either diet measurement approach.

Conclusion Red meat and processed meat, but not poultry intake, were associated with an increased risk of T2D development. Compared to FFQs, the use of 7dDDs shows a stronger red meat-T2D association but weaker processed meat -T2D association, which suggests that either standard dietary tool might not be optimal for measuring subtypes of meat intake and highlights the evidence triangulation using different dietary measurement approaches in nutritional epidemiological studies.

2.1 Introduction

Diabetes is one of the most common chronic diseases in the UK and worldwide, and its prevalence is increasing globally. Nutrition recommendations have advised the general public to consume less red and processed meat for preventing and managing type 2 diabetes (T2D), based on accumulating evidence of positive associations of red meat and processed meat intake with the risk of T2D in large population-based studies[25,33,39,40,55]. Inconsistent results about poultry intake and the risk of incident T2D have been reported from limited studies[40,48,152]. In most of these epidemiological studies about meat and T2D, quantities of the exposure were mainly measured using food frequency questionnaires (FFQs) in which the frequencies and serving sizes of participants consuming food groups (e.g., fruit, red meat, and dairy) over a specific period (e.g., the past year) were collected.

The dietary record, such as 7-day diet diary (7dDD), is another commonly used self-administered approach to assess actual dietary information in details during a specific period using an open-ended questionnaire in population-based studies. Although the FFQ has its place in nutritional epidemiological studies and 7dDD is more expensive and time-consuming in data interpretation and dishes disaggregation, 7dDD has advantages, such as higher correlation with biomarker data, and without recall bias[127]. It has been shown to improve precision in the estimation of the amount of meat intake using national surveys and cohort data[100,105,153,154]. However, studies on meat intake and incident T2D using 7dDDs are infrequent.

Accurate measurement of exposures and disease outcomes is essential to ensure valid estimation of diet-disease risk and will aid in the formulation of nutritional and public health advice. To our knowledge, the EPIC-Norfolk study is the largest cohort up-to-date (~25,000 British population) which has used both an FFQ and a 7dDD to assess dietary intake. Furthermore, the EPIC-Norfolk cohort has been linked with multiple external electronic health records (EHRs) data. The linkage with these routinely collected data has scientific

strength which can allow to follow up health endpoints of participants in the whole cohort, especially for T2D.

Therefore, in this study, I first identified and ascertained incident T2D cases using survey data and multiple linked EHRs data in EPIC-Norfolk, and then evaluated the association between consumption of meat (red meat, processed meat and poultry) and incident T2D using dietary information assessed by 7dDDs and FFQs.

2.2 Methods

2.2.1 Study population

The study design and details of the EPIC-Norfolk study have been described previously in **Chapter 1.4.1**[98]. Briefly, the EPIC-Norfolk study included 25,636 men and women aged 40-79 years from 1993-1998 in Norfolk, England. For the current study, I included participants who completed 7dDDs and FFQs and excluded those who had either of 1) potential invalid dietary information, indicated by total energy intake of <500 and >3500 kcal/day for females, <800 and >4200 kcal/day for males, 2) missing information in any covariates, or 3) prevalent diabetes at recruitment **(*Supplementary Figure 2.1 and 2.2***).

2.2.2 Dietary assessment

We measured food and nutrient intake using a 130-item FFQ and a 7dDD for all participants at recruitment [100,155]. The details of the measurement of dietary variables has been described in **Chapter 1.4.1**. In an FFQ, participants were asked to choose one of nine categories (ranging from never to over 6 times/day) for each food item. In data processing, portion weights were assigned to each item. A 7dDD recorded everything a participant ate (food types, brands, amounts, recipes, and cooking methods) for 7 consecutive days since the day of a baseline health examination. The process of data entry, data cleaning, food classification and disaggregation can be found elsewhere[100]. For this analysis, meat was classified as red meat (unprocessed beef, pork, lamb, veal, venison, etc.), processed meat (bacon, ham, sausages, etc.) and poultry (chicken, turkey, duck, goose, etc.). All dietary information has been quantified in the unit of grams/day (g/d) for analysis.

2.2.3 Ascertainment of T2D cases

The previous case ascertainment was conducted in 2005-2006 and that work contributed to

a designed case-cohort for diabetes within EPIC-Norfolk and EPIC-InterAct, which comprised 816 incident cases of T2D and a comparison subcohort of 1,018 participants. Incident cases of T2D were ascertained by reviewing evidence from multiple sources, including self-report, linkage to primary and secondary care registers, medication use from drug registers, hospital admissions, and mortality data. All self-reported cases were verified with independent evidence. Person time of follow up was determined from the date of baseline assessment to the date of diagnosis, date of death, or 31 December 2006, whichever came first.

In this study, I expanded the T2D cases acertainment to the whole EPIC-Norfolk study. I identified and ascertained T2D cases by reviewing multi-source data, which not only involved conventional data, such as follow-up surveys of the cohort, linkage with hospital admission and outpatient data from Hospital Episode Statistics (HES), and linkage with death certification from the Office of National Statistics Death Registry, but also several unique external linked data routinely collected from diabetic eye screening programme (DESP), and clinical biochemistry measurements of HbA1c in Norfolk. The overview and timeline of each data source are described in the ***Supplementary Information*** and ***Supplementary Figure 2.1***.

Prevalent diabetes cases were defined as those with any of the following evidence: self-reported diabetes diagnosis, self-reported anti-diabetic drug usage, glycated haemoglobin (HbA1c) ≥48mmol/mol, with medical records of diabetes in HES or with records of attending DESP or relevant treatment in eye clinic before baseline. Participants with prevalent diabetes were excluded from this analysis.

Incident T2D was defined as fulfilling any one or more of the following criteria: (1) ascertained by linkage to a medical record or registry, including HES, Mortality Registry, DESP or relevant treatment in the eye clinic, and clinical biochemistry measurements of HbA1c≥48 mmol/mol; (2) HbA1c≥48mmol/mol in the 2nd Health Examination; (3) self-report of diabetes diagnosis or anti-diabetic medication, verified by any one of additional sources from (1) or (2) above. The workflow of how we define prevalent and incident diabetes in EPIC-Norfolk is shown in ***Supplementary Figure 2.2***.

The diagnosis date for an incident event was set as the earliest date that diabetes was recorded in HES, death registry or DESP, a suboptimal HbA1c level was first measured, or the date of the questionnaire in which diabetes diagnosis or diabetes medication was first reported. For the diagnosis date which was not available, the midpoint between the latest date could be found without evidence of diabetes and censoring was used (e.g., for those who died of diabetes or had self-report of diabetes diagnosis). The follow-up was censored at the date of T2D occurrence, 31 March 2020 or death, whichever occurred first.

2.2.4 Assessment of covariates

Covariates to adjust for in the current analyses were selected based on their established associations with both the dietary exposures and the T2D risk, their biological plausibility in influencing the observed association, or whether or not adjustment influenced the association of interest. Potential dietary confounders were selected based on the availability of corresponding measurements in both 7dDD and FFQ to ensure the comparability of association results. I considered the following variables to be confounding factors, age, sex, smoking status (never, former, and current smokers), physical activity (inactive, moderately inactive, moderately active, active), alcohol drinking (g/d), education (primary school or no qualification, middle school or equivalent, high school or equivalent, college degree and above), total energy intake (kcal/d), BMI (kg/m^2) and dietary intake (g/d) of fruits, vegetables, fatty fish, white fish, nuts, dairy, legumes, eggs, and sugar-sweetened beverages.

2.2.5 Statistical analysis

We described baseline characteristics of participants in total and by red meat intake categories in tertiles. We examined the association between consumption of each type of meat and incident risk of T2D using Cox proportional hazards regression[156] to estimate the hazard ratio (HR) and its 95% confidence intervals (CIs) per standard deviation (SD) of meat intake (or 100 g/d of red meat intake, 50 g/d of processed meat intake and 100 g/d of

poultry intake). We first adjusted for age and sex in a basic model, and then further adjusted for smoking status, physical activity, alcohol drinking, education, BMI, total energy intake and other food groups (fruits, vegetables, white fish, fatty fish, nuts, legumes, eggs, dairy and sugar-sweetened beverages). The independence of the associations of specific meat types was tested by mutually adjusting for other meat types.

2.3 Results

2.3.1 Baseline characteristics

We present the baseline characteristics of all eligible participants (n=23,406) and by red meat consumption categories in ***Table 2.1***. A total of 697 participants with prevalent diabetes were excluded from the analysis (***Supplementary Figure 2.2***). The study population had a mean age of 56.9 years at baseline, and 56% were women. The meat consumption (mean ± SD) was 33.8 ± 29.8 g/d for red meat, 22.2 ± 21.0 for processed meat and 25.2 ± 27.9 for poultry. Participants with higher red meat consumption were more likely to be male, current or former smokers, and had higher alcohol consumption and total energy intake, but with lower consumption of fruits and fish.

Table 2.1 Baseline characteristics of the study participants by tertiles of red meat consumption and in total in the EPIC-Norfolk study (n=23,406)

	Tertile 1 [0.0, 17.7) (n=7,813)	Tertile 2 [17.7, 41.6) (n=7,719)	Tertile 3 [41.6,425.6] (n=7,802)	Total (n=23,406)
Red meat intake, g/d	5.8 ± 6.3	29.2 ± 6.8	66.6 ± 26.4	33.8 ± 29.8
Processed meat intake, g/d	20.0 ± 22.9	23.0 ± 18.9	23.7 ± 20.7	22.2 ± 21.0
Poultry intake, g/d	29.0 ± 34.2	25.4 ± 23.7	21.3 ± 24.0	25.2 ± 27.9
Age, y	58.5 ± 9.4	59.6 ± 9.5	59.1 ± 9.2	59.1 ± 9.3
Sex, n(%) women	5011 (64 %)	4553 (58 %)	3354 (43 %)	12918 (55 %)
Smoking				
Current	855 (10 %)	901 (11 %)	1122 (14 %)	2878 (12 %)
Former	3201 (39 %)	3335 (41 %)	3717 (46 %)	10253 (42 %)
Never	4099 (50 %)	3919 (48 %)	3315 (41 %)	11333 (46 %)
Alcohol intake, g/d	9.9 ± 15.9	11.2 ± 16.3	14.9 ± 19.9	12.0 ± 17.6
Physical activity				
Inactive	2397 (29 %)	2457 (30 %)	2470 (30 %)	7324 (30 %)
Moderately inactive	2463 (30 %)	2317 (28 %)	2285 (28 %)	7065 (29 %)
Moderately active	1864 (23 %)	1892 (23 %)	1836 (23 %)	5592 (23 %)
Active	1431 (18 %)	1489 (18 %)	1563 (19 %)	4483 (18 %)
BMI, kg/m^2	26.0 ± 4.0	26.3 ± 3.8	26.6 ± 3.8	26.3 ± 3.9
Education				
primary school or no qualification	2842 (35 %)	3101 (38 %)	2972 (36 %)	8915 (36 %)
middle school or equivalent	860 (11 %)	853 (10 %)	816 (10 %)	2529 (10 %)
high school or equivalent	3249 (40 %)	3203 (39 %)	3401 (42 %)	9853 (40 %)
college degree and above	1204 (15 %)	998 (12 %)	965 (12 %)	3167 (13 %)
Total Energy Intake, kcal/d	1830.1 ± 505.7	1920.4 ± 495.5	2079.7 ± 532.0	1940.3 ± 521.6
Fatty fish intake, g/d	14.2 ± 24.4	12.5 ± 18.8	10.6 ± 17.9	12.4 ± 20.6
White fish intake, g/d	16.3 ± 22.8	15.8 ± 16.6	13.6 ± 16.3	15.3 ± 18.8
Fruit intake, g/d	184.2 ± 144.4	165.1 ± 117.3	151.4 ± 123.1	167.4 ± 129.2
Vegetable intake, g/d	150.1 ± 88.0	147.4 ± 68.4	158.2 ± 73.6	151.3 ± 77.4
Legumes intake, g/d	27.6 ± 33.5	28.0 ± 26.8	29.9 ± 29.9	28.5 ± 30.2
Nuts intake, g/d	2.36 ± 6.71	2.19 ± 5.93	2.25 ± 6.81	2.27 ± 6.49
Dairy intake, g/d	223.5 ± 152.0	220.8 ± 141.2	221.2 ± 148.4	221.9 ± 147.3
Egg intake, g/d	13.9 ± 19.1	13.8 ± 14.8	14.4 ± 17.3	14.1 ± 17.2
Sugar-sweetened beverages intake, g/d	30.2 ± 83.3	33.9 ± 71.9	37.4 ± 79.8	33.8 ± 78.5

Tertile cut-offs are based on absolute intakes; continuous variables were reported in mean ± SD; categorical variable were reported in %. BMI, body mass index. Dietary information was estimated using 7-day diet diaries.

2.3.2 Association of red meat consumption with incident T2D

During a median follow-up of 23 years (484,280 person-years), 3,019 participants developed diabetes, with an incidence rate of 7.7 cases per 1000 person-years. The number of newly

ascertained diabetes cases by sources is shown in ***Supplementary Figure 2.3.*** Linked records to clinical biochemistry measurements (n cases=2,708), eye screening records (n cases=2,583) and HES (n cases=2,484) contributed most to the ascertained cases, and a large proportion (70%) of new diabetes events were ascertained from at least two data sources.

In the prospective analysis with dietary intake assessed by 7dDDs, higher red meat intake was associated with an increased risk of incident T2D (HR per SD 1.06, 95% CI 1.03 to 1.10, *P*<0.001) with adjustment of potential confounders (***Table 2.2***). The associations between processed meat, poultry consumption and incident T2D were not significant with adjustment of confounding factors (HR per SD for processed meat 1.04, 95% CI 1.00 to 1.07, *P*=0.06; HR per SD for poultry 1.02, 95% CI 0.98 to 1.05, *P*=0.29). When we used FFQs to assess food consumption, we observed a significant positive association between processed meat and incident T2D (HR per SD 1.08, 95% CI 1.04 to 1.12, *P*<0.001), but not for red meat, nor poultry intake (HR per SD for red meat 1.02, 95% CI 0.99 to 1.06, *P*=0.36; HR per SD for poultry 1.00, 95% CI 0.96 to 1.03, *P*=0.94). Associations did not change substantially when we further mutually adjusted for other meat types.

Table 2.2 Association between types of meat consumption and incident T2D measured by 7-day diet diary (7dDD) or food frequency questionnaire (FFQ): the EPIC-Norfolk study*

Exposure	Model	7dDDs			FFQs		
		Intake†	**HR‡**	***P* value**	**Intake**	**HR**	***P* value**
Red meat	model1	33.8 ± 29.8	1.10 (1.06, 1.14)	1.97E-07	40.1 ± 28.9	1.07 (1.04, 1.10)	6.14E-05
	model2		1.06 (1.03, 1.10)	6.52E-04		1.02 (0.99, 1.06)	0.36
Processed meat	model1	22.2 ± 21.0	1.10 (1.07, 1.14)	1.61E-09	18.4 ± 15.5	1.13 (1.09, 1.16)	2.66E-13
	model2		1.04 (1.00, 1.07)	0.06		1.08 (1.04, 1.12)	5.08E-05
Poultry	model1	25.2 ± 27.9	1.03 (1.00, 1.07)	0.10	26.3 ± 21.0	1.00 (0.97, 1.04)	0.94
	model2		1.02 (0.98, 1.05)	0.29		1.00 (0.96, 1.03)	0.94

*N participants=23,406, n cases=3,019. Cox proportional hazards models were adjusted for age, sex in model 1, and further adjusted for physical activity, smoking status, alcohol drinking, alcohol drinking squared, education, BMI, BMI squared, total energy intake, fruits, vegetables, fatty fish and white fish, nuts, dairy, legumes, eggs and sugar-sweetened beverages in model 2. †Intake, the mean ± standard deviation of each dietary exposure in grams per day. ‡HR, hazard ratio per standard deviation of the exposure.

2.4 Discussion

In this large prospective study, we investigated associations between different types of meat consumption and incident T2D using dietary data assessed by 7dDDs and FFQs. I reported that a SD higher red meat intake measured by 7dDDs was associated with a 6% increased risk of developing incident diabetes; 1SD higher intake of processed meat assessed with FFQs was associated with 8% higher risk of incident T2D during over 20 years of follow-up. I did not find a significant association between poultry consumption and incident T2D using either 7dDDs or FFQs.

To the best of my knowledge, this is the first study examining the association between meat intake and T2D using both multiple-day diet diaries and FFQ data. In this study, I reported positive associations of red and processed meat consumption with incident T2D, which was in line with results from several meta-analyses[33,35,38,40,55].

Moreover, I observed different effect estimates for red meat and processed meat on T2D risk when using 7dDDs compared to when using FFQs. The association between red meat and incident T2D was stronger when using 7dDD-measured dietary intake than when using FFQs, whereas the processed meat-T2D association performed stronger when measured by FFQs compared to that by 7dDDs. In published cohort studies, processed meat was consistently associated with an increased risk of T2D, and had a larger coefficient than that for red meat intake. Most of these studies used FFQs to assess meat consumption. The dissimilar results of meat and T2D using 7dDDs versus FFQs suggest that different dietary assessment approaches had their pros and cons when measuring different types of foods, especially meat. Compared to red meat, processed meat tends to be consumed more episodically in populations which may result in larger day-to-day variations. The prospective nature of food diaries enables the recording of diverse information of non-predefined foods and amounts at the time of consumption, and also allows for the disaggregation of different meat types from composite dishes. Although FFQs collect dietary information from a predefined food

list and are prone to recall bias, they may be good at capturing foods that are consumed irregularly, such as processed meat.

The evidence about the link between poultry consumption and T2D was less conclusive in the literature. Many studies reported null associations[44,68,73,74], which are similar to what I found in this study using both FFQs and 7dDDs, whereas others observed a positive association between poultry consumption and the development of T2D[64,157]. Ibsen et al. also reported that the replacement of red meat with poultry was estimated to reduce the risk of T2D[47]. The inconclusive association between poultry and T2D is largely unexplained, which could be due to differences in population characteristics, or different cooking methods (grilling, stir-frying or steaming) for poultry used in different populations. Liu et al. found that the open-flame and high-temperature cooking methods (e.g., grilling, broiling) frequency of chicken was associated with incident T2D in three US populations[97].

About 1 in 3 of T2D in the population may be undiagnosed because T2D usually has a slow onset with imperceptible symptoms[7]. One strength of this study is that I ascertained T2D outcomes using both common and unique external data sources to detect as many cases as possible in this large cohort. In epidemiological studies, the use of electronic health records has been suggested as a pragmatic and efficient approach for detecting diabetes cases in the UK, due to reduced burden on participants and researchers, relative low costs, and less selection bias in follow-up studies[158,159]. For the EPIC-Norfolk study, linkages with the eye screening programme and clinical biochemistry data in the Norfolk area provide vital and novel sources for identifying diabetes cases. Apart from these unique non-national data sources, we also used nationwide data sources of HES and mortality registry, which could identify cases who might have moved out of the local area after recruitment and therefore are not captured by the above area-specific data sources. All of these complementary data sources enabled a comprehensive case ascertainment for diabetes in this study, which could benefit understanding diet-disease associations.

This study has several limitations. First, there might be residual confounding even though we

have adjusted for a range of covariates. Second, we only assessed diet intake at recruitment thus our results cannot justify potential dietary changes over the follow-up period. Third, the study population is predominantly British and the generalisation of results to other populations is limited. Additionally, linkage with EHRs has an inherent bias in administrative data toward people who use health services or attend for screening. In this study, the potential misclassification of incident T2D is likely be non-differential across meat consumption levels. Consequently, the random misclassification may have little bias to the point estimates of associations but could widen the confidence intervals[160]. TThe use of longitudinal biochemistry measurements allowed for the inclusion of those undiagnosed diabetes cases and provided less biased information for diabetes.

In conclusion, in this large prospective study, higher consumption of red meat and processed meat, but not poultry, was associated with an increased risk of T2D. We observed a stronger red meat-T2D association but a weaker processed meat-T2D association when using 7dDDs compared with using FFQs, suggesting that 7dDDs might not always be better than FFQs, especially when measuring foods that are consumed episodically or with large variations in populations. Further studies about the comparison of 7dDDs and FFQs in diet-disease associations are needed in more diverse food groups, which could help better understand the roles of dietary habits in health process.

Chapter 3

Development and validation of a metabolite score for red meat intake: an observational cohort study and a randomised controlled dietary intervention

Publication related to this chapter

Li C, Imamura F, Wedekind R, Stewart ID, Pietzner M, Wheeler E, Forouhi NG, Langenberg C, Scalbert A, Wareham NJ. Development and validation of a metabolite score for red meat intake: an observational cohort study and randomised controlled dietary intervention. Am J Clin Nutr. 2022 Aug 4;116(2):511-522. doi: 10.1093/ajcn/nqac094. PMID: 35754192; PMCID: PMC9348983.

Contributions and collaborations

N.J.W, F.I, and I designed the research; I analysed the data using data from the EPIC-Norfolk study; R.W conducted the laboratory analyses of the intervention study; I worked with F.I, R.W, A.S, and N.J.W to interpret the results; I drafted the original manuscript; N.J.W, F.I and R.W edited the manuscript; I.D.S, M.P, E.W, N.G.F, and C.L provided administrative, technical or material support; and all authors revised and approved the final manuscript.

Abstract

Background Self-reported meat consumption is associated with disease risk but objective assessment of different dimensions of this heterogeneous dietary exposure in observational and interventional studies remains challenging. This study aimed to derive and validate scores based on plasma metabolites for types of meat consumption. For the most predictive score, I aimed to test whether the included metabolites varied with change in meat consumption, and whether the score was associated with incidence of type 2 diabetes (T2D) and other non-communicable diseases.

Methods I derived scores based on 781 plasma metabolites for red meat, processed meat and poultry consumption assessed with 7-day food records among 11,432 participants in the European Prospective Investigation into Cancer and Nutrition-Norfolk (EPIC-Norfolk) cohort. The scores were then tested for internal validity in an independent subset (n=853) of the same cohort. In focused analysis on the red meat metabolite score, whether the metabolites constituting the score were also associated with meat intake in a randomised cross-over dietary intervention on meat (n=12, Lyon, France, NCT03354130) were examined. In the EPIC-Norfolk study, I assessed the association of the red meat metabolite score with T2D incidence (n=1,478) and other health endpoints.

Results The best performing score was for red meat, comprising 139 metabolites which accounted for 17% explained variance of red meat consumption in the validation set. In the intervention, 11 top-ranking metabolites in the red meat metabolite score increased significantly after red meat consumption. In the EPIC-Norfolk study, the red meat metabolite score was associated with T2D incidence (adjusted hazard ratio per standard deviation 1.17, 95% confidence interval 1.10 to 1.24).

Conclusions The red meat metabolite score derived and validated in this study contains metabolites directly derived from meat consumption and is associated with T2D risk. These findings suggest the potential for objective assessment of dietary components and their application for understanding diet-disease associations.

3.1 Introduction

Meat is an important component of the human diet and high consumption is a risk factor for many non-communicable diseases, including type 2 diabetes (T2D)[2,40,48,161,162]. However, meat consumption is a heterogeneous exposure and assessing total meat intake and specific subtypes such as red meat in epidemiological studies that evaluate its influence on health outcomes remains challenging.

Metabolite profiling is a promising approach for quantifying habitual meat intake and can be a complementary approach to self-reported dietary assessment methods (e.g., food frequency questionnaires (FFQs) or dietary records)[125,163]. Diet is an important determinant of the plasma metabolome and one study estimated that it accounts for 50% of the explainable variance, compared to 2% of the variance explained by lifestyle factors, including smoking status, exercise time, etc.[163]. Measurement of metabolites as a complement to self-reported assessment methods has other theoretical advantages, including diminishing social desirability bias and recall bias, and greater comparability across populations[127,134].

Several individual metabolites have previously been reported to be significantly associated with different types of meat consumption[164–167]. However, few studies have examined how combinations of metabolites can predict meat consumption. Cuparencu et al. reported that a combination of six metabolite biomarkers were able to assign people to a binary classification of red meat consumption in a 2-day feeding trial. However, the study was small and the result may be liable to overfitting[166].

In the current study, we aimed to develop and test metabolite scores for different types of meat consumption by combining 781 blood metabolites in the European Prospective Investigation into Cancer and Nutrition-Norfolk (EPIC-Norfolk) cohort and then to take forward the red meat metabolite score to potential replication in a short-term randomised controlled trial (RCT) that measured metabolites after a red meat and a non-meat diet. Finally we tested whether the meat metabolite score associated with the risk of incident T2D and other non-communicable diseases to explore the potential utility of the score in understanding disease risk.

3.2 Methods

3.2.1 Data source and study design

The overall design of the project includes a derivation and validation phase in an observational study, a test of change in an RCT and a test of association with incident health outcomes in a prospective study as shown in ***Figure 3.1***.

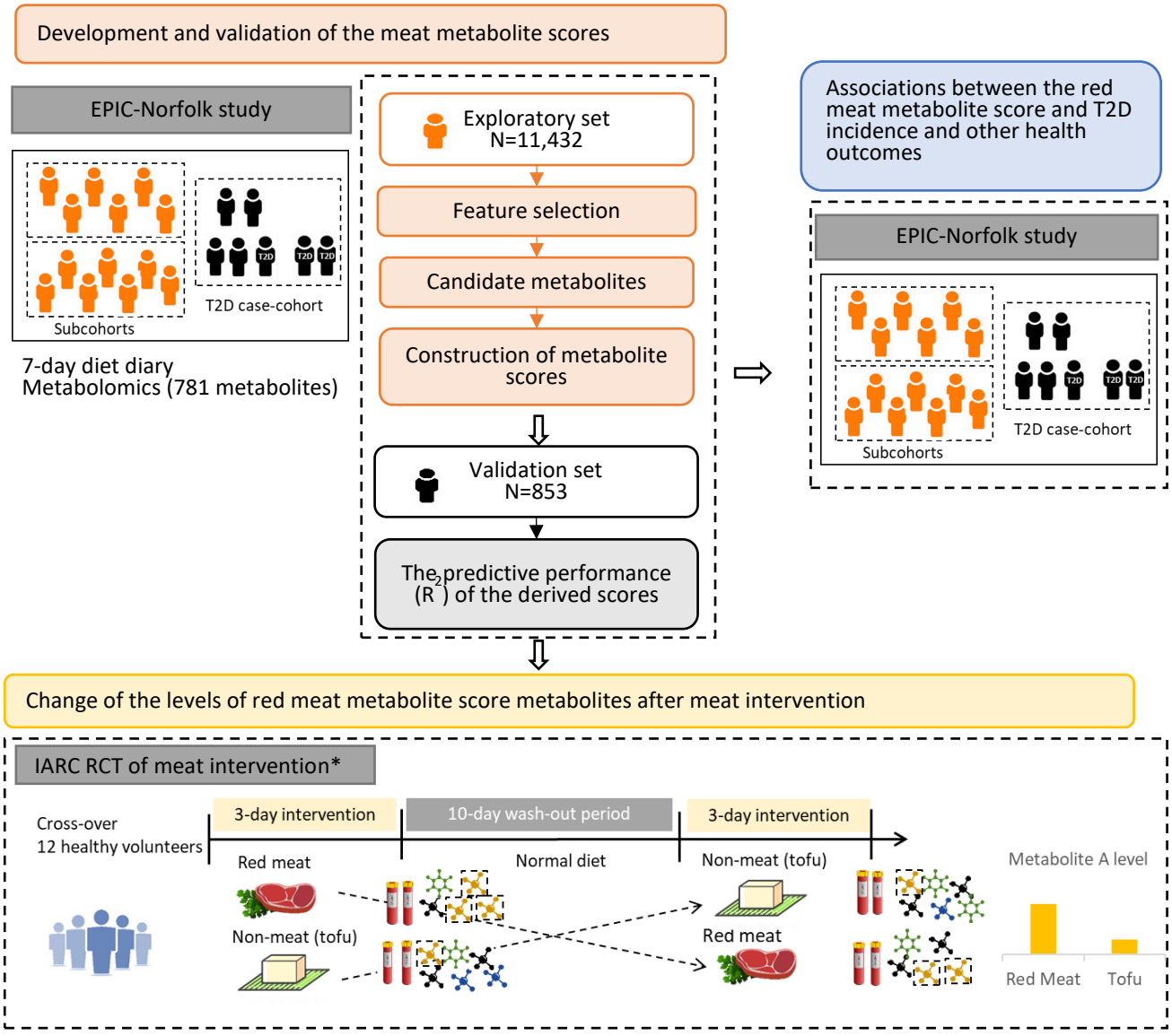


Figure 3.1 Flow chart for the overall analytic approach for development and validation of the meat metabolomics score. *the visualisation simplifies the design of RCT, only two out of five arms are shown.

3.2.2 Observational data for the derivation and validation of the metabolite scores: the EPIC-Norfolk study

I developed and validated the metabolite scores for three types of meat consumption (red meat, processed meat and poultry), using baseline data from the EPIC-Norfolk study which originally recruited 25,636 men and women aged 40-79 years between 1993 and 1998 in the United Kingdom. Details of the recruitment procedures and data collection have been described previously[98] and also in **Chapter 1.4.1**.

I developed metabolite scores for different types of meat consumption in an exploratory set which included 11,432 participants who had both untargeted metabolomics and dietary data. I excluded from this exploratory dataset individuals who were part of a nested case-cohort study for incident T2D; those with extreme energy intake measures (<500 and >3500 kcal/d for women, <800 and >4200 kcal/d for men); or those with prevalent diabetes at baseline. Participants from the subcohort of an independent nested T2D case-cohort study[168] were used as a validation set, which included 853 participants after exclusions.

Metabolomics measurement and data processing in the EPIC-Norfolk study

Untargeted metabolomics data were measured using ultra-performance liquid chromatography-tandem mass spectrometry (UPLC MS/MS) on the Metabolon DiscoveryHD4® platform from plasma samples collected at baseline. The measurement of metabolites was performed in three subsets in March 2015, January 2016 and March 2017 successively. The data quality control and processing methods have been described previously[126] and are summarised in the ***Supplementary Information***. After data quality control and data management, three subsets included 1,503, 5,992 and 5,980 individuals, in which 944, 1,168 and 1,219 metabolites were measured, respectively, and 781 metabolites were identical across all subsets.

Before analysis, values of each metabolite underwent the following steps within each subset: log-transformation, replacement of outliers with 5 standard deviations (SDs) from the mean (winsorisation) and standardisation to a mean of 0 and SD of 1. For metabolite concentrations that were assumed to be below the limit of detection, we imputed them with the lowest values of that metabolite[169]. The different subsets in the exploratory dataset underwent the metabolomics assays in different time points, and the time difference was adjusted for in the

subsequent analysis with a regression technique.

Assessment of meat consumption in the EPIC-Norfolk study

Meat consumption and other dietary exposures were assessed with a 7-day diet diary (7dDD) as documented previously[100]. The details about using 7dDDs for dietary assessment and nitrition data processing are described in **Chapter 1.4.1**. For this study, the meat related categories were all disaggregated from composite dishes including red meat (unprocessed beef, lamb, pork, veal, rabbit, venison etc.), processed meat (bacon, ham and sausages etc., smoked, cured, salted or chemically-preserved), and poultry (chicken, turkey, goose, duck, guinea fowl, pheasant etc.) in the unit of grams per day (g/d). Participants were also asked whether they followed a special diet (vegetarian, other diet or no special diet).

Development and validation of metabolite scores of self-reported red, processed meat and poultry consumption

In the EPIC-Norfolk study, 781 metabolites were evaluated simultaneously for the prediction of red meat consumption. In the exploratory set, I applied elastic net regression[170] with a bootstrapping approach[171,172] to select a combination of metabolites for the prediction of red meat consumption; and ridge regression[173] to estimate penalised weights of these candidate metabolites (***Supplementary Information***). I applied the weights of all candidate metabolites and constructed a metabolite score for each individual in both of the derivation and validation datasets. The score was standardised to a mean of 0 and SD of 1 for further analysis. The metabolite scores for processed meat and poultry were derived and tested using the same process.

3.2.3 Randomised controlled trial of meat consumption

Given the availability of trial-based data for meat consumption, I further investigated associations of metabolites in the score from the observational EPIC-Norfolk study with red meat consumption in an RCT previously conducted in Lyon, France in 2018. The details of this RCT have been reported previously[174]. In brief, 12 healthy adults consumed in random order 5 different foods (fried pork, hot dogs, bacon, salami and tofu) as part of a controlled diet. For this analysis, the differences in metabolites levels between fried pork (unprocessed red meat) and tofu control arms were examined. In this trial, fasting plasma samples were collected in

the morning after the last meal of each test period. Participants provided informed consent and procedures were carried out according to the Declaration of Helsinki. The study was approved by the International Agency for Research on Cancer (IARC) Ethics Committee (IEC Project 17–12) and registered at *clinicaltrials.gov* (NCT03354130).

Test of candidate metabolites of red meat intake in the RCT

In the RCT, whether metabolites that were part of the metabolite score for red meat intake were increased after intake of fried pork (red meat) compared to tofu was evaluated. The process of identification of metabolites that make up the red meat metabolite score in the RCT is shown in ***Supplementary Figure 3.1***. First, the primary focus was on metabolites that had been annotated successfully in the IARC laboratory and had a positive coefficient in the metabolite score. Corresponding signal intensities were extracted with Agilent Profinder 10.0 (Agilent Technologies, Santa Clara, CA, USA) using the find-by-formula method ($[M+H]^+$ and $[M-H]^-$ ions only, exact mass +/- 8ppm, Retention time +/- 0.05 min). Metabolites were carried forward for statistical analysis if they were detected in >75 % of the samples collected after pork intake. Then paired Welch's t-tests were used to assess whether metabolites were significantly (P<0.05) elevated in plasma samples collected after pork intake compared to tofu intake. Second, for metabolites not previously identified in the IARC laboratory, only those with a coefficient of >1.0 in the meat intake score were extracted from the raw data by chemical formula to test for their increase in plasma samples after pork intake. Compounds were confirmed by comparison of MS/MS spectra with those in the literature (annotation confidence level 2 or 3)[175].

3.2.4 Prospective cohort analysis of the association of the red meat metabolite score with incident disease outcomes in the EPIC-Norfolk study

I also examined the association of the red meat metabolite score and the relevant self-reported consumption parameter with the risk of incident T2D in a case-cohort study nested in the EPIC-Norfolk cohort[168]. This comprised a total of 659 incident cases of T2D and a comparison subcohort of 846 participants, which had an overlap by design of 27 individuals with the case set, after we excluded participants who had extreme energy intake measures or missing covariate data. Participants were followed up from baseline to 31 December 2006.

The details of the study design of the case-corhort for T2D in EPIC-Norfolk and the ascertainmetn of T2D cases are described in **Chapter 2.2.3**.

Assessment of covariates in the EPIC-Norfolk study

Information about health behaviors and clinical risk factors were collected by trained nurses during a health check at baseline. Information obtained included age, sex, education level (primary school or no qualifications, middle school or equivalent, high school or equivalent, college degree and above), smoking status (never, former, and current smokers), alcohol drinking (g/d), physical activity (inactive, moderately inactive, moderately active, active), height (m), weight (kg), and other food group consumption in g/d (fruits, vegetables, fatty fish, white fish, dairy, legumes, nuts, eggs and sugar-sweetened beverages). BMI was calculated as weight divided by the square of height (kg/m^2). Total energy intake was calculated from 7dDDs.

Statistical methods for the assessment of the association with incident T2D

I analysed the association of a standardised metabolite score for red meat consumption with incident T2D in the case-cohort study using Prentice-weighted Cox regression[176] to estimate the hazard ratio (HR) for T2D and its 95% confidence interval per SD of the exposure.

I considered the effect of potential confounders in a model adjusting for age, sex, and then further adjusted for education, smoking status, alcohol drinking, BMI and dietary factors (consumption of fruits, vegetables, fatty fish and white fish, sugary beverages, dairy, legumes, nuts, eggs and total energy intake). For alcohol drinking and BMI, their linear and squared terms were included to account for their potential non-linear associations with each outcome.

Ascertainment of other non-communicable diseases outcomes in the EPIC-Norfolk study

I investigated the incident outcomes of six health conditions including cardiovascular diseases (including ischemic heart disease, hemorrhagic stroke, cerebral stroke, heart failure, and atrial fibrillation); gastrointestinal cancers (including colon cancer, rectal cancer, stomach cancer); liver disease, renal disease, fractures, and deaths due to any causes[126]. Outcome data were obtained by linkage to Hospital Episode Statistics, the cancer registry and the Office of National Statistics. Follow-up ended on March 31st, 2016. Prevalent and incident cases for each disease were identified with the International Classification of Diseases 10th revision as listed in ***Supplementary Table 3.1***.

Statistical methods for the assessment of the association with multiple disease outcomes

In an exploratory analysis I tested the association of the red meat metabolite score with incident health outcomes using standard Cox regression after excluding the prevalent cases for each clinical outcome (see ***Supplementary Table 3.1)***. I adjusted for the same sets of potential confounders as considered in the association with T2D.

3.3 Results

3.3.1 Baseline characteristics and meat consumption of study participants in the EPIC-Norfolk study

The baseline characteristics of the participants in the exploratory and validation sets within the EPIC-Norfolk study are shown in ***Table 3.1***. Among the 11,432 participants in the exploratory set, 46% were male and the mean (SD) age at baseline was 59.6 (9.0) years. The mean (SD) meat consumption in g/d was 34.4 (29.3) for red meat, 22.5 (21.0) for processed meat, and 24.8 (27.5) for poultry. The characteristics in the validation set (n=853) were broadly similar to those in the exploratory set.

Table 3.1 Baseline characteristics of the study participants for development and validation of meat metabolite scores in the EPIC-Norfolk study[π]

	Exploratory set (n=11,432)	Validation set (n=853)
Age, y	59.6 ± 9.0	59.0 ± 9.4
Female	6204 (54 %)	494 (58 %)
Red meat intake, g/d	34.4 ± 29.3	33.6 ± 29.1
Processed meat intake, g/d	22.5 ± 21.0	21.7 ± 19.7
Poultry intake, g/d	24.8 ± 27.5	26.0 ± 25.5
Education		
No	4345 (38 %)	326 (38 %)
Olevel	1155 (10 %)	79 (9 %)
Alevel	4541 (40 %)	330 (39 %)
Degree	1385 (12 %)	117 (14 %)
Missing	6 (0.1%)	1 (0.1%)
Smoking		
Current	1290 (11 %)	112 (13 %)
Former	4826 (42 %)	329 (39 %)
Never	5224 (46 %)	407 (48 %)
Missing	92 (0.8%)	5 (0.6%)
Alcohol intake, g/d	11.9 ± 17.8	11.6 ± 16.6
Physical activity		
Inactive	3325 (29 %)	238 (28 %)
Moderately inactive	3243 (28 %)	246 (29 %)
Moderately active	2658 (23 %)	206 (24 %)
Active	2206 (19 %)	163 (19 %)
BMI, kg/m^2	26.1 ± 3.7	26.1 ± 3.7
Total Energy Intake, kcal/d	1950.2 ± 526.1	1940.0 ± 517.3
Fruit intake, g/d	166.0 ± 126.4	168.2 ± 125.3
Vegetable intake, g/d	152.2 ± 76.9	150.1 ± 68.6
Fatty fish intake, g/d	12.3 ± 20.4	13.3 ± 22.3
White fish intake, g/d	15.5 ± 18.5	15.9 ± 17.6
Legumes intake, g/d	28.6 ± 30.2	26.7 ± 26.9
Nuts intake, g/d	2.3 ± 6.5	2.2 ± 5.6
Dairy intake, g/d	222.4 ± 146.0	217.1 ± 142.3
Egg intake, g/d	14.3 ± 17.4	14.0 ± 17.0
Sugar-sweetened beverages intake, g/d	32.9 ± 78.6	30.8 ± 65.5

[π]Values are mean ± SD for continuous variables and n (%) for categorical variables.

3.3.2 Development and validation of metabolite scores for meat consumption

In the exploratory set in the EPIC-Norfolk study, 139 metabolites were identified to be associated with red meat consumption, and they were assembled into a composite red meat metabolite score. This score was made up of 49 (19.3%) lipids and 30 (22.2%) amino acids, other metabolite classes such as xenobiotics (n=14, 12.5%) and 36 (18.4%) unknown metabolites (***Figure 3.2***). The top 5 metabolites with positive coefficients were 1-(1-enyl-stearoyl)-2-arachidonoyl-glycerophosphoethanolamine (GPE) (P-18:0/20:4), 1-(1-enyl-stearoyl)-2-arachidonoyl-glycerophosphocholine (GPC) (P-18:0/20:4), 1-margaroyl-2-oleoyl-GPC (17:0/18:1), trans-4-hydroxyproline, and verapamil. The derived metabolite score for red meat consumption achieved an explained variance of 24% and 17% in the exploratory and validation sets. The metabolite score for red meat intake was associated with quintiles of self-reported meat intake (***Figure 3.3***). It was also significantly higher in the subgroups of self-reported red meat consumers and non-vegetarians, compared to non-consumers of red meat and vegetarians, respectively.

The metabolite scores for processed meat consumption and poultry consumption consisted of 82 and 49 predictive metabolites, respectively, and were made up predominantly of lipids and amino acids (***Figure 3.2***). The overlapping and distinct sets of metabolites that were associated with red meat, processed meat and poultry consumption are shown in *Figure 2*. Six metabolites were included in all three metabolite scores: trans-4-hydroxyproline, trimethylamine N-oxide (TMAO), methionine sulfone, sphingomyelin (d18:2/14:0, d18:1/14:1), N-acetylputrescine, and X-11849. Overall the 7dDD meat intake variances explained by the corresponding metabolite scores in the validation set were 15% for processed meat and 13% for poultry (***Supplementary Figure 3.2***).

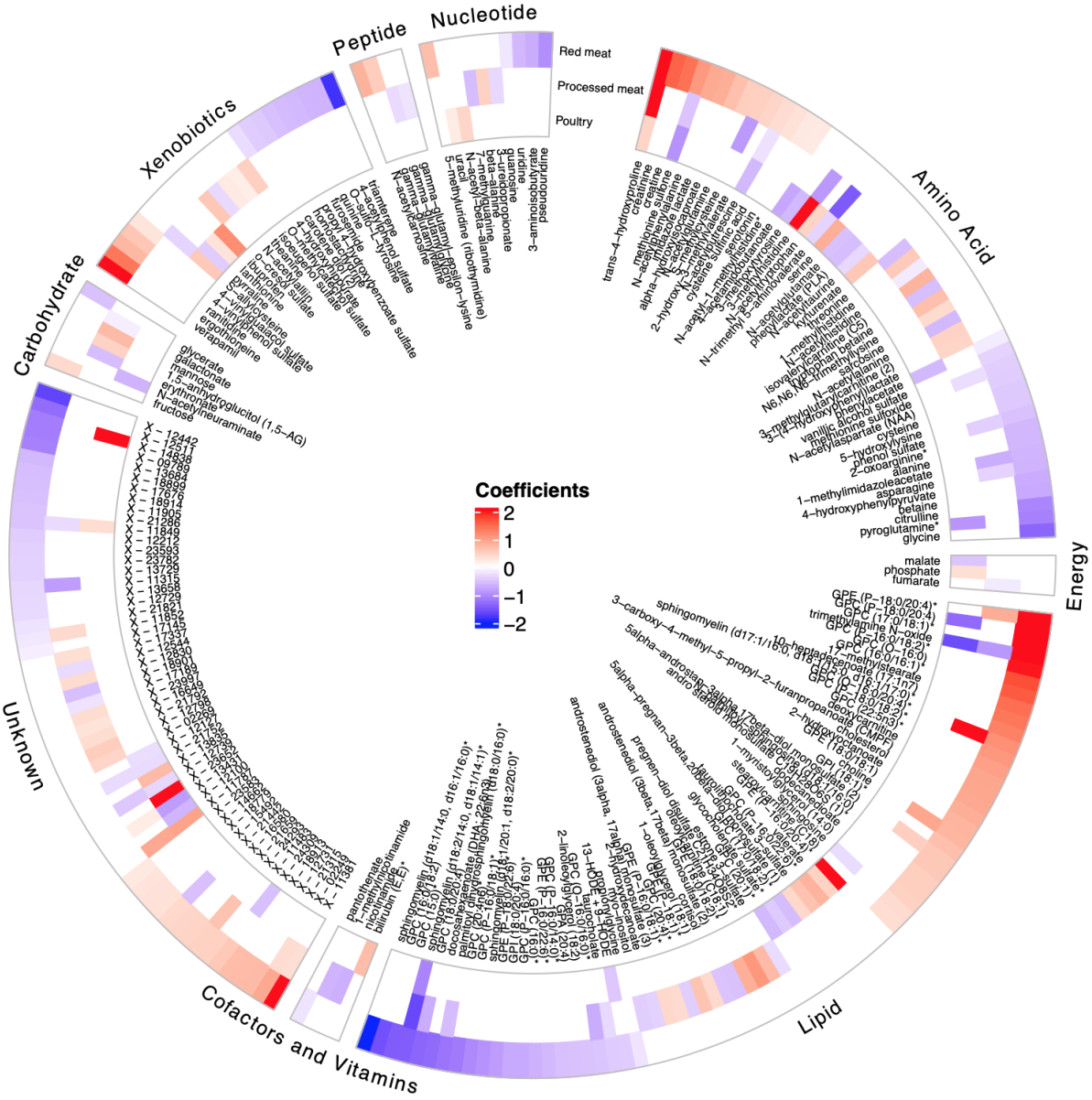


Figure 3.2 Coefficients of metabolites with self-reported red and processed meat and poultry intake: the EPIC-Norfolk study (n=11,432). The colors represent the coefficients (weights) of each metabolite in each metabolite score; red means positive association and blue means negative association.

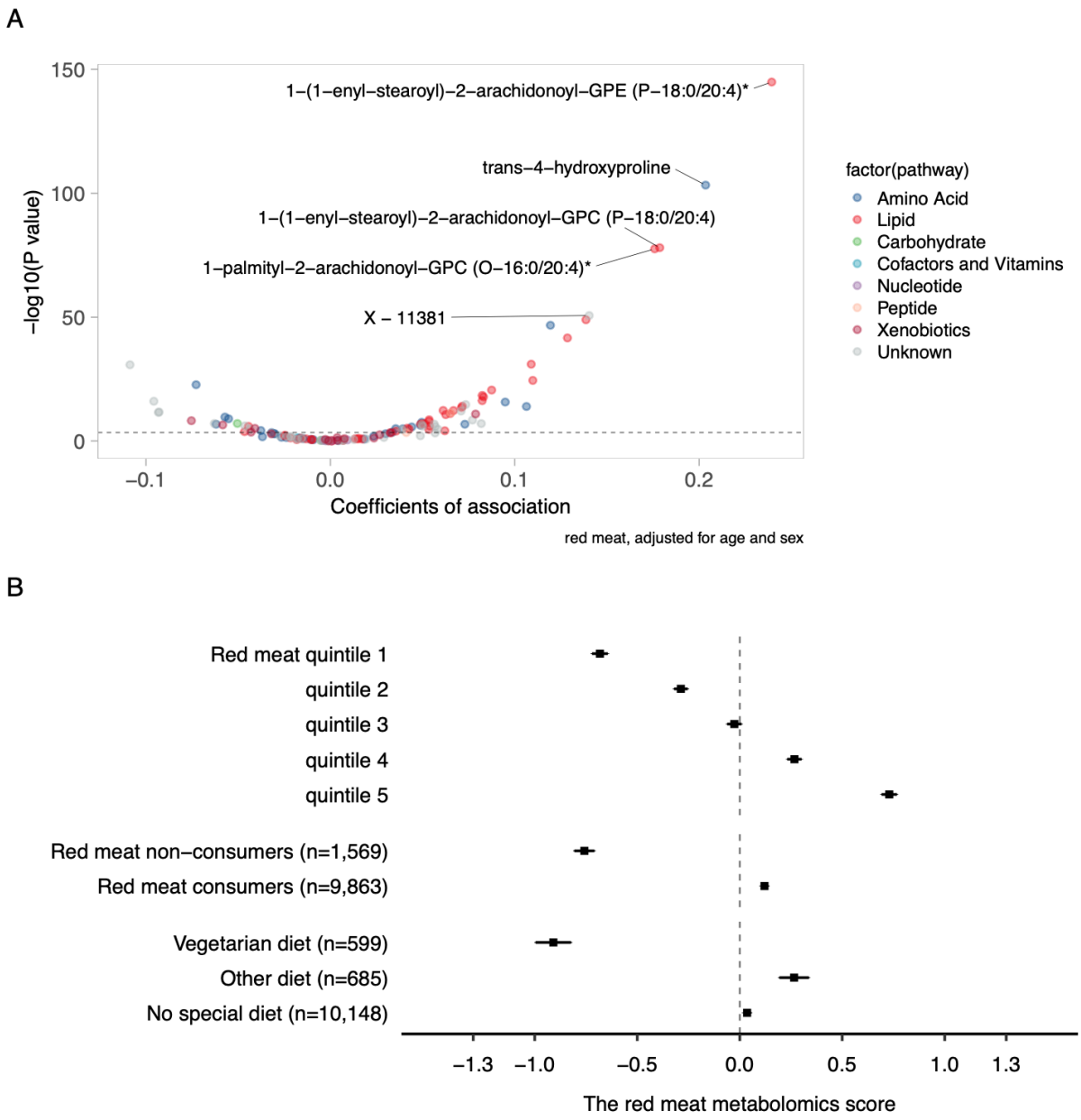


Figure 3.3 Volcano plot of candidate metabolites for red meat intake (n=139) with self-reported red meat intake and comparison of the red meat metabolite score across different categories of meat consumer groups: the EPIC Norfolk study (n= 11,432)

3A. Metabolites had strongest associations (top 5) with self-reported red meat intake after adjusting for age and sex were annotated in the volcano plot; 3B. A red meat non-consumer was defined as a participant with red meat consumption equals to zero (n=1,569) and a red meat consumer was a participants with red meat consumption over zero (n=9,863). Participants with vegetarian diet, other diet or no special diet were identified by self-reported questionnaires

3.3.3 Associations of metabolites in the red meat metabolite score with meat intake in an RCT

For the metabolites that were part of the metabolite score for red meat intake, I used untargeted plasma metabolomics data from a meat RCT to investigate the differences of metabolite concentrations after a 3-day red meat intervention compared to a non-meat diet. Out of the 50 known metabolites positively associated with self-reported red meat consumption in the EPIC-Norfolk study, 11 were identified in the RCT and significantly increased after fried pork (red meat) intake compared to tofu: 4-hydroxyproline, TMAO, stearoylcarnitine, deoxycarnitine, creatine, and several glycerophospholipids (***Table 3.2, Supplementary Figure 3.3 and 3.4***). The correlations between these top-ranking metabolites and types of meat consumption in the EPIC-Norfolk study are shown in ***Supplementary Figure 3.5***. Of the top 8 metabolites that had the highest coefficients in the red meat metabolite score in the EPIC-Norfolk study, 6 were replicated in the RCT.

Table 3.2 Metabolites from the red meat metabolomics score that were positively associated with red meat consumption in both the EPIC-Norfolk and the randomised cross-over trial.

Name	Formula	Fold-change[Π]	*P* value	Chromatographic Method[ɪ]	Retention time, min	Confidence level of identification[§]	MS fragments for identification	Rank[*]
1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4)	$C_{43}H_{78}NO_7P$	2.52	1.36×10^{-6}	RP	9.04, 9.43	Level 2	361.2741 611.5296 392.2934	1
1-(1-enyl-stearoyl)-2-arachidonoyl-GPC (P-18:0/20:4)	$C_{46}H_{84}NO_7P$	2.00	6.69×10^{-6}	RP	9.1	Level 3	184.0733	2
4-Hydroxyproline	$C_5H_9NO_3$	6.27	1.06×10^{-4}	HILIC	5.74	Level 1	68.0498 86.0601	4
TMAO	C_3H_9NO	1.56	6.30×10^{-3}	HILIC	3.62	Level 1	42.0329	7
1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)	$C_{42}H_{80}NO_7P$	1.32	1.94×10^{-4}	RP	8.97	Level 3	184.0733	8
1-palmityl-GPC (O-16:0)	$C_{24}H_{52}NO_6P$	2.01	3.64×10^{-6}	RP	7.18	Level 2	104.1072 184.0770 341.3025	9
Creatine	$C_4H_9N_3O_2$	1.50	4.88×10^{-2}	RP	0.7	Level 1	44.0482 90.0538	13
1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4)	$C_{44}H_{82}NO_7P$	2.44	4.30×10^{-6}	RP	9.04	Level 3	184.0733	17
1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2)	$C_{44}H_{84}NO_7P$	1.96	1.00×10^{-3}	RP	9.19	Level 3	184.0733	18

Following the last page

Name	Formula	Fold-change[Π]	*P* value	Chromatographic Method[ł]	Retention time, min	Confidence level of identification[§]	MS fragments for identification	Rank*
Deoxycarnitine	$C_7H_{15}NO_2$	1.23	6.12×10^{-3}	HILIC	5.18	Level 2	43.0179 60.0811 87.0445	21
Stearoylcarnitine	$C_{25}H_{49}NO_4$	1.52	7.36×10^{-3}	RP	6.47	Level 1	85.0277 60.0813	57

[Π]Fold change in signal intensity in the RCT after fried pork intake compared with the tofu diet. The variation of metabolites intensity after consumption of pork vs tofu is shown in Supplementary Figure 3.

[ł]RP: reverse phase chromatography; HILIC: Hydrophilic Interaction Liquid Chromatography. The chromatographic tracing of selected metabolites in the blood after consumption of pork vs tofu are shown in Supplementary Figure 3.4.

[§]Level of confidence in metabolite identification according to Sumner et al.[175]: level 1, matching of mass, retention time and mass fragmentation pattern with authentic chemical standard; level 2, matching of accurate mass and mass fragmentation pattern with the corresponding compound in a database; level 3, matching of mass and fragmentation pattern with the corresponding compound a database, due to the non-specific fragment, only the functional group, but not the length of each carbon chains can be determined.

*Rank: The rank of coefficients out of 139 metabolites in the red meat metabolite score in the EPIC-Norfolk study.

3.3.4 Association of the red meat metabolite score with T2D

The baseline characteristics of the participants in the T2D case-cohort are presented in ***Table 3.3***. In the subcohort, participants with higher metabolite scores of red meat consumption were more likely to be male, current smokers, have higher BMI, higher consumption of alcohol, sugar-sweetened beverages and total energy, and have lower levels of fruit, legumes, and fish consumption, compared to participants with lower metabolite scores.

Table 3.3 Characteristics of the study participants from baseline of the T2D case-cohort in the EPIC-Norfolk cohort[π]

	Subcohort						T2D cases (n=659)
	Total (n=846)	Q1[ǂ] (n=169)	Q2 (n=169)	Q3 (n=169)	Q4 (n=169)	Q5 (n=170)	
Red meat intake, g/d	33.6 ± 29.1	20.7 ± 20.7	25.8 ± 22.2	29.4 ± 22.7	40.9 ± 28.0	51.3 ± 37.8	39.3 ± 30.6
Age, y	59.0 ± 9.4	59.3 ± 9.5	58.5 ± 9.4	58.9 ± 9.5	59.4 ± 9.3	58.7 ± 9.3	61.8 ± 8.3
Female	489 (58 %)	133 (79 %)	115 (68 %)	89 (53 %)	90 (53 %)	62 (36 %)	275 (42 %)
Education							
No	321 (38 %)	69 (41 %)	62 (37 %)	59 (35 %)	72 (43 %)	59 (35 %)	309 (47 %)
Olevel	79 (9 %)	20 (12 %)	14 (8 %)	17 (10 %)	11 (7 %)	17 (10 %)	54 (8 %)
Alevel	329 (39 %)	60 (36 %)	64 (38 %)	72 (43 %)	66 (39 %)	67 (39 %)	229 (35 %)
Degree	117 (14 %)	20 (12 %)	29 (17 %)	21 (12 %)	20 (12 %)	27 (16 %)	67 (10 %)
Smoking							
Current	112 (13 %)	15 (9 %)	17 (10 %)	19 (11 %)	27 (16 %)	34 (20 %)	79 (12 %)
Former	328 (39 %)	53 (31 %)	59 (35 %)	63 (37 %)	69 (41 %)	84 (49 %)	320 (49 %)
Never	406 (48 %)	101 (60 %)	93 (55 %)	87 (51 %)	73 (43 %)	52 (31 %)	260 (39 %)
Alcohol intake, g/d	11.7 ± 16.7	6.33 ± 8.71	11.0 ± 16.3	12.8 ± 17.0	10.6 ± 15.7	17.8 ± 21.2	11.4 ± 19.0
Physical activity							
Inactive	234 (28 %)	54 (32 %)	37 (22 %)	51 (30 %)	42 (25 %)	50 (29 %)	290 (44 %)
Moderately inactive	244 (29 %)	46 (27 %)	65 (38 %)	39 (23 %)	46 (27 %)	48 (28 %)	157 (24 %)
Moderately active	206 (24 %)	39 (23 %)	37 (22 %)	45 (27 %)	44 (26 %)	41 (24 %)	122 (19 %)
Active	162 (19 %)	30 (18 %)	30 (18 %)	34 (20 %)	37 (22 %)	31 (18 %)	90 (14 %)
BMI, kg/m^2	26.0 ± 3.7	25.3 ± 3.4	26.1 ± 3.9	26.6 ± 3.8	26.0 ± 3.7	26.2 ± 3.7	29.6 ± 4.5
Total Energy, kcal/d	1939.5 ± 516.2	1790.1 ± 433.7	1850.2 ± 444.0	1979.8 ± 543.3	2030.4 ± 559.6	2060.9 ± 536.7	1940.0 ± 538.4
Processed meat intake, g/d	21.7 ± 19.7	16.3 ± 19.2	19.1 ± 17.2	19.5 ± 17.2	25.7 ± 21.5	28.0 ± 20.9	25.1 ± 21.1
Poultry intake, g/d	25.8 ± 25.3	19.6 ± 21.7	27.0 ± 25.6	26.2 ± 25.1	28.0 ± 24.5	28.2 ± 28.3	24.0 ± 26.5
Fruit intake, g/d	167.1 ± 124.0	205.0 ± 138.3	177.2 ± 117.1	171.3 ± 119.0	158.0 ± 128.2	124.2 ± 99.6	151.3 ± 137.1
Vegetable intake, g/d	150.2 ± 68.6	152.1 ± 63.5	149.0 ± 69.8	152.3 ± 72.1	148.1 ± 67.3	147.2 ± 70.5	146.4 ± 80.9
Fatty fish intake, g/d	13.3 ± 22.3	15.9 ± 22.5	15.1 ± 28.9	12.5 ± 17.6	12.2 ± 22.7	10.7 ± 17.7	13.9 ± 27.6
White fish intake, g/d	15.9 ± 17.6	15.1 ± 17.0	13.5 ± 15.0	16.7 ± 18.5	18.4 ± 21.0	15.7 ± 15.8	16.3 ± 22.3
Legumes intake, g/d	26.7 ± 26.9	26.7 ± 27.7	22.8 ± 23.4	26.4 ± 26.3	29.9 ± 31.4	27.6 ± 25.0	28.7 ± 29.8
Nuts intake, g/d	2.2 ± 5.7	2.4 ± 6.2	2.2 ± 4.8	2.2 ± 5.6	1.6 ± 4.5	2.6 ± 6.9	2.0 ± 7.3
Dairy intake, g/d	218.2 ± 142.0	220.0 ± 140.2	210.1 ± 146.4	215.4 ± 132.8	245.9 ± 134.5	197.2 ± 151.7	216.4 ± 158.8
Egg intake, g/d	14.0 ± 17.0	11.9 ± 17.6	12.2 ± 13.3	14.3 ± 17.4	13.9 ± 15.3	17.8 ± 20.3	15.3 ± 17.3
Sugar-sweetened beverages intake, g/d	30.9 ± 65.7	19.9 ± 51.6	31.2 ± 62.1	37.9 ± 74.6	29.3 ± 56.1	36.4 ± 78.8	45.1 ± 127

[π]Values are mean ± SD for continuous variables and n (%) for categorical variables.
[ǂ]Q: the red meat metabolite score in quintiles.

In a prospective analysis with a median follow-up of 10 years, the metabolite score for red meat consumption was positively associated with a higher risk of incident T2D (HR per SD 1.17, 95% CI 1.10 to 1.24) after adjusting for potential confounding factors (***Figure 3.4***). There was a significant association between self-reported red meat consumption and incident T2D (1.08, 1.03 to 1.14).

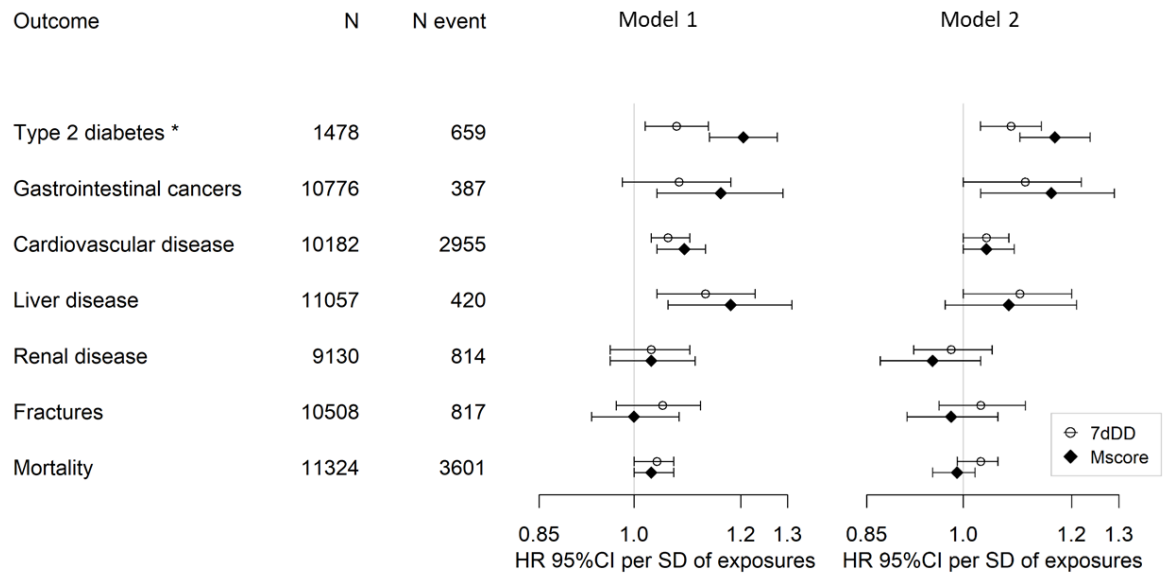


Figure 3.4 The associations of the red meat metabolite score and self-reported red meat intake with incident type 2 diabetes in a nested case-cohort study and exploratory analyses of multiple other health outcomes in the EPIC-Norfolk study.

The regression model 1 adjusted for age and sex; the regression model 2 adjusted for the following potential confounders: age, sex, education, smoking status, alcohol drinking, alcohol drinking squared, body-mass index, body-mass index squared, and dietary factors (consumption of fruits, vegetables, fatty fish and white fish, sugary beverages, dairy, legumes, nuts, eggs and total energy intake). The definition incident cases and exclusion of prevalent cases are reported in *Supplementary Table 3.1*. Abbreviations: 7dDD, 7-day diet diary; CI, confidence interval; Mscore, red meat metabolite score; SD, standard deviation. *, the association with incident type 2 diabetes was conducted in a nested case-cohort study in the EPIC-Norfolk study, associations with other exploratory health outcomes were conducted in the EPIC-Norfolk study after exclusion of participants involved in the case-cohort.

3.3.5 Association of the red meat metabolite score with other health outcomes

In an exploratory analysis, we examined the association of the red meat metabolite score with

six health outcomes. In an adjusted analysis, a higher red meat metabolite score was significantly associated with higher risk of incident cardiovascular disease (1.04, 1.00 to 1.09) and gastrointestinal cancers (1.16, 1.03 to 1.29). The estimates of associations for meat intake using 7dDD measurements were similar to those using the derived scores but the *P* values were generally smaller (*Figure 3.4*).

3.4. Discussion

In this paper I report the development and validation of metabolite scores for three different types of meat consumption: red meat, processed meat and poultry, based on untargeted plasma metabolomics data and 7dDD data in a large British cohort with comprehensive phenotypes. In focused analysis on the red meat metabolite score, I found that eleven top-ranking metabolites in the score were associated with red meat intake in an RCT suggesting a causal link between red meat intake and change of these metabolites. Finally, I found that the red meat metabolite score was associated with T2D incidence and potentially also associated with other cardiometabolic diseases. The association of the metabolite score with these outcomes was comparable with that for self-reported dietary intake.

Metabolite scores of meat consumption

Previous evidence on combining biomarkers into scores to measure meat intake is limited. A feeding trial in Denmark indicated that combinations of several metabolic biomarkers of red meat intake were more efficient than a single biomarker to classify red meat consumers compared to other participants[166]. However, previous studies have not evaluated a dose-response association between meat consumption and a combination of biomarkers. In this large population-based study, I estimated the absolute amounts of meat consumption with 7dDD, which is thought to provide more accurate estimates than an FFQ to rank consumption levels[155]. Our results indicate the utility of untargeted plasma metabolomics to generate an overall score to predict the level of meat consumption rather than only being able to discriminate between consumers and non-consumers.

The metabolite scores of meat consumption were characterised by a wide range of metabolites, including lipids, amino acids, and xenobiotics. Several metabolites that constitute the derived scores have been identified by previous studies, such as TMAO, trans-4-hydroxyproline, creatine, and stearoylcarnitine[164,165,177]. Specifically, an RCT in the United States (n=113) reported that TMAO in plasma significantly increased after red meat consumption compared to consumption of poultry or non-meat products. Positive associations of plasma TMAO levels with risk of cardiovascular disease, diabetes and all-cause mortality have been reported in several meta-analyses of clinical studies[82,85,178]. It suggested that TMAO might be involved as part of underlying mechanisms between red meat intake and

the development of chronic disease. In addition to metabolites in the score of red meat intake, several metabolites specific to processed meat (e.g., o-cresol sulphate)[179,180] or poultry consumption (e.g., 3-methylhistidine)[164] in our study were also reported by previous intervention studies.

I also identified several as yet unreported metabolites that were associated with red meat consumption in both the observational study and the RCT, in particular several plasmalogens, such as 1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4), 1-margaroyl-2-oleoyl-GPC (17:0/18:1) and 1-palmityl-GPC (O-16:0). Plasmalogens, a subclass of membrane glycerophospholipids, contain a vinyl-ether bond at the sn-1 position and are enriched in polyunsaturated fatty acids at the sn-2 position of the glycerol backbone[181]. Mazzilli et al. reported that several plasmalogens were correlated with self-reported red meat [182]. However, most of the plasmalogens identified in our study were not reported in that previous study, partly due to different platforms used to measure and annotate metabolites in different studies. These compounds present a very promising group of potential new biomarkers for meat intake. Their role in meat metabolism and disease development is largely unknown and warrants additional investigation. Some drug metabolites were also identified in the red meat metabolite score, such as verapamil and ranitidine. These metabolites were detected in only a small number of participants, so they could be indicators of participants with chronic disease who were taking these drugs (verapamil for cardiac illness and ranitidine for gastrointestinal illness). These drugs may be the consequence of meat consumption and its association with disease.

One group of metabolites with high contributions to the red meat metabolite score are small meat-derived molecules with short half-lives, such as TMAO, trans-4-hydroxyproline or creatine. These compounds are unlikely to be good long-term biomarkers for rarely consumed foods as they are cleared from the body within one or several days but they may reflect regular red meat consumption well. The second group of metabolites that rank highest in the red meat score are lipophilic compounds, such as plasmalogens. These compounds have half-lives of days or even weeks and can serve as long-term dietary biomarkers[183,184]. They might be useful in the identification of foods that are consumed rarely. These two groups of metabolites in the meat metabolite score ensure that the score reflected not just recent food intake but is indicative of the diet over a longer time frame. This observation also reflected our study

strength of using 7dDDs that captured both short-term and habitual dietary intakes[185].

Associations with T2D risk

The red meat metabolite score, as a proxy for red meat consumption, showed a positive association with incident T2D risk consistent with results from several large cohort studies that have reported associations with T2D risk with self-reported intake as dietary exposures[33,34,40,48]. The score-derived association appeared to be comparable with or stronger than that using 7dDD-measured meat intake. Similar results were reported in a nutritional metabolomics study on a metabolic signature of the Mediterranean diet and its association with risk of cardiovascular diseases[186]. Comparisons between traditional dietary assessment methods and applications of metabolomics are of future interest in various aspects, including clinical utility, cost-effectiveness, and utility to predict disease outcomes and understand pathophysiology.

Strengths and limitations

To our knowledge, this study was the first of this kind to develop and replicate a metabolite score for red meat intake in a large population study which has comprehensive dietary measurements and metabolomics data. Metabolite profiling provided a complementary approach to assess different types of meat consumption objectively. The application of metabolomics to a meat intervention trial provided additional evidence on biological plausibility and reproducibility of the red meat metabolite score. Additionally, in the EPIC-Norfolk prospective cohort study, a long follow-up with detailed information of multiple incident diseases enabled us to examine the association between the red meat metabolite score and multiple health conditions simultaneously.

Several limitations warrant discussion. Firstly, the study was based on a British population so generalisability is limited for other populations and further validation studies should be considered. Secondly, although I have adjusted for a comprehensive set of confounders to examine the association between the red meat metabolite score and risks of non-communicable diseases, the results may be affected by residual confounding. Thirdly, while I have tested the change of metabolites after meat intervention in a trial, the limited number of red meat products and the limited size of the trial hindered a comprehensive validation analysis. The potential causal links between meat intake and most of the candidate

metabolites are largely unknown. Many metabolites in the score are probably not directly influenced by meat intake, but affected by factors that are correlated with meat intake, such as BMI or derived from metabolic or physiological processes. Also, this study might be unable to validate metabolites that reflect long-term diets because the feeding study tested short-term exposures. However, the most important metabolites were validated in the RCT and the score correlated well with meat intake in the validation set. Further validation studies with a wider range of confirmed metabolites in other populations are needed.

In conclusion, this study suggests that a metabolite score derived from untargeted metabolomics profile in plasma has the potential to reflect red meat consumption and inform the association of red meat consumption, assessed objectively, with clinical outcomes.

Chapter 4

Investigation of causal associations between meat-related metabolites and type 2 diabetes

Abstract

Background The 11 metabolites that are included in the meat score in Chapter 3 show the potential for metabolites to predict meat intake and are validated in the external meat intervention study, but the causal nature of associations with T2D has not been assessed.

Methods I studied the associations of 11 metabolites decribed in Chapter 3 with T2D risk. Observational analyses were undertaken using data from the EPIC-Norfolk study of 11,432 individuals with no known history of T2D at baseline and were followed up from baseline (1993-1998) to March 2020. The Cox proportional hazard model was applied to estimate the hazard ratio per standard deviation increase of plasma metabolites. Mendelian randomisation (MR) analyses were performed for the 11 metabolites using the largest published genome-wide association studies (GWASs) and the GWAS performed in the local data. The summary statistics for instrumental single nucleotide polymorphisms (SNPs) for T2D outcome were obtained from DIAbetes Genetics Replication And Meta-analysis GWAS, which includes 898,130 European-descent adults (n cases=74,124) across 32 studies. I used two-sample random-effect inverse variance-weighted and Wald ratio methods to derive MR effect estimates.

Results In observational analyses, 1,188 new-onset cases were reported (median follow-up of 23.5 years). A decreased risk of T2D was reported with 5 metabolites: 1-(1-enyl-stearoyl)-2-arachidonoyl-GPC (P-18:0/20:4), 1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2), 1-palmityl-GPC (O-16:0), and 1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2). The associations for other metabolites were not significant. In MR analyses, genetically predicted high levels of 1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4) was associated with higher risk of T2D (odds ratio 1.11, 95% CI 1.05 to 1.18). An increased risk was also observed for genetically predicted 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4) (1.06, 1.02 to 1.10). Conversely, lower genetically predicted levels of deoxycarnitine were associated with low risk of T2D (0.93, 0.89 to 0.96).

Conclusion This study shows weak evidence of causal associations between meat-related metabolites and incident T2D. This may be due to the limited power of the genetic instruments for meat metabolites.

4.1 Introduction

In the previous Chapter, I identified and validated a metabolite score for red meat intake using observational and interventional data. Among the metabolites that made up for the score, 11 metabolites were associated with red meat consumption in an intervention study, suggesting that these metabolites could be promising biomarkers for assessing meat intake. I also reported that the combination of all meat biomarkers as a metabolite score for red meat intake was associated with the risk of incident T2D. However, how each of these meat-related metabolites is associated with T2D risk has yet been evaluated. Understanding this relationship could help understand the biological pathways that underlie the association between meat consumption and health outcomes.

For some of the meat related metabolites, available evidence from observational and interventional studies has suggested inconsistent associations with T2D. For example, creatine and deoxycarnitine might have protective effects on T2D[187–190]. The association between trimethylamine N-oxide (TMAO) and T2D risk is controversial[85,191–194]. Data on meat biomarkers in the pathway of glycerophospholipids (e.g., plasmalogens) are limited[195,196]. Moreover, it remains uncertain whether these meat-related metabolites were causally related to T2D or whether observational associations were presented due to confounding or reverse causality.

Mendelian randomisation (MR) analysis has been widely applied for causal inference using genetic variants as indicators for the exposure[138,139,197]. Genetic variants are unlikely to be confounded by environmental or lifestyle factors as they are randomly assorted at meiosis, and can avoid reverse causality as the direction of association is from genetic variants to traits. Therefore, genetic variants associated with metabolite levels can be used to examine the aetiological associations between metabolites and disease outcomes such as T2D.

Therefore, in this study, I focused on the roles of 11 meat related metabolites on T2D risk. I first investigated associations of these metabolites with T2D incidence using observational cohort data; then, I used Mendelian randomisation to explore their potential causal roles in T2D by utilising large-scale genetics data.

4.2 Methods

4.2.1 Observational association analyses

Study population

Participants with metabolomics data measured in 2016 and 2017 were included for the observations analyses (n=11,432). The study design, data collection and untargeted metabolomics measurements have been described in **Chapter 1.4.1** and **Chapeter 3.2.2**. The ascertainment of T2D outcomes is reported in **Chapter 2.2.3**.

Statistical analyses

The associations between 11 meat-related metabolites and incident T2D were examined individually using observational data from the EPIC-Norfolk study. Cox proportional hazard model was used to estimate the hazard ratio (HR) and 95% confidence interval (CI) for each metabolite. HRs were reported per 1-SD increase of each plasma metabolite. Potential confounding factors were prespecified based on literature review and clinical knowledge. The model 1 included age, sex, education, physical activity, smoking, dietary total energy intake, and model 2 further adjusted for BMI. The detailed information about the covariates has been described in **Chapter 3.2.4**.

4.2.2 Mendelian randomisation analyses

Genotyping and imputation in EPIC-Norfolk

In EPIC-Norfolk, the participants' genome was genotyped using the UK Biobank Affymetrix Axiom Array and imputed based on Haplotype Reference Consortium[198] and the combined UK10K[199]/1000 Genomes[200] panels. Genetic variants were excluded if the imputation quality INFO score was <0.4 in EPIC-Norfolk.

Genetic instruments selection for metabolites

The summary statistics for 11 candidate metabolites (exposure) were preferentially extracted from published genome-wide association studies (GWASs) if the corresponding sample size for a target metabolite was larger than the EPIC-Norfolk cohort; otherwise, they were

obtained through the conduct of GWAS based on the in-house individual data from EPIC-Norfolk study (n=9,497 with both genetics and metabolomics data). Those published GWASs were (1) Metabolon GWAS: the most extensive GWAS for 913 plasma untargeted metabolites, including 14,296 participants from the EPIC-Norfolk and INTERVAL studies[201]; (2) cross-platform GWAS for metabolites: the largest GWAS for 174 plasma metabolites measured by multiple platforms involving up to 86,507 participants across 6 cohorts. Several metabolites, 1-(1-enyl-stearoyl)-2-arachidonoyl-GPC (P-18:0/20:4), 1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2), 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4), 1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2) and trimethylamine N-oxide (TMAO) were not covered by these two data sources. Therefore, I performed a GWAS for each metabolite using SNPTEST v2.5.2[202]. The model adjusted for age, sex, and population structure of the first 10 principal genetic components.

Genetic instrumental variables for each exposure metabolite were defined as SNP/s that passed a conventional genome-wide significance threshold ($P<5\times10^{-8}$) in the GWAS. Those candidate SNPs were then clumped based on linkage disequilibrium (LD) ($r^2>0.1$) to ensure the independence of included SNPs and to avoid 'double counting' effects. The effect allele was defined as the one associated with increased levels of a metabolite.

Published GWAS for T2D

The GWAS summary statistics for T2D was based on a GWAS meta-analysis conducted by the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium, which includes 898,130 European-descent adults (n cases=74,124) across 32 studies[203]. I extracted summary information from this GWAS, including SNP alleles, effect allele frequency, effect size, standard error, and *P* value. Those instrumental SNPs not available in the T2D GWAS meta-analysis were replaced by proxy SNPs in LD ($r^2 \geq 0.8$) based on 1000 Genomes data of European samples[200].

Two sample MR analyses

The genetic summary statistics on the SNP-metabolite and SNP-T2D associations were harmonised to ensure that effect estimates represented the genetic associations of the same allele[147]. The two-sample mendelian randomisation analyses were performed using the TwosampleMR R package. The inverse variance weighted method was used to estimate each "genetically predicted metabolite level to T2D" association in the primary analyses[138]. The

inverse variance weighted method can provide more precise estimates with relatively improved statistical power, but is susceptible to the pleiotropic effects of genetic variations. Other approaches have recently been developed to account for different assumptions regarding the validity of instrumental variables, such as the weighted median[146], Mendelian randomisation Egger Regression (MR-Egger)[145], and weighted mode methods[148], which were all conducted as sensitivity analyses. The heterogeneity between individual SNP effect estimates was assessed using Cochrane's Q test for inverse variance weighted analyses and Rücker's Q test for MR-Egger analyses. The Wald ratio method was used by default for metabolites with only one SNP.

All analyses were conducted using R (https://cran.r-project.org/).

4.3 Results

4.3.1 Observational associations

In the EPIC-Norfolk study, 11,432 participants were included to study the association between red meat-related metabolites and incident T2D. The characteristics of participants at baseline are shown in ***Table 3.1*** **(Chapter 3)**. During a median follow-up of 23.5 years, 1,188 new-onset cases were reported. Observational associations of these 11 metabolites with T2D risk are presented in ***Table 4.1***. An increased risk of T2D was observed with 4 metabolites in model 1: trans-4-hydroxyproline, TMAO, creatine and 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4). The associations with T2D for trans-4-hydroxyproline and TMAO were attenuated to null after further adjusted for BMI. A decreased risk of T2D was reported with 5 metabolites in both models 1 and 2: 1-(1-enyl-stearoyl)-2-arachidonoyl-GPC (P-18:0/20:4), 1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2), 1-palmityl-GPC (O-16:0), 1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2) and deoxycarnitine. The associations of creatine, 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4) and deoxycarnitine with T2D risk (in model 2) were not significant after multiple correction (P>0.005). Associations of the remaining metabolites with T2D were not significant in either models.

Table 4.1 The associations of the red meat related metabolites with incident type 2 diabetes in the EPIC-Norfolk study (n=11,432)

Metabolite	Model	HR (95% CI)	*P* value
1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4)	model1	1.03 (0.97, 1.09)	4.02E-01
	model2	1.01 (0.96, 1.08)	6.36E-01
1-(1-enyl-stearoyl)-2-arachidonoyl-GPC (P-18:0/20:4)	model1	0.86 (0.82, 0.91)	1.35E-07
	model2	0.90 (0.85, 0.95)	2.16E-04
Trans-4-hydroxyproline	model1	1.08 (1.02, 1.14)	6.35E-03
	model2	1.04 (0.98, 1.10)	1.55E-01
Trimethylamine N-oxide	model1	1.06 (1.01, 1.13)	2.78E-02
	model2	1.04 (0.98, 1.10)	1.98E-01
1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)	model1	0.70 (0.66, 0.74)	1.74E-38
	model2	0.76 (0.72, 0.81)	5.30E-21
1-palmityl-GPC (O-16:0)	model1	0.81 (0.76, 0.86)	2.52E-12
	model2	0.86 (0.81, 0.91)	1.86E-06
Creatine	model1	1.12 (1.05, 1.20)	6.37E-04
	model2	1.07 (1.00, 1.15)	3.68E-02
1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4)	model1	1.07 (1.01, 1.14)	1.78E-02
	model2	1.06 (1.00, 1.12)	6.23E-02
1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2)	model1	0.77 (0.75, 0.80)	4.96E-40
	model2	0.83 (0.80, 0.87)	2.13E-18
Deoxycarnitine	model1	0.90 (0.84, 0.95)	5.74E-04
	model2	0.92 (0.86, 0.98)	7.54E-03
Stearoylcarnitine	model1	0.96 (0.90, 1.01)	1.33E-01
	model2	0.97 (0.91, 1.03)	3.20E-01

Model 1 adjusted for age, sex, education, physical activity, smoking, dietary total energy intake, and model 2 further adjusted for BMI. HR, hazard ratio. CI, confidence Interval.

4.3.2 Identification of genetic instruments for metabolites

For the 11 metabolites related to meat consumption, GWAS summary statistics of 4 metabolites were extracted from the Metabolon GWAS, including creatine, stearoylcarnitine, deoxycarnitine and 1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4); GWAS summary data of trans-4-hydroxyproline were extracted from the cross-platform GWAS; GWAS data for the remaining metabolites were performed in the EPIC-Norfolk study (n=9,497). ***Table 4.2*** presents the information of independent instrumental SNPs for each metabolite, following harmonisation of the instruments information for metabolites (exposures) with the corresponding T2D (outcome) data and LD clumping. The instruments for metabolites were used in the following MR analyses. There was no instrumental SNP for 1-palmityl-GPC (O-16:0) available after LD clumping and alleles alignment between the exposure and the outcome

variables. Given that MR assumes that there is no pleiotropic effect beyond that on the trait of interest, I excluded the SNP (rs3761097) at the pleiotropic *PRODH2* locus from the MR analyses for trans-4-hydroxyproline and TMAO. Therefore, MR analyses of 1-palmityl-GPC (O-16:0) and trans-4-hydroxyproline with the risk of T2D risk were not assessed in this study owing to the absence of instruments for exposures.

Table 4.2 Summary statistics for the genetic instrumental variants of the meat related metabolites and their association with T2D

Metabolite	Instrumental SNP	Chr	Position	Effect/ Other Alleles	EAF	Nearest Gene	Beta of metabolites level per Allele	*P* value	OR (95% CI) for T2D per Allele
Creatine	rs1047891	2	211540507	A/C	0.32	*CPS1*	0.17 (0.01)	6.98E-42	0.99 (0.53, 1.84)
	rs2486274	15	45666228	T/G	0.38	*GATM*	0.11 (0.01)	2.13E-19	0.98 (0.47, 2.06)
Stearoylcarnitine	rs1171617	10	61467182	T/G	0.76	*SLC16A9*	0.12 (0.01)	4.71E-17	0.99 (0.62, 1.57)
	rs603424	10	102075479	A/G	0.18	*PKD2L1*	0.12 (0.02)	1.78E-15	1.02 (0.71, 1.45)
	rs72939920	6	110762453	A/T	0.76	*SLC22A16*	0.23 (0.01)	2.08E-61	0.99 (0.62, 1.59)
Deoxycarnitine	rs10774021	12	349298	T/C	0.66	*SLC6A13*	0.36 (0.01)	3.81E-182	0.98 (0.50, 1.90)
	rs1171614	10	61469538	C/T	0.77	*SLC16A9*	0.16 (0.01)	1.99E-29	0.99 (0.63, 1.57)
	rs34400381	11	65143892	G/A	0.97	*SLC25A45*	0.26 (0.03)	2.03E-15	0.96 (0.90, 1.03)
1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4)	rs102275	11	61557803	T/C	0.65	*TMEM258*	0.25 (0.01)	1.37E-92	1.03 (0.52, 2.02)
	rs4374298	19	55738746	G/A	0.81	*TMEM86B: AC010327.2*	0.09 (0.01)	6.09E-09	1.01 (0.69, 1.48)
Trans-4-hydroxyproline	rs3761097	19	36290977	T/C	0.05	*PRODH2*	0.18 (0.02)	1.54E-16	1.02 (0.98, 1.05)
Trimethylamine N-oxide	rs3761097	19	36290977	T/C	0.05	*PRODH2*	0.23 (0.03)	7.90E-12	1.02 (0.98, 1.05)
	rs77796333	2	65216743	C/T	0.04	*SLC1A4*	0.20 (0.03)	2.19E-08	1.04 (1.00, 1.08)
1-(1-enyl-stearoyl)-2-arachidonoyl-GPC (P-18:0/20:4)	rs174533	11	61549025	G/A	0.34	*MYRF: TMEM258*	0.48 (0.01)	1.00E-200	1.03 (1.01, 1.04)
	rs1885041	14	67976325	C/T	0.48	*TMEM229B*	0.11 (0.01)	2.55E-15	0.99 (0.97, 1.00)
1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)	rs553997864	21	29854726	C/A	1.00	*AF131217.1*	4.46 (0.75)	2.75E-09	0.97 (0.39, 2.39)
	rs964184	11	116648917	C/G	0.87	*ZNF259*	0.12 (0.02)	1.78E-09	0.98 (0.96, 1.00)
	rs9939224	16	57002732	G/T	0.80	*CETP*	0.13 (0.02)	2.33E-13	1.00 (0.98, 1.01)
1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4)	rs174564	11	61588305	A/G	0.66	*FADS2: FADS1*	0.44 (0.01)	1.26E-187	1.03 (1.01, 1.04)
1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2)	rs148086989	20	50560354	C/G	0.99	*RN7SL603P*	0.48 (0.08)	1.65E-09	1.01 (0.94, 1.09)
	rs1864163	16	56997233	G/A	0.75	*CETP*	0.12 (0.02)	3.42E-11	0.99 (0.97, 1.00)
	rs549018368	10	4221248	T/A	1.00	*LINC00702*	2.59 (0.46)	1.29E-08	0.97 (0.54, 1.74)

SNP, single nucleotide polymorphism; Chr: chromosome. EAF, effect allele frequency; OR, odds ratio; SE, standard error.

4.3.3 MR associations

The associations of the nine red meat-related metabolites evaluated in this MR study with the risk of T2D are shown in ***Table 4.3***. The results indicated a possible causal association between 1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4) and T2D (odds ratio (OR) 1.11, 95% CI 1.05 to 1.18, *P*=0.0004) using the inverse variance weighted approach. T2D risk also associated with higher concentrations of 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4) (1.06, 1.02 to 1.10, *P*=0.001). The association of TMAO with T2D risk was not significant after multiple correction (1.13, 1.00 to 1.28, *P*=0.04). The MR analyses also indicated that higher concentrations of deoxycarnitine were associated with a lower risk of T2D (0.93, 0.89 to 0.96, *P*=0.0002). The association for deoxycarnitine with T2D was inconsistent in the MR-Egger analysis after adjustment for pleiotropy, compared with associations using other approaches. The association between genetically predicted creatine with T2D risk was not significant after multiple correction (0.90, 0.84 to 0.97, *P*=0.01).

No evidence of heterogeneity was detected between each SNP effect estimate in analyses for all targeted metabolites. There is little evidence for the presence of pleiotropic effects across the multi-locus (≥3 SNPs) analyses for stearoylcarnitine, deoxycarnitine, 1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2) and 1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2) based on the MR-Egger intercept test. Due to limited intruments (≤ 2 SNPs), other MR approaches apart from IVW were not applicable in associations for creatine, 1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4), 1-(1-enyl-stearoyl)-2-arachidonoyl-GPC (P-18:0/20:4), TMAO, and 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4) with T2D. For the association of 1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4) with T2D risk, asymmetry was observed in the funnel plots of instrument precision, which indicated potential directional pleiotropy.

Table 4.3 Causal effects of mea-related metabolites on T2D risk

	OR	95% CI	*P* value
Creatine			
Inverse variance weighted	0.90	(0.84, 0.97)	0.01
Stearoylcarnitine			
Inverse variance weighted	0.99	(0.90, 1.08)	0.75
Weighted median	0.97	(0.91, 1.05)	0.47
Weighted mode	0.97	(0.90, 1.04)	0.48
Simple mode	0.95	(0.87, 1.04)	0.38
MR Egger	0.95	(0.65, 1.39)	0.83
Deoxycarnitine			
Inverse variance weighted	0.93	(0.89, 0.96)	0.0002
Weighted median	0.93	(0.90, 0.97)	0.001
Weighted mode	0.93	(0.90, 0.97)	0.08
Simple mode	0.94	(0.88, 1.00)	0.18
MR Egger	0.93	(0.82, 1.06)	0.46
1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4)			
Inverse variance weighted	1.11	(1.05, 1.18)	0.0004
Trimethylamine N-oxide			
Inverse variance weighted	1.13	(1.00, 1.28)	0.04
1-(1-enyl-stearoyl)-2-arachidonoyl-GPC (P-18:0/20:4)			
Inverse variance weighted	1.05	(0.96, 1.14)	0.30
1-(1-enyl-palmitoyl)-2-linoleoyl-GPC (P-16:0/18:2)			
Inverse variance weighted	0.94	(0.85, 1.03)	0.20
Weighted median	0.96	(0.86, 1.08)	0.50
Simple mode	0.98	(0.85, 1.14)	0.84
Weighted mode	0.98	(0.85, 1.13)	0.78
MR Egger	1.00	(0.80, 1.25)	0.99
1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4)			
Wald ratio	1.06	(1.02, 1.10)	0.001
1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2)			
Inverse variance weighted	0.96	(0.88, 1.05)	0.40
Weighted median	0.98	(0.88, 1.09)	0.69
Weighted mode	1.01	0.87, 1.17)	0.91
Simple mode	1.00	(0.87, 1.16)	0.97
MR Egger	1.03	(0.88, 1.21)	0.77

Odds ratios (ORs) of T2D are presented per standard deviation increase in candidate metabolites as estimated by two-sample Mendelian randomisation methods. MR-Egger, Mendelian randomisation Egger. SNP, single nucleotide polymorphism.

4.4 Discussion

In this study, I characterised the associations of 11 meat-related metabolites with T2D incidence using observational data and reported that 4 metabolites were associated with increased risk of T2D and 5 metabolites were inversely associated with T2D risk. However, all positive metabolite-T2D associations were attenuated to the null after adjustment for BMI with correction for multiple tests. I then investigated potential causal effects of meat-related metabolites on T2D risk using MR. An increased T2D risk was associated with of higher concentrations 1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4), 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4), and lower concentrations of deoxycarnitine.

I found suggestive evidence that genetically predicted 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4) was associated with T2D risk. The effect was driven by the SNP (rs174564) in the *FADS1* and *FADS2* (fatty acid desaturase 1 and 2) gene sets. The *FADS* gene sets encode delta-5- and delta-6-desaturases enzymes that regulate the biosynthesis of long-chain polyunsaturated fatty acids (PUFAs)[204,205]. PUFAs are known to activate peroxisome proliferator-activated receptor-gamma (PPAR-γ), which is associated with the pathogenesis of diabetes[206]. The genetic polymorphisms in *FADS* gene (e.g., rs174575) were associated with desaturase activity in participants of Caucasians and Asian ancestries[207], and have been linked to the risk of diabetes in different populations (n>400)[208–210]. A study in an Iranian population did not find a significant association between rs174583 (*FADS*) and diabetes, possibly due to the specific effect of that SNP of interest or a smaller sample size (~200), compared with other studies[211]. In this study, I further tested the LD correlation of variants rs174564 and rs174575 on *FADS*1/*FADS*2 loci and found that they are moderately correlated (r^2=0.62). This suggests that the likely causal association between 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4) and T2D might be affected by a pleiotropic effect of *FADS*1/*FADS*2 gene sets on fatty acids, although the LD is not very high. The effects of rs174564 in the function of *FADS* gene expression, 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4) metabolism and the pathogenesis of T2D need further investigation in more studies.

To our knowledge, this study is the first that has reported the association between 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4) and T2D risk in both observational and genetics studies. 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4) is a phosphatidylcholine in the plasmalogen

pathway. In the MR analysis, another plasmalogen, 1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4), also presented a potential causal association with an increased risk of T2D. Plasmalogens play important roles in multiple biological functions, such as membrane construction and anti-oxidation[212]. Plasmalogens have been linked with neurological and metabolic diseases in lipidomics and metabolomics studies. However, conflicting evidence exists in the literature and the causal effects of plasmagens on the development of diseases have not been established[196,213,214].

TMAO has attracted attention given its emerging effects in the development of cardiovascular disease and its enrichment in animal-sourced foods, particularly red meat. The evidence for the impact of TMAO in the pathogenesis of T2D was inconclusive based on observational studies. A recent Mendelian randomisation study of TMAO and T2D reported that T2D can increase the levels of TMAO[215]. However, the instruments for TMAO in that MR study were based on a small GWAS (n=2,076) and not all of the instruments reached conventional statistical significance thresholds ($P<5 \times 10^{-8}$)[216]. In this study, I used independent instruments for TMAO extracted from the largest GWAS (n=9,497) and found a weak association between genetically predicted TAMO levels and the risk of T2D (*P* value=0.04). However, the positive association did not persist after correction for multiple tests.

MR is a useful approach for causal inference and one key condition to consider is the validity of the instruments. For metabolites (e.g., deoxycarnitine) with multiple instrumental SNPs (n≥3), the results were roughly consistent across different MR methods based on different assumptions of horizontal pleiotropy, suggesting that horizontal pleiotropy is less likely to driven the findings. The MR-Egger intercept term test also suggested there is no evidence of pleiotropy. However, an assessment of the presence of pleiotropy was not possible for metabolites with a small number of instruments (n≤2) because I was unable to perform different MR methods and it is difficult to interpret asymmetry in instrument precision. Therefore, the interpretation for the causal roles of these metabolites in T2D risk need to be treated with caution and need further investigation.

This study has several strengths. First, I used genetic information as the instrumental variable for the exposure which is unlikely to be affected by confounding factors and reverse causation, compared with conventional observational study designs. Second, the largest possible GWAS studies were used to identify instruments for the exposure and model the outcome.

However, this study has a few limitations. In the observational analysis, the participants included were quansi-randomly selected from the EPIC-Norfolk after excluding participants who were included in a case-cohort for T2D in 2005. This means that the study population represents those who were relatively healthy without the development of T2D during the first 5-10 years of follow-up. Although I have leveraged the largest possible GWAS data to identify instruments for metabolites of interests, the current GWASs for metabolites still have limited power compared to GWASs for other traits (e.g., n=~20k for metabolomics GWAS vs. n>500k for BMI GWAS). The small number of instruments available for exposures might limit the ability to observe the effects of metabolites on T2D risk. Furthermore, there is considerable growth in the number of metabolome genetic studies in the literature. Future well powered biobank data will improve the understanding of the genetic basis for metabolite profiling, which will then enable further investigation of the potential causal effects of metabolites in the pathogenesis of diabetes. In this study, I focused on the causal role of each red meat related metabolites on T2D. These mmetabolites might be correlated with each other and related to other intermediate traits, such as glycaemic traits, lipid levels and anthropometric measurements. Therefore, it might be helpful to investigate the roles of other related traits in the link between metabolites and T2D risk.

In summary, I investigated potential causal associations of red meat related metabolites with T2D risk using Mendelian randomisation. The results suggested that TMAO and several plasmalogens are likely to play causal roles in the pathogenesis of T2D. The causal inference of these red meat biomarkers on the risk of T2D needs to be interpreted with caution given the weak genetics instruments used in the current study. Further genetic and functional research is warranted to explain the underlying mechanisms of these metabolites in the aetiology of T2D.

Chapter 5

Differential impacts of meat consumption on type 2 diabetes risk in population subgroups: in EPIC-InterAct and UK Biobank

Abstract

Background High meat intake has been associated with the risk of developing type 2 diabetes (T2D) in the general population, but whether this association differs in sub-populations with varying genetic and clinical risks is understudied. This study aimed to investigate the association between red meat intake and incident T2D within genetic and clinical risk groups.

Methods Participants from the EPIC-InterAct and UK Biobank (UKBB) prospective cohorts were followed from the recruitment until 2007 and 2020, respectively. Three generic risk scores (GRSs) for T2D, insulin resistance and BMI, and two clinical indexes (HbA1c and Cambridge diabetes risk score) were then used to define subgroups with varying baseline risks. The GRSs were constructed based on significant single nucleotide polymorphisms (SNPs) from published genome-wide association studies. Prespecified confounders were adjusted using multivariate Prentice-weighted Cox and standard Cox models. Both hazard ratios (HRs) and absolute risk increases (ARIs) over the follow-up were calculated.

Results Of 20,628 and 316,222 participants in EPIC-InterAct and UKBB, 9,086 and 10,518 new-onset T2D cases were identified during a median follow-up of 10.3 and 11.9 years, respectively. After meta-analyses, higher red meat intake was associated with incident T2D (HR 1.09, 95% CI 1.01 to 1.17). The interactions between meat and genetic or clinical risks were not significant. Compared with those with low meat intake and low risks, having a high meat intake combined with a high Cambridge risk score, HbA1c or T2D genetic susceptibility was associated with a 12-fold, 6-fold or 3-fold increased risk of T2D, respectively.

Conclusion In the two large population-based cohorts, red meat intake was associated with a higher risk of incident T2D independently of genetic and clinical risks. These findings support the necessity of encouraging a reduction of meat intake to prevent T2D in the entire population.

5.1 Introduction

Nutrition plays a vital role in the improvement of public health and prevention of diseases. Red meat, one of the key dietary factors, has been linked with increased risk of chronic diseases, including type 2 diabetes (T2D)[33,38,55,80,217]. Dietary guidelines have been applied to limit the amount of red meat consumption in many populations in order to reduce disease burdens. However, applying a one-fit-for-all dietary intervention in the whole population remains challenging with limited impact in real world due to complex reasons, such as societal influences, discrepancies in adherence to the dietary guidance, food availability and affordability. Moreover, there are differences in people's response to foods that lead to dissimilarities in disease progress even if they have eaten the same diet.

More effective dietary intervention strategies in populations are needed. Tailored dietary advices in subgroups with high risks could be a more effective and pragmatic approach in altering dietary behaviours and these subgroups might benefit more from a targeted prevention[218,219]. However, in currently available trials, the extent to which changes in diet can benefit health outcomes is unclear, partly because it is challenging to carry out a trial of personalised nutrition intervention over a sufficiently long follow-up time with enough power to evaluate its effect on health.

Whether the effects of red meat intake on incident T2D varies across different subgroups in large observational cohort studies is unknown. A recent study reported an association between diet quality and incident T2D among individuals with high genetic predisposition to T2D, but not among those with low genetic risks[220]. A previous study found high meat intake was associated with increased T2D risk in a subgroup with high T2D genetic risk scores defined by 10 Single nucleotide polymorphisms (SNPs), but the study was conducted in men only[221]. Whether the association between red meat intake and incident T2D is different in subpopulations with different T2D risks needs to be investigated in larger studies with more comprehensive information for population classification.

In this study, we aimed to examine the effects of red meat intake on incident T2D by subgroups defined by genetic or clinical risk of T2D in two large prospective studies of middle aged adults of European populations.

5.2 Methods

5.2.1 Study populations

The analyses of this Chapter used data from two large population-based studies: the European Prospective Investigation into Cancer and Nutrition-InterAct (EPIC-InterAct) and the UKB biobank (UKBB). Study designs and general inforamtion of the two studies have been described in **Chapter 1.4.2** and **Chapter 1.4.3**. EPIC-InterAct is a large case-cohort study of incident T2D, consisting of a case group of 12,403 ascertained incident T2D and a sub-cohort of 16,154 individuals as the control group who were representative of all eligible pan-Europe EPIC participants (n=340,234) who had diabetes information and stored blood samples[114]. UKBB is a prospective cohort of over half million individuals recruited from England, Scotland and Wales between 2006-2010[120,121]. For current analyses, participants of each study were excluded if they had either 1) missing information in dietary intake, genetic and covariates data, 2) invalid dietary records, and 3) prevalent diabetes at baseline.

5.2.2 Dietary assessment

In EPIC-InterAct, individual habitual food intake was assessed using country-specific dietary questionnaires at baseline, which included up to 260 food items. The questionnaires were validated within each country and country-specific dietary data were standardised using a 24-hour dietary recall program (EPIC-SOFT) to provide comparable dietary data across participating countries[114–118]. The main exposure of red meat was calculated as a sum intake of unprocessed beef, lamb or mutton, and pork in the unit of grams per day (g/d). In UKBB, dietary consumption frequency of 29 food groups was evaluated using a touchscreen dietary questionnaire at recruitment for all participants. Besides, about 200k UKBB participants were followed up via emails to complete at least one online 24-hour dietary recall questionnaire, in which the actual amount of diet consumption was collected. The performance of the touchscreen dietary questionnaire was evaluated by a previous study showing a good agreement with the 24-hour recall data regarding the ability to rank participants' food consumption[122]. Quantitative red meat consumption (in g/d) was calculated directly for participants who had at least two complete 24-hour recall questionnaires and a mean value

was assigned to those without 24-hour dietary data according to their touchscreen report category in the meat category.

In this study, we defined meat consumption below 33th percentile over each cohort (the subcohort was used for EPIC-InterAct) as low meat intake, and that above the 67th percentile as high meat intake.

5.2.3 Genotyping and polygenic risk scores

Participants from EPIC-InterAct were genotyped using two array-based chips: Illumina 660W-Quad Bead chip (553,115 variants, 10,023 participants) and Illumina core-exome chips (366,044 variants, 13,474 participants). Genetic data were imputed using Haplotype Reference Consortium (HRC) reference panel[198]. Similarly, the UKBB genotype data were assayed using two closely related chips, UK BiLEVE Axiom Array (807,411 variants, n=49,950 participants) and UK Biobank Axiom Array (825,927 variants, n=438,427 participants)[121]. The two arrays are very similar; hence the genotype data were merged for further analyses. Missing genotypes in UKBB were imputed using the UK10K and 1000 Genomes reference panels[199,200]. Information of genotyping, imputation, sample quality control in EPIC-InterAct and UKBB has been detailed elsewhere [121,222].

In this study, we constructed three genetic risk scores (GRSs) for T2D[203], insulin resistance[223] and BMI[224], respectively, based on significant SNPs identified from published large genome-wide association studies (GWASs). The SNPs included in each scores are provided in ***Supplementary Table 5.1***. In general, the GRS was calculated by summing the number of candidate risk alleles, weighted by their relative effect sizes extracted from the reference GWAS[225]. The scores were then standardised to a mean value of 0 and standard deviation of 1 in each cohort for the following analyses.

5.2.4 HbA1c and Cambridge diabetes risk score

We used glycated hemoglobin (HbA1c) and the Cambridge diabetes risk score to classify individuals' clinical risk for T2D. HbA1c levels were measured from erythrocyte samples in EPIC-InterAct[226] and UKBB[227], using ion exchange high-performance liquid chromatography

(HPLC) on a Tosoh G8 and HPLC analysis on a Bio-Rad VARIANT II Turbo, respectively. The Cambridge diabetes risk score was developed as an effective and simple tool for identification of individuals at risk of developing diabetes. The score is made up of several simple non-biochemical information that are available from routinely collected data in General Practices, including age, sex, prescribed antihypertensive medication, prescribed steroids, BMI, family history of diabetes, and smoking status. The details on construction of the score are given in a previous study[228].

5.2.5 Outcome ascertainment

Incident T2D were identified by repetitive follow-up surveys and by linkage to multiple routinely collected health databases, including general practitioners' primary-care practices, hospital secondary-care, national diabetes registry and death registry, wherever available for different countries. Criteria for defining T2D cases from each databases are described in the ***Supplementary Information***. Follow-up was censored at the date of T2D occurrence, 31 December 2007 for EPIC-InterAct, 30 November 2020 for UKBB, or death, whichever occurred first. Participants with any historical T2D records before the baseline were excluded.

5.2.6 Assessment of covariates

We prespecified a list of potential confounding factors, including sociodemographic factors (age, sex), smoking status (never, former, current, unknown), physical activity, body mass index (BMI), and family history of diabetes. Total energy intake was estimated from European food composition tables in EPIC-InterAct, whereas it was not available in UK Biobank.

5.2.7 Statistical analysis

We described baseline characteristics of eligible participants overall and by different meat intake categories in tertiles, with counts (percentages) for categorical variables, and mean (standard deviation) for continuous variables. For the overall associations between meat consumption and incident T2D, we estimated the hazard ratio (HR) using Prentice-weighted Cox proportional hazard (PH) regression[176] in the EPIC-InterAct case-cohort while standard

Cox PH regression[156] was used in the UKBB cohort. We simultaneously adjusted for age, sex, genetic ancestry (the first 10 principal components), smoking status, physical activity, BMI, and family history of diabetes in both cohorts and further included country and total energy intake in the modelling of EPIC-InterAct data. The interaction effects between meat intake and prespecified genetic and clinical risk scores were examined based on the multiplicative scale by including a product term of meat intake and each study score. The results from the two studies were meta-analysed using a random effects meta-analysis. We considered two-sided *P* values less than 0.01 (*P* values of less than 0.05 divided by the number of tests, i.e., 0.05/5) as statistically significant.

To identify potential high incident T2D burden groups associated with high meat intake, we first stratified participants into three sub-groups according to their baseline T2D risk defined by each genetic and clinical score. The cut-off of each group in EPIC-InterAct was calculated using the subcohort. The HbA1c groups were classified according to thresholds for diagnosing diabetes(<42 mmol/mol as the normal group, ≥42 and <48 mmol/mol as the high group. Participants with a HbA1c≥48 mmol/mol were not included in the analysis) as they had prevalent but undiagnosed diabetes at baseline[7]. We then estimated HR and absolute risk increase (ARI) within subgroups using the formula $I_0 \times (HR - 1)$, where I_0 is the incidence in the reference group (normal meat intake). The covariates adjusted were the same in each subgroup, except for those subgroups stratified by Cambridge risk score where age, sex, BMI, family history of diabetes, and smoking status were already incorporated in the generation of the score and thus were not included in the Cox model. In EPIC-InterAct, to estimate the cumulative incidence of T2D within strata defined by subgroups of the genetic or clinical risk factors in EPIC-InterAct, we recreated the full cohort by resampling with replacement from the subcohort, according to the distributions of the stratum variables within the subcohort[229].

We further studied the joint effect of meat intake with each genetic and clinical risk factor on T2D incidence (9 categories with low meat intake and low baseline risk as the reference). Statistical significance was determined by a 95% confidence interval (CI) that excluded 1 for HR and 0 for ARI. Analyses were predominantly performed using R and Stata (version 15) was used for the estimation of absolute risks.

5.3 Results

Table 5.1 shows the baseline characteristics of the participants from the EPIC-InterAct subcohort and UKBB. The populations included 20,628 unrelated participants in EPIC-InterAct (mean age 53.9 years, 57% female) and 316,222 in UKBB (mean age 56.5 years, 55 % female). Individuals with high meat intake were more likely to be male, be physically active, be current or former smokers, and have a family history of diabetes. A total of 9,086 T2D cases were observed in EPIC-InterAct (median follow-up 10.3 years) and 10,518 cases in UKBB (median follow-up 11.9 years).

Table 5.1 Characteristics of participants at baseline in total and by tertiles of meat intake in the EPIC-InterAct (n=20,628) and the UK Biobank (n=316,222) studies

Characteristics	EPIC-InterAct				UK Biobank			
	Total (n=20,628)	Low (n=6,373)	Medium (n=6,863)	High (n=7,392)	Total (n=316,222)	Low (n=31,065)	Medium (n=125,808)	High (n=159,349)
Age at recruitment, y	53.9 ± 8.70	52.7 ± 9.55	54.3 ± 8.82	54.6 ± 7.65	56.5 ± 8.02	54.8 ± 8.18	56.3 ± 8.04	57.0 ± 7.93
Female	11839 (57 %)	4330 (68 %)	4260 (62 %)	3249 (44 %)	173317 (55 %)	21223 (68 %)	71030 (56 %)	81064 (51 %)
BMI, kg/m²	27.5 ± 4.8	27.0 ± 5.0	27.5 ± 4.7	27.9 ± 4.6	27.1 ± 4.6	25.9 ± 4.6	27.0 ± 4.6	27.5 ± 4.6
Red meat intake, g/d	47.8 ± 36.7	12.5 ± 7.4	38.5 ± 8.4	86.7 ± 31.4	118.0 ± 57.3	24.4 ± 0	74.8 ± 0	170.0 ± 24.6
Smoking status								
Never	9121 (44 %)	3357 (53 %)	3114 (45 %)	2650 (35 %)	173556 (55 %)	17670 (57 %)	70019 (56 %)	85867 (54 %)
Former	5954 (29 %)	1672 (26 %)	1977 (29 %)	2305 (31 %)	110131 (35 %)	10602 (34 %)	43446 (35 %)	56083 (35 %)
Current	5553 (27 %)	1344 (21 %)	1772 (26 %)	2437 (33 %)	32535 (10 %)	2793 (9 %)	12343 (10 %)	17399 (11 %)
Physical Activity*								
Inactive	5289 (26 %)	1872 (29 %)	1837 (27 %)	1580 (21 %)				
Moderately inactive	6903 (33 %)	2153 (34 %)	2324 (34 %)	2426 (33 %)	46647 (15 %)	3895 (13 %)	18623 (15 %)	24129 (15 %)
Moderately active	4502 (22 %)	1348 (21 %)	1501 (22 %)	1653 (22 %)	106160 (34 %)	9954 (32 %)	42382 (34 %)	53824 (34 %)
Active	3934 (19 %)	1000 (16 %)	1201 (17 %)	1733 (23 %)	105396 (33 %)	11582 (37 %)	41449 (33 %)	52365 (33 %)
Missing					58019 (18.3%)	5634 (18.1%)	23354 (18.6%)	29031 (18.2%)
Family history of diabetes								
No	18178 (88 %)	5767 (90 %)	6036 (88 %)	6375 (86 %)	256012 (81 %)	25217 (81 %)	101604 (81 %)	129191 (81 %)
Parent or sibling	2082 (10 %)	520 (8 %)	697 (10 %)	865 (12 %)	54476 (17 %)	5329 (17 %)	21903 (17 %)	27244 (17 %)
Parent and sibling	368 (2 %)	86 (1 %)	130 (2 %)	152 (2 %)	5734 (2 %)	519 (2 %)	2301 (2 %)	2914 (2 %)
Hypertension treatment	3592 (17 %)	1125 (18 %)	1279 (19 %)	1188 (16 %)	26792 (8 %)	2432 (8 %)	10766 (9 %)	13594 (9 %)
Lipid lowering treatment	721 (3 %)	222 (3 %)	270 (4 %)	229 (3 %)	17523 (6 %)	1614 (5 %)	7121 (6 %)	8788 (6 %)

*Low, 0-0.9 times/week, medium, 1-3.9 times/week; high, ≥4 times/week. Values are mean ± SD for continuous variables and n (%) for categorical variables.

5.3.1 Associations with incident T2D

In both studies, compared to low red meat consumption, higher meat consumption was associated with higher risk of T2D, with a HR of 1.23 (95% CI 1.08 to 1.40) in EPIC-InterAct and 1.02 (0.93 to 1.12) in UKBB for high meat intake; and 1.16 (1.04 to 1.29) in EPIC-InterAct and 0.97 (0.88 to 1.06) in UKBB for intermediate intake. We also tested associations with incident T2D for GRSs for T2D, insulin resistance and BMI, as well as the Cambridge diabetes risk score and HbA1c. All of these T2D risk indexes except BMI GRS were associated with risk of T2D (***Supplementary Table 5.2***) and were therefore taken them forward for subsequent analyses.

In subgroup analyses, the relative risk of meat intake on incident T2D was comparable across strata defined by genetic and clinical risks (***Figure 5.1, Supplementary Table 5.3***) in EPIC-InterAct and UKBB, and the effect estimates were stronger in EPIC-InterAct than those in UKBB, apart from associations between HbA1c and T2D risk. We did not observe any evidence of significant interactions between red meat intake and potential modifying factors in either study or in the pooled analysis.

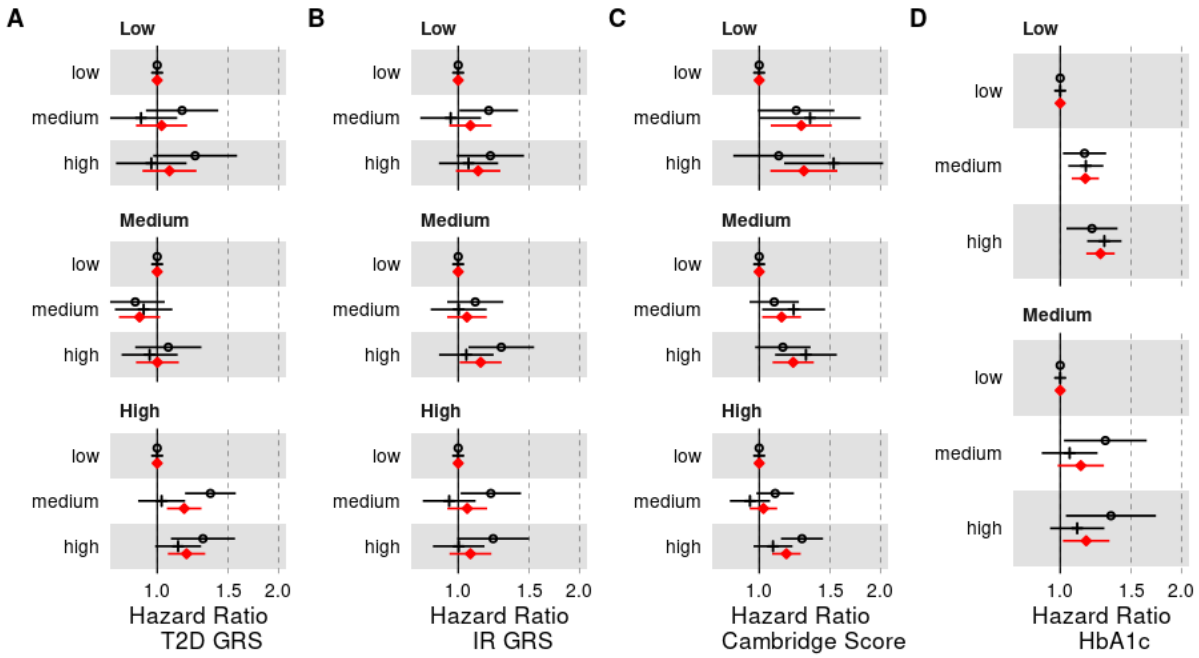


Figure 5.1 Relative risk (adjusted hazard ratios) for T2D in subpopulations defined by meat intake and other risk factors in the EPIC-InterAct (n=20,628) and the UK Biobank (n=316,222) studies. In these comparisons, participants with low meat intake within each subgroups defined by tertiles of genetic or clinical risk factors were regarded as the reference group. The HbA1c groups were classified according to thresholds for diagnosing diabetes (<42 mmol/mol as the low group, ≥42 and <48 mmol/mol as the medium group, ≥48 mmol/mol were not included for analysis). There was no evidence of a significant interaction (multiplicative) between meat intake and genetic or clinical risk factors in both studies. GRS, genetic risk score; IR, insulin resistance; Cambridge Score, Cambridge diabetes risk score. The black circle shows results in EPIC-InterAct, the black cross shows results in UK Biobank and the red circle shows results in a random effects meta-nalaysis.

5.3.2 Absolute risk of meat intake with T2D by groups

We estimated cumulative incidence rates and ARI of T2D by strata (***Figure 5.2***, ***Supplementary Figures 5.1 and 5.2, and Supplementary Table 5.4***). The ARIs of T2D were increased with baseline diabetes risks, and were more apparent in the groups with a high T2D risk defined by clinical indexes (HbA1c and Cambridge diabetes risk score) and T2D GRS. For example, in EPIC-InterAct, the cumulative incidence rates of developing T2D over 10 years in the high HbA1c group (prediabetes) were 7.7%, 9.7% and 10.8% across groups with low, intermediate and high meat intake, compared to 1.0%, 1.3% and 1.6% in the normal HbA1c group. The ARI between high and low meat intake after adjustment for multi-covariates were 2.6% in the high HbA1c group compared to 0.2% in the normal HbA1c group in EPIC-InterAct. For the Cambridge diabetes risk score, participants had considerable higher incidence rates of T2D (4.7%, 5.2% and 5.6%) in subgroups with low, intermediate and high meat take, compared those with low scores (0.4%, 0.5% and 0.5%). The adjusted ARIs between high and low meat intake were 1.3% vs. 0.04% in high vs. low T2D risk groups defined by the Cambridge diabetes risk score, 0.9% vs. 0.2% for T2D GRS and 0.5% vs. 0.3% for insulin resistance GRS (***Supplementary Table 5.4***).

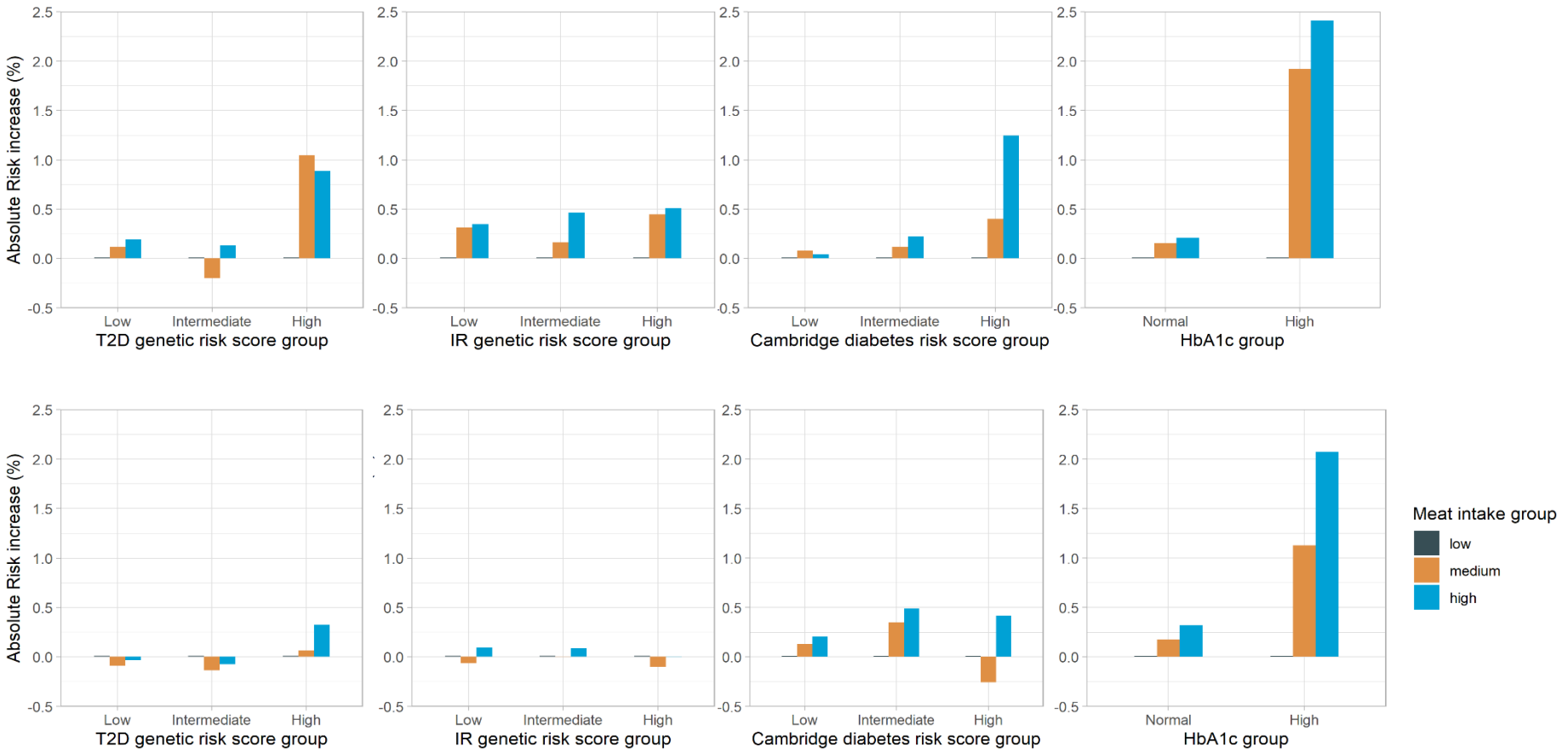


Figure 5.2 Adjusted absolute risk differences in the EPIC-InterAct (n=20,628) and the UK Biobank (n=316,222) studies. Differences of 10-year cumulative incidence in subgroups by meat intake and genetic or clinical risk factors are presented. The upper panel shows results in EPIC-InterAct and the lower panel shows results in UK Biobank. In these comparisons, participants with low meat intake within each subgroups defined by tertiles of genetic or clinical risk factors were regarded as the reference group. The estimates adjusted for age, sex, centres, smoking, physical activity, BMI, total energy intake and family history of diabetes. The HbA1c groups were classified according to thresholds for diagnosing diabetes (<42 mmol/mol as the normal group, ≥42 and <48 mmol/mol as the high group, ≥48 mmol/mol were not included for analysis).

5.3.3 Joint associations of meat intake and genetic or clinical risks with incident T2D

Figure 5.3 shows the combined association of meat intake with genetic or clinical risk on the risk of incident T2D. An additive effect on T2D risk was found for meat intake and genetic or clinical risks. The highest risk was observed in individuals with high meat intake and high genetic or clinical risk, especially for those with high Cambridge diabetes risk scores. For example, having high meat consumption and a high Cambridge diabetes risk score was associated with 12-fold high risk of T2D (HR 12.5, 95% CI 10.9 to 14.4), compared with low meat intake and a low score. Compared with individuals in the low meat intake and low HbA1c level or genetic risk groups, the HRs for those in the high meat intake and high risk groups were 6.19 (5.01 to 7.65) for HbA1c, 3.39 (2.97 to 3.86) for T2D GRS and 1.49 (1.32 to 1.68) for the insulin resistance GRS, after meta-analysing the results in EPIC-InterAct and UKBB.

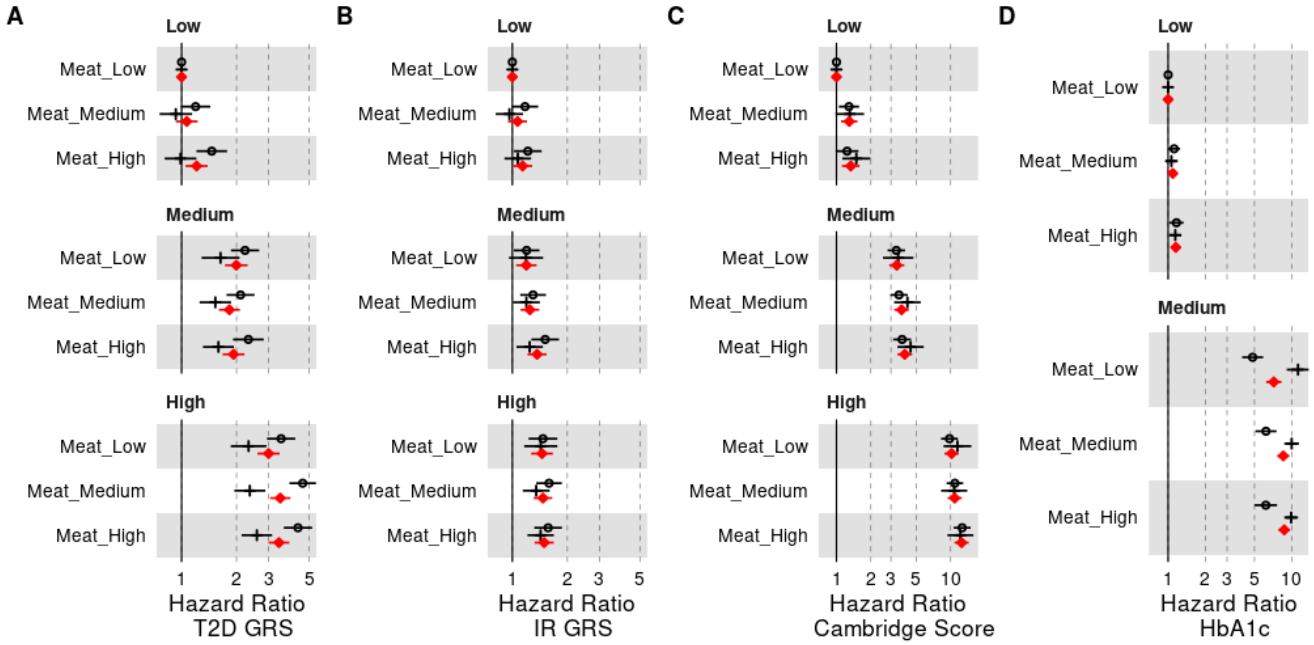


Figure 5.3 Joint effects of meat intake and genetic or clinical risk on the risk of incident T2D in the EPIC-InterAct (n=20,628) and the UK Biobank (n=316,222) studies and in meta-analysis. Adjusted hazard ratios and 95% confidence intervals were reported. In these comparisons, participants with low meat intake and low diabetes risk defined by tertiles of genetic or clinical risk factors were regarded as the reference group. The HbA1c groups were classified according to thresholds for diagnosing diabetes (<42 mmol/mol as the low group, ≥42 and <48 mmol/mol as the medium group, ≥48 mmol/mol were not included for analysis). GRS, genetic risk score; IR, insulin resistance; Cambridge Score, Cambridge diabetes risk score. The black circle shows results in EPIC-InterAct, the black cross shows results in UK Biobank and the red circle shows results in a random effects meta-nalaysis.

5.4 Discussion

In two large population-based prospective studies including 336,850 participants (19,604 incident T2D cases), we reported that high red meat intake was associated with increased risk of developing T2D independently of genetic susceptibility and clinical risks of diabetes. The relative effects of meat intake were comparable between subgroups. No interaction effects in multiplicative models were observed between meat intake and genetic or clinical risks of T2D. The absolute effects of meat consumption were more pronounced in subpopulations who were at higher risk of T2D defined either by genetic or clinical indicators. Furthermore, having a high meat intake combined with high Cambridge risk score, HbA1c or T2D genetic susceptibility was associated with 12-fold, 6-fold or 3-fold increased risk of T2D respectively, compared with having a low meat intake and low genetic or clinical risk.

Comparison with previous studies

To our knowledge, this is the first study of this kind to quantitatively evaluate the association of meat intake with T2D incidence in different subpopulations defined not only by genetic susceptibility, but also by clinical metrics of diabetes risk using a HbA1c measurement or a validated diabetes risk score (the Cambridge diabetes risk score). The association between meat and T2D has been established in the general population in large cohort studies, including the EPIC-InterAct and the UKBB[33]

[,80]. However, whether the positive association varies across different strata has not been established.

The role of gene-diet interaction in disease development has been widely accepted but there is limited evidence about the interplay between genetic composition and dietary behaviour of meat intake on T2D risk or glycaemic traits[230]. In this study, we observed the associations between meat intake and incident T2D were comparable across strata and the interaction effects between genetic predisposition and meat intake were not significant, which is in line with a previous report by Qi et al., which studied a smaller population of 2,533 male health professionals with 1,196 diabetes cases[221]. The general pattern of association across strata was similar in both studies but most associations in the previous report by Qi et al. were not significant except for the group with the highest quintile of T2D GRS, which is likely due to the small sample size in subgroups[221]. The CHARGE Consortium studied gene-meat interactions

on glycaemic traits in 50,000 people from 14 studies. That meta-analysis reported significant associations of meat intake with fasting glucose and insulin that were consistent across different GRS groups. However, the gene-meat interactions on incident T2D were not evaluated[231]. Furthermore, compared with previous studies, the present study derived comprehensive genetic risk scores by including more SNPs (400 vs 10 for T2D, 53 vs 9 for insulin resistance) to improve the power for estimating genetic risks with the advance of latest GWAS studies.

In this study, the general association patterns were similar in EPIC-InterAct and UKBB, but the effect estimates in EPIC-InterAct were generally larger than those in UKBB. This could potentially be because that the EPIC-InterAct study has more accurate estimation of dietary variables and has used multiple data sources to ascertain the diabetes cases, compared with the UKBB, in which the amount of meat intake were estimated using information from follow-up surveys which are available for a subset of participants.

The implications of absolute risks

Although the relative effects of meat intake on T2D incidence were similar across strata, the absolute risks of meat intake were greater in groups with high T2D risk defined either by the Cambridge diabetes risk score, HbA1c, or T2D GRS. From the perspective of personalised dietary intervention, individuals can gain more from preventions if they have high absolute risks because the number needed to treat (the inverse of the absolute risk) for intervention would be lower. For example, one T2D event could be prevented if 32 individuals with high HbA1c (prediabetes) could reduce their dietary meat intake from high to low levels over 10 years, compared with 167 individuals with normal HbA1c levels who would need to modify their meat eating behaviour to achieve the same effect.

The absolute effects of meat intake on T2D incidence were larger in high risk groups defined by clinical indexes (HbA1c or Cambridge risk score) than using GRSs of T2D. The marked effects of clinical factors on the absolute risk of T2D compared to that of genetic scores suggests the clinical indexes of diabetes risk are valid tools to classify populations with higher risk which are more sensitive to the effect of dietary meat intake. However, the use of GRSs could be helpful to identify high-risk populations because genetic information is less affected by environmental factors and potentially useful in early preventions before the change of any clinical indexes.

Strengths and limitations

The major strength of this study is the large population size and number of incident T2D cases, which enables the study of the combination of genetic risks, clinical factors and meat intake in detail. In addition, we constructed comprehensive genetic risk scores to accurately predict genetic risk, using large numbers of variants identified to be associated with T2D and factors on the pathological pathway of diabetes in the latest large GWAS meta-analyses. We also evaluated the risk of diabetes using clinical indexes, including both biochemical measurements of HbA1C and an effective and simple risk score composed of non-biochemical information.

Several limitations need to be noted. First, this study was based on European descent, and thus generalisation of results to other populations is limited. Studies are needed that examine these associations in diverse populations. Second, ascertainment of incident T2D cases in UK Biobank was based on hospital episode statistics, and thus some new T2D cases may be missed if they were identified in General Practice. However, this misclassification is likely to be independent of diet and GRS and thus might not cause serious bias. We found the association patterns were similar in UKBB, and in EPIC-InterAct, which used a comprehensive multi-source approach to ascertain diabetes cases. Third, residual confounding due to measurement errors or unmeasured factors may still exist, even though we have adjusted for various potential confounders.

In summary, a higher meat intake was associated with increased risk of incident T2D independently of T2D genetic susceptibility and clinical markers of T2D risk. This result highlights the benefit of reducing meat intake for T2D burden in the whole population. The Cambridge diabetes score and HbA1c measurements are efficient tools to identify high-risk populations compared with genetic instruments in clinical practice but genetic instruments may have an advantage in early prediction of disease risks. The effects of providing stratified nutritional recommendations in improving diabetes outcomes need to be evaluated in future studies.

Chapter 6

Associations of meat consumption with incident type 2 diabetes in 1.5 million participants from 23 studies: a federated meta-analysis in InterConnect

Abstract

Background The association of meat consumption and its subtypes with new onset type 2 diabetes (T2D) has been investigated mainly in Europe and North America. Whether the association is the same for populations in other geographic regions is not clear.

Methods We examined the association between types of meat consumption and T2D using individual participant data from 23 cohort studies from America (9), Europe (10), the Western Pacific (3), and the Eastern Mediterranean (1), comprising 1,501,177 participants. Hazard ratios (HRs) and 95% confidence intervals (CIs) for associations of red meat, processed meat and poultry intake with incident T2D were estimated for each study with adjustment for a consistent set of confounders and were combined across studies using a random effects meta-analysis.

Results A total of 68,779 incident cases of T2D were included, with a median of 13 years of follow-up. Red meat was the dominant type of meat consumption in most areas. Processed meat consumption was relatively higher in European populations, especially in Germany (median intake of 56.9 g/d in the EPIC-InterAct Germany study). Poultry was highly consumed in North America and Eastern Mediterranean cohorts and was the dominant meat type consumed in Iran (median intake of 48.1 g/d in the Golestan study). Consumption of red meat (HR per 100 g/d 1.11, 95% CI 1.07 to 1.16), processed meat (HR per 50 g/d 1.14, 95% CI 1.09 to 1.18) and poultry (HR per 100 g/d 1.09, 95% CI 1.02 to 1.16) were all significantly associated with risk of T2D. There was no significant effect modification by age, sex, or BMI on the meat-T2D association ($P_{interaction}$>0.05).

Conclusion Higher red meat, processed meat and poultry consumption is associated with increased risk of T2D. Our findings support current dietary guidelines limiting the consumption of red and processed meat products as a behavioural intervention for T2D prevention. The robust data on T2D risk associated with poultry intake has particular implications for public health, calling for urgent attention from policymakers on T2D prevention.

6.1 Introduction

Meat consumption has rapidly increased worldwide and is above what is considered to be the optimal intake level in many regions[25]. Livestock production has substantial adverse impacts on global warming, and increased meat consumption has been linked to enormous burdens of non-communicable diseases[2,4]. Habitual consumption of red and processed meat has been shown to be associated with risk of type 2 diabetes (T2D) according to meta-analyses of prospective studies[31–39]. Compared to red and processed meat, poultry consumption has been considered a potential alternative to red and processed meat consumption[46,47]. However, the association of poultry consumption with T2D risk has only been characterised in a few studies with inconclusive results[39,40,48]. This evidence needs to be updated with more studies and standardised methods.

Moreover, research evidence on the meat-T2D association has been highly heterogeneous, reflecting variations in meat intake habits or T2D risks between populations. The available evidence from the literature-based meta-analyses included studies predominantly from North America and Europe, with a few from Asia and Australia and none from other areas. Data in different geographic regions is required to characterise the association of meat consumption with the risk of incident T2D, especially in understudied populations, and to investigate between-population differences.

The InterConnect Project provides an opportunity to study associations between types of meat consumption and the T2D risk using global data, was enabled by the use of a federated-meta analysis approach to avoid restrictions on transferring data allowing cross-cohort analyses securely and flexibly. Therefore, in this study, I assessed the meat-T2D association in cohort studies worldwide based on the InterConnect platform to investigate whether the association between meat intake and T2D is the same across populations.

6.2 Methods

6.2.1 Study design and participants

InterConnect is an international research project aiming to optimise the use of individual participant data by enabling cross-cohort analyses without pooling data at a central location (see details in **Chapter 1.4.4**). InterConnect takes a federated meta-analysis approach (**Chapter 1.5.3**) to build consortia within this infrastructure to answer specific research questions.

For this study, we used the following sources to identify potentially relevant studies: 1) the InterConnect Data Discovery registry (*http://www.interconnect-diabetes.eu/data-discovery/*), which was compiled using systematic searches of the literature alongside surveys of other online study registries, surveys of websites relating to consortia of studies, and searches of the literature to identify unpublished data; 2) published studies containing information on T2D or meat intake; and 3) studies that have participated previous exemplars in InterConnect.

We attempted to contact 121 cohorts, and 53 studies agreed to participate in this consortium. Other studies declined to join for various reasons, including being unable to establish contact (59), having insufficient data on exposure, outcome, or covariates (4), and having no capacity to contribute (5).

23 cohorts from 12 countries participated in the collaborative group (***Supplementary Table 6.1***). Data for 11 cohorts were obtained by approval of data sharing requests, 9 cohorts uploaded data to a local server to allow federated meta-analysis, and 3 studies shared summarised statistics following the same analysis protocol. We classified regions according to the WHO and included 9 cohorts in America (North and Latin America), 10 in Europe, 3 in the Western Pacific (Australia, China), and 1 from the Eastern Mediterranean (Iran). The distribution of participating studies is shown in ***Figure 6.1***. All cohorts obtained ethical review board approval at the host institution and written informed consent from participants.

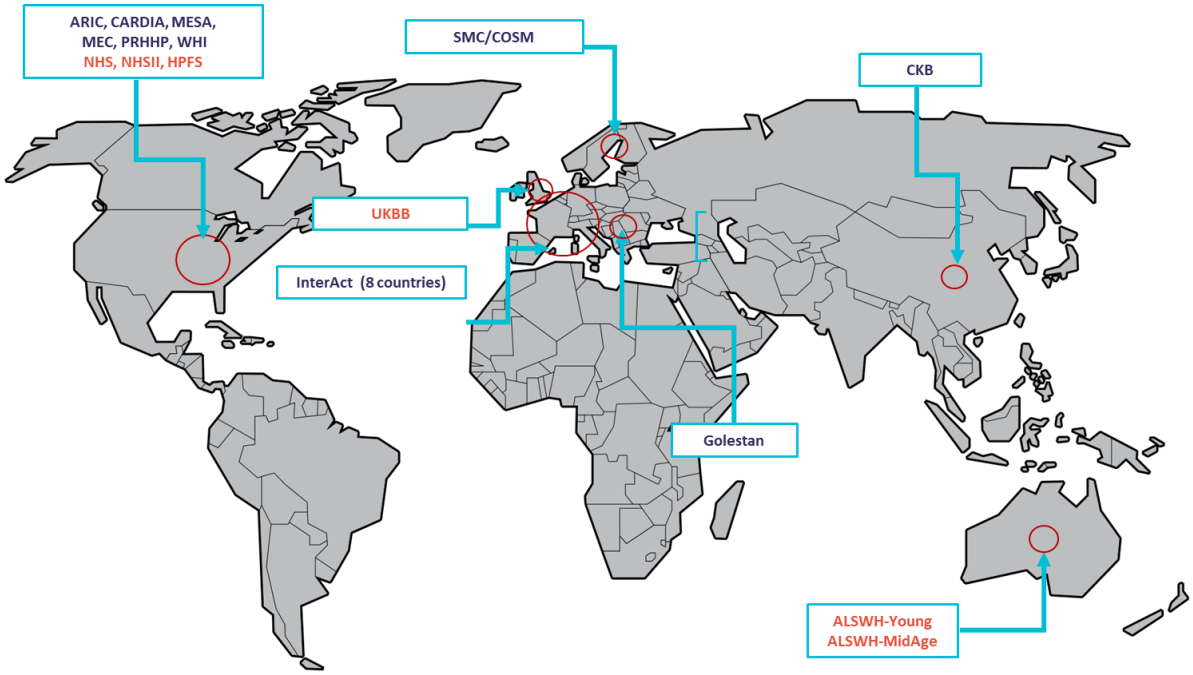


Figure 6.1. Map of the distribution of included studies (n=23). Study names in blue are cohorts that participated in the previous InterConnect projects; and those in red are new studies that joined this project. ALSWH, Australian Longitudinal Study on Women's Health; ARIC, Atherosclerosis Young Adults Study; CARDIA, Coronary Artery Risk Development in Young Adults; CKB, China Kadoorie Biobank; COSM, Cohort of Swedish Men; EPIC, the European Prospective Investigation into Cancer; HPFS, Health Professionals Follow-up Study; MEC, Multiethnic Cohort Study; MESA, Multi-Ethnic Study of Atherosclerosis; NHS, Nurses' Health Study; PRHHP, Puerto Rico Heart Health Program; SMC, Swedish Mammography Cohort; WHI, Women's Health Initiative Study; UKBB, UK Biobank.

6.2.2 Dietary assessment

Dietary information in participating studies was collected by self-reported approaches: 16 studies used FFQs, 4 studies used quantitative dietary questionnaires, 1 studies used interviewer-administered dietary history, 1 study used 24-hour dietary recalls, and 1 study used either FFQs or quantitative dietary questionnaires depending on location (***Supplementary Table 6.2***). The majority of cohorts provided exposure data in the unit of g/d. We used consumption in g/d for the exposure data: including red meat (e.g., beef, pork, lamb, veal); processed meat (e.g., bacon, ham, sausage, hot dog), and poultry (e.g., chicken, turkey, duck, goose). For studies that didn't provide derived exposure variables, we calculated them by summing the consumption of separately reported foods. For studies that provided data in other formats, we transformed data to g/d using variable-specific standard portion sizes sourced from the databases in the United States Department of Agriculture (USDA) (https://www.usda.gov/) (***Supplementary Table 6.3***).

6.2.3 Incident T2D ascertainment

We adopt the definitions of T2D used in previous InterConnect exemplars[232,233].

The primary outcome was clinically incident T2D which was confirmed if a case fulfilled any one or more of the following criteria: (1) ascertained by linkage to a registry or medical record; (2) confirmed antidiabetic medication usage; (3) self-report of physician diagnosis or antidiabetic medication, verified by any of the following: a) ≥1 additional source from (1) or (2) above; b) an abnormal biochemical measurement (e.g., fasting glucose level ≥7.0 mmol/l or glycated haemoglobin (HbA1c) ≥48 mmol/mol)[6,7]; c) a validation study with high concordance.

We defined a more inclusive outcome as incident T2D (the secondary outcome) if a participant was confirmed by any of the following criteria: (1) ascertained by linkage to a registry or medical record; (2) confirmed antidiabetic medication usage; (3) self-report of physician diagnosis or antidiabetic medication; (4) an abnormal biochemical measurement.

6.2.4 Covariates

We specified potential confounding factors, including socio-demographic characteristics (age, sex, education level), lifestyle behaviours (smoking, alcohol drinking, physical activity), dietary information (total energy intake, fruit, vegetables, sugar-sweetened beverages, fish, dairy products, cereal products, legumes, nuts and seeds, other types of meat, fibre, eggs, coffee, and tea), body mass index (BMI), comorbidity at baseline (myocardial infarction, stroke, cancer, hypertension) and family history of diabetes. ***Supplementary Table 6.4*** shows details of variable information for each study.

6.2.5 Statistical analysis

For each study, we excluded participants with a diagnosis of prevalent diabetes at baseline, those with incident type one diabetes during follow-up, and those with possibly invalid dietary questionnaires defined by extreme energy intakes (<500 or >3500 kcal/d for women and <800 or >4200 kcal/d for men), and those with missing values for any of the exposures, outcomes or main confounders (age, sex, education, smoking, physical activity, alcohol intake, BMI, dietary consumption of fruit, vegetables, fish, sugar-sweetend beverages, dairy products, cereal products, legumes, nuts and seeds, fibre, eggs, coffee, tea and total energy intake).

Association analysis

We expressed observed associations per 100 g/d for red meat and poultry and 50 g/d for processed meat. Hazard ratios (HRs) and 95% confidence intervals (CIs) for risk of T2D incidence were estimated using Cox proportional hazards regression[156] in each study. For the eight case-cohort studies in the EPIC-InterAct, Prentice-weighted Cox regression was applied[176]. Multiple models were conducted: model 1 adjusted for age and sex; model 2 further adjusted for education, smoking, physical activity, alcohol intake, squared alcohol intake, total energy intake, BMI, squared BMI, and dietary factors, including information on fruit, vegetables, fish, other types of meat, sugar-sweetened beverages, dairy products, eggs, cereal products, fibre, coffee, and tea. Some potential confounders (e.g., family history of diabetes, waist circumference, comorbidity) were not available for many studies and were only examined in additional analyses.

Effect modification study

We investigated effect modification by age (age<60 years as zero), sex, and BMI (<25 kg/m^2 as zero) by adding an interaction term in model 2 within each cohort. Then I meta-analysed the regression estimates of the interaction terms using random effects estimation. Analyses were further stratified by age, sex, or BMI if there was evidence of significant interaction.

We further presented results by geographic regions according to WHO classification: Americas, including North and Latin America; Europe; the Eastern Mediterranean; and the Western Pacific, including China and Australia. Studies from Africa and South East Asia were not available in the current study, and thus could not be included in this analysis.

Federated meta-analysis

We estimated pooled effects across all cohorts by combining effect estimates of each participating study using a random-effect meta-analysis. Heterogeneity was calculated with τ^2 and I^2 statistics. To investigate the potential sources of heterogeneity, we performed a meta-regression of effect estimates on cohort-level characteristics, including median intakes of meat, geographic areas, and the duration of follow-up.

Hardware and software construction

We performed a federated meta-analysis using DataSHIELD based on a series of R (R Core Team, Vienna, Austria) packages. Details of DataSHIELD are described in **Chapter 1.5.3.** In brief, DataSHIELD is an infrastructure and a series of R packages that allow the conduct of individual-level meta-analysis remotely without needing data to be transferred from the local institutions to a central location. Through DataSHIELD, commands were passed from the analysis server to each data server, and the results were returned to the analysis server. We then combine the results from each cohort to calculate the overall estimates.

In this study, we took a hybrid approach to analyses. We adopted a preferred federated approach for most studies. Each of these studies has set up a data server with data uploaded. A central server connected with each data server was also set up in Cambridge to manage, harmonise, and execute analysis on the data servers. For several studies with difficulties in setting up servers, the analysis was performed by an analyst in each study who had access to the data following the agreed analysis plan and the results were returned to the coordinator in Cambridge for meta-analysis.

6.3 Results

A total of 1,501,177 individuals from 23 collaborative cohorts were included in the analysis. Nine out of the 23 studies had not previously published results on this study topic. A similar proportion of participants were from studies in America (34%), Europe (32%), and the West Pacific (33%), and with fewer than 1% from the Eastern Mediterranean (Iran). Seven studies comprised only women (WHI, EPIC-InterAct France, ALSWH-MidAge, ALSWH-Young, MTC, NHS I, and NHS II), and two studies consisted of only men (PRHHP, HPFS) (***Table 6.1***). Most studies involved older participants with mean ages ranging from 50-60 years, and two studies (CARDIA and ALSWH-Young) included young adults (mean ages of 25 and 28 years).

Table 6.1. Participant characteristics in the cohorts participating in the InterConnect project on the association between meat consumption and type 2 diabetes (n=1,501,177).

Country	Study Name	N	Age, y	N female	Primary outcome		Secondary outcome	
					N event	Follow-up, y	N event	Follow-up, y
America								
USA	MEC	143,811	59.2 (51.0, 67.5)	61,985	7,856	17.0 (15.5, 17.6)	7,856	17.0 (15.5, 17.6)
USA	ARIC	11,895	54.1 (49.3, 58.9)	6,617	804	9.1 (8.8, 23.3)	2,339	9.0 (7.3, 23.1)
USA	WHI	83,491	63.9 (58.0, 69.3)	83,491	7,404	7.9 (6.8, 8.9)	7,721	7.9 (6.8, 8.9)
USA	NHS I	69,698	45.6 (38.8, 52.4)	69,698	5,644	21.3 (18.9, 22.6)	5,644	21.3 (18.9, 22.6)
USA	NHS II	90,746	36.1 (32.8, 40.6)	90,746	7,411	25.4 (21.9, 27.4)	7,411	25.4 (21.9, 27.4)
USA	HPFS	45,302	52.4 (48.9, 65.2)	-	4,385	23.5 (20.3, 25.2)	4,385	23.5 (20.3, 25.2)
USA	MESA	4,923	61.6 (53.2, 70.0)	2,660	228	9.0 (6.6, 10.3)	692	9.1 (6.7, 10.2)
USA	CARDIA	3,923	25.1 (22.3, 28.9)	2,301	198	25.1 (19.0, 25.5)	396	25.0 (16.3, 25.1)
Puerto Rico	PRHHP	6,977	52.7 (47.3, 57.4)	-	213	5.0 (5.0, 5.0)	825	5.0. (5.0, 5.0)
Europe								
France	EPIC-InterAct France	795	55.8 (51.3, 61.8)	795	257	9.3 (7.3, 10.5)	257	9.3 (7.3, 10.5)
Italy	EPIC-InterAct Italy	3,112	51.7 (45.5, 57.5)	2,023	1,271	10.9 (6.8, 12.7)	1,271	10.9 (6.8, 12.7)
Spain	EPIC-InterAct Spain	5,584	50.1 (44.3, 56.8)	3,159	2,354	12.5 (9, 13.6)	2,354	12.5 (9, 13.6)
Netherlands	EPIC-InterAct Netherlands	2,067	55.0 (50.2, 61.2)	1,707	741	11.1 (6.4, 12.6)	741	11.1 (6.4, 12.6)
Germany	EPIC-InterAct Germany	3,448	53.6 (45.9, 59.5)	1,741	1,505	9.5 (4.9, 11.2)	1,505	9.5 (4.9, 11.2)
Sweden	EPIC-InterAct Sweden	5,192	55.4 (49.9, 61.6)	2,732	2,383	11.9 (9.2, 13.6)	2,383	11.9 (9.2, 13.6)
Denmark	EPIC-InterAct-Denmark	3,896	56.7 (52.9, 60.8)	1,730	1,970	10.4 (6.3, 11.6)	1,970	10.4 (6.3, 11.6)
UK	EPIC-InterAct UK	1,858	59.1 (50.3, 66.3)	988	608	10.6 (6.3, 12.2)	608	10.6 (6.3, 12.2)
UK	UK Biobank	456,708	58.0 (38.0, 73.0)	252,174	16,592	7.2 (6.3, 8.1)	16,592	7.2 (6.3, 8.1)
Sweden	SMC/COSM	49,461	58.1 (52.3, 66.9)	22,727	4,831	17.0 (16.1, 18.0)	4,910	17.0 (16.1, 18.0)
Western Pacific								
Australia	ALSWH-MidAge	8,617	52.5 (51.2, 53.8)	8,617	869	15.2 (12.3, 15.3)	869	15.2 (12.3, 15.3)
Australia	ALSWH-Young	6,939	27.6 (26.3, 28.8)	6,939	118	14.6 (12.3, 15)	118	14.6 (12.3, 15)
China	CKB	482,423	51.6 (30.6, 41.2)	84,629	17,043	11.1 (10.3, 12.1)	17,043	11.1 (10.3, 12.1)
Eastern Mediterranean								
Iran	Golestan	10,146	49.9 (45.2, 56.3)	5,241	686	4.2 (3.6, 5.6)	1,191	4.1 (3.4, 5.6)

Following the last page

Data are median (interquartile range, IQR), or n (%). ALSWH, Australian Longitudinal Study on Women's Health; ARIC, Atherosclerosis Young Adults Study; CARDIA, Coronary Artery Risk Development in Young Adults; CKB, China Kadoorie Biobank; COSM, Cohort of Swedish Men; EPIC, European Prospective Investigation into Cancer; HPFS, Health Professionals Follow-up Study; MEC, Multiethnic Cohort Study; MESA, Multi-Ethnic Study of Atherosclerosis; NHS, Nurses' Health Study; PRHHP, Puerto Rico Heart Health Program; SMC, Swedish Mammography Cohort; WHI, Women's Health Initiative Study. The SMC and COSM used the same protocol and were made available as one combined dataset.

Median consumption of red meat ranged from 8.6 g/d in the Golestan study to 106.2 g/d in the CARDIA study. Median processed meat intake ranged from 0 g/d in the Golestan study and the PRHHP study to 56.9 g/d in the EPIC-InterAct Germany study. Poultry intake ranged from 0 g/d in PRHHP to 63.3 g/d in CARDIA. Red meat was reported as the most frequently consumed meat type in most areas. Processed meat consumption was higher in European populations, especially in Germany (EPIC-InterAct Germany, median of 56.9 g/d), compared with other regions. Poultry was the primary meat type consumed in Iran (the Golestan study) and was consumed in similar quantities as red meat in other areas (***Table 6.2***).

Table 6.2. The consumption types of meat intake in 23 cohorts participating the InterConnect project on the association between meat consumption and incident type 2 diabetes (n=1,501,177).

Country	Study Name	Red meat	Processed meat	Poultry
America				
USA	MEC	30.0 (16.0, 49.5)	13.1 (6.3, 23.2)	32.4 (19.2, 54.1)
USA	ARIC	66.4 (39.6, 119.5)	16 (8, 36.5)	31.6 (15.8, 48.6)
USA	WHI	25.4 (11.5, 48.5)	7.2 (2.2, 16.2)	25.2 (12.4, 48.3)
USA	NHS I	63.2 (45.2, 106.3)	8.3 (2.1, 22.0)	42.8 (27.4, 76.0)
USA	NHS II	57.6 (38.7, 98.2)	8.3 (1.7, 19.4)	52.3 (29.3, 86.4)
USA	HPFS	61.2 (41.1, 114.3)	10.5 (3.2, 31.6)	48.5 (29.6, 87.4)
USA	MESA	31.9 (15.8, 56.5)	4.1 (0, 10.5)	34.1 (17.8, 58.9)
USA	CARDIA	106.2 (47.5, 193.2)	16.7 (6.5, 33.4)	63.3 (29.4, 118.7)
Puerto Rico	PRHHP	28.3 (0, 84.9)	0 (0, 28.3)	0 (0, 56.6)
Europe				
France	EPIC-InterAct France	47.1 (22.2, 72.2)	26.8 (14.8, 42.7)	18.7 (4.3, 31.1)
Italy	EPIC-InterAct Italy	45.3 (26.8, 65.4)	18.7 (10.7, 32.2)	22.8 (12.9, 35.9)
Spain	EPIC-InterAct Spain	37.8 (18.3, 63.1)	29.7 (14.3, 53.1)	31.5 (18.2, 51.2)
Netherlands	EPIC-InterAct Netherlands	56.1 (33.6, 81.8)	20.8 (10.9, 37)	9.1 (3.7, 16.4)
Germany	EPIC-InterAct Germany	27.3 (16.5, 43.4)	56.9 (34.7, 84.2)	9.2 (4.7, 17.3)
Sweden	EPIC-InterAct Sweden	25.1 (12.2, 44.3)	37.5 (22.1, 58.4)	7.2 (0, 16.4)
Denmark	EPIC-InterAct Denmark	73.7 (53.3, 99.1)	26.2 (15.2, 42.3)	17.3 (10.2, 26.6)
UK	EPIC-InterAct UK	32.2 (16.1, 65.6)	19.9 (11, 34.2)	16.1 (8.1, 49.5)
UK	UK Biobank	71.2 (33.5, 109.1)	38.5 (0, 80.0)	36.4 (0, 56.5)
Sweden	SMC/COSM	44.4 (26.0, 63.1)	32.2 (19.4, 45.0)	8.0 (7.6, 10.1)
Western Pacific				
Australia	ALSWH-MidAge	54.9 (31.1, 89.8)	12.4 (5.4, 22.8)	22.4 (12.3, 37.2)
Australia	ALSWH-Young	48.3 (24.3, 82.1)	14.9 (6.3, 28.4)	25.6 (12.8, 43.2)
China	CKB	51.7 (23.8, 75.4)	-	11.7 (6.8, 17.6)
Eastern Mediterranean				
Iran	Golestan	8.6 (3.9, 16.6)	0 (0, 2.1)	48.1 (25.6, 76.9)

Data are median (interquartile range, IQR). ALSWH, Australian Longitudinal Study on Women's Health; ARIC, Atherosclerosis Young Adults Study; CARDIA, Coronary Artery Risk Development in Young Adults; CKB, China Kadoorie Biobank; COSM, the Cohort of Swedish Men; EPIC, European Prospective Investigation into Cancer; HPFS, Health Professionals Follow-up Study; MEC, Multiethnic Cohort Study; MESA, Multi-Ethnic Study of Atherosclerosis; NHS, Nurses' Health Study; PRHHP, Puerto Rico Heart Health Program; SMC, Swedish Mammography Cohort; WHI, Women's Health Initiative Study. The SMC and COSM used the same protocol and were made available as one combined dataset.

A total of 68,779 clinically incident T2D cases (the primary outcome) and 78,479 incident T2D cases (the secondary outcome) were included, with a median of 13 years of follow-up. Higher red meat consumption was associated with an increased risk of clinically incident T2D in the model adjusted for age and sex (HR per 100 g/d higher intake 1.31, 95% CI 1.21 to 1.42, model 1). The positive association between red meat intake and incident T2D was attenuated with additional adjustment for behavioural factors, BMI and dietary factors but remained significant (1.11, 1.07 to 1.16, model 2) (***Figure 6.2***). The results showed low heterogeneity between studies for red meat-T2D associations (τ^2=0.0031, I^2=53.0%). Regional analysis showed positive associations in America (1.14, 1.08 to 1.21) and Europe (1.07, 1.04 to 1.10). There was no association in the Western Pacific or the Eastern Mediterranean.

Higher processed meat consumption was associated with an increased risk of incident T2D (HR per 50 g/d higher intake 1.40, 95% CI 1.26 to 1.55) in model 1, and the association was attenuated to 1.14 (1.09 to 1.18) in model 2, with low heterogeneity between studies (τ^2=0.0049, I^2=69%). The association was mainly driven by studies in America (1.19, 1.11 to 1.29) and Europe (1.10, 1.05 to 1.16) (***Figure 6.3***).

We also found a positive association between poultry consumption and incident T2D (HR per 100 g/d higher intake 1.29, 95% CI 1.21 to 1.38 in model 1; 1.09, 1.02 to 1.16 in model 2). The results showed moderate heterogeneity in the poultry-T2D association (τ^2=0.131, I^2=72.0%). Results by region showed a significant positive association in the Eastern Mediterranean Region (1.37, 1.19 to 1.59) and no evidence of association in other areas (***Figure 6.4***).

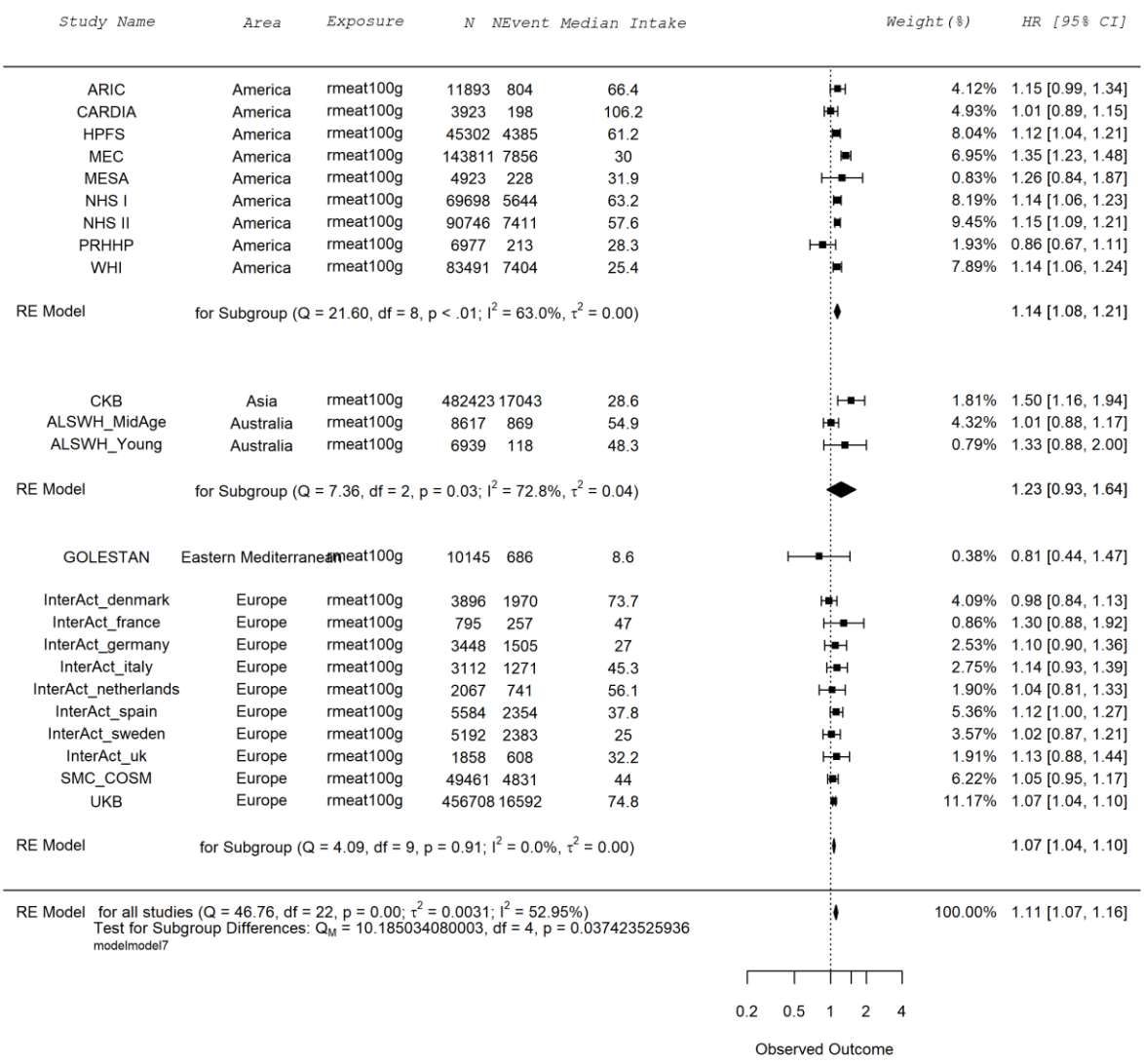


Figure 6.2 Hazard ratios and 95% confidence intervals (CIs) for association between red meat consumption (per 100 g/d) and incident type 2 diabetes (primary outcome) in the InterConnect project. Combined n=1,501,177; total incident type 2 diabetes cases=68,779. Associations are adjusted for age, sex, education, smoking, physical activity, alcohol intake, total energy intake, BMI, and other food intakes.

ALSWH, Australian Longitudinal Study on Women's Health; ARIC, Atherosclerosis Young Adults Study; CARDIA, Coronary Artery Risk Development in Young Adults; CKB, China Kadoorie Biobank; COSM, Cohort of Swedish Men; EPIC, the European Prospective Investigation into Cancer; HPFS, Health Professionals Follow-up Study; MEC, Multiethnic Cohort Study; MESA, Multi-Ethnic Study of Atherosclerosis; NHS, Nurses' Health Study; PRHHP, Puerto Rico Heart Health Program; SMC, Swedish Mammography Cohort; WHI, Women's Health Initiative Study; UKB,UK Biobank. The SMC and COSM used the same protocol and were made available as one combined dataset.

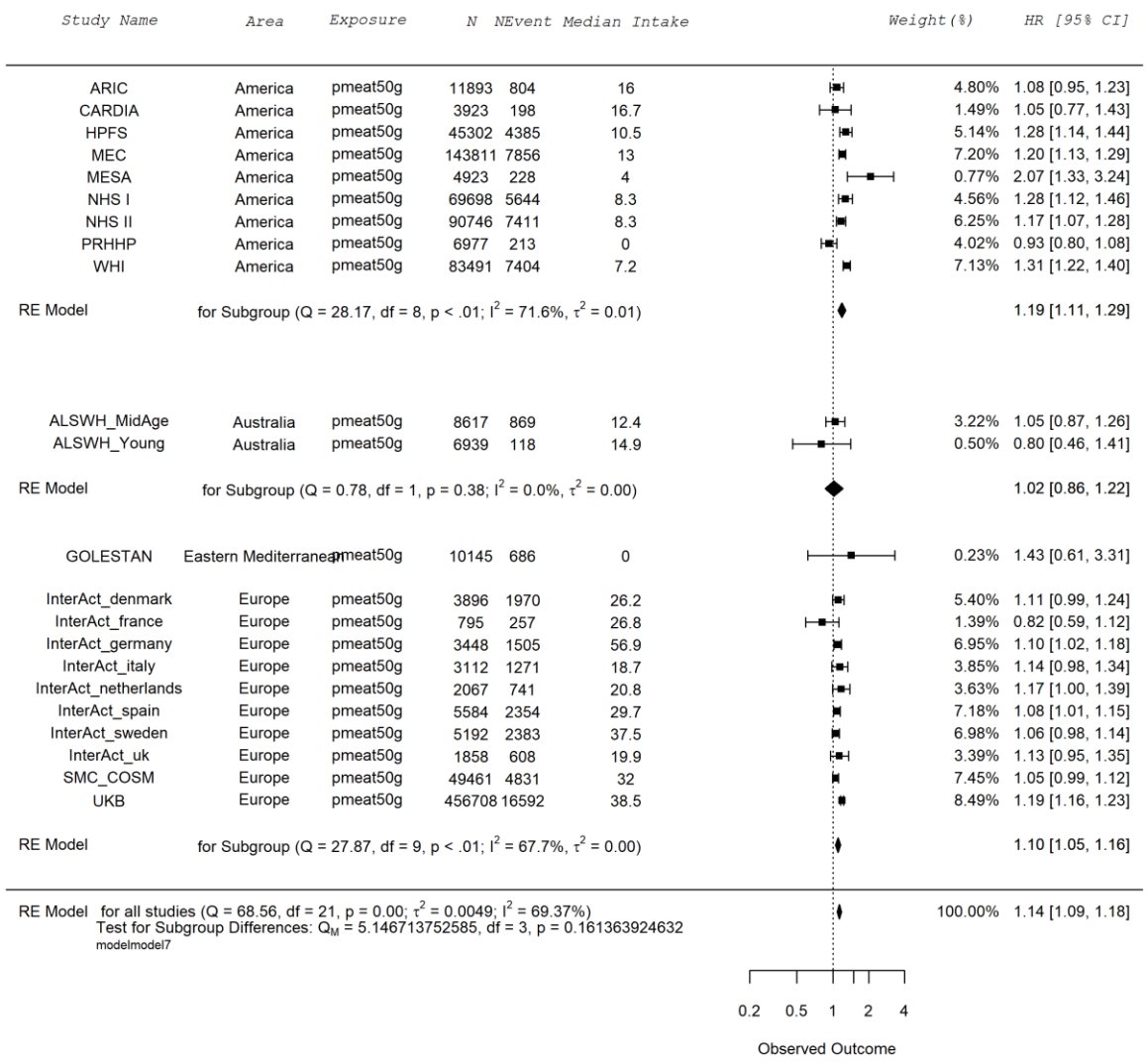


Figure 6.3 Hazard ratios and 95% confidence intervals (CIs) for association between processed meat consumption (per 50 g/d) and incident type 2 diabetes (primary outcome) in the InterConnect project. Combined n=1,018,754; total incident type 2 diabetes cases=51,736. Associations are adjusted for age, sex, education, smoking, physical activity, alcohol intake, total energy intake, BMI, and other food intakes.

ALSWH, Australian Longitudinal Study on Women's Health; ARIC, Atherosclerosis Young Adults Study; CARDIA, Coronary Artery Risk Development in Young Adults; CKB, China Kadoorie Biobank; COSM, Cohort of Swedish Men; EPIC, the European Prospective Investigation into Cancer; HPFS, Health Professionals Follow-up Study; MEC, Multiethnic Cohort Study; MESA, Multi-Ethnic Study of Atherosclerosis; NHS, Nurses' Health Study; PRHHP, Puerto Rico Heart Health Program; SMC, Swedish Mammography Cohort; WHI, Women's Health Initiative Study; UKB,UK Biobank. The SMC and COSM used the same protocol and were made available as one combined dataset.

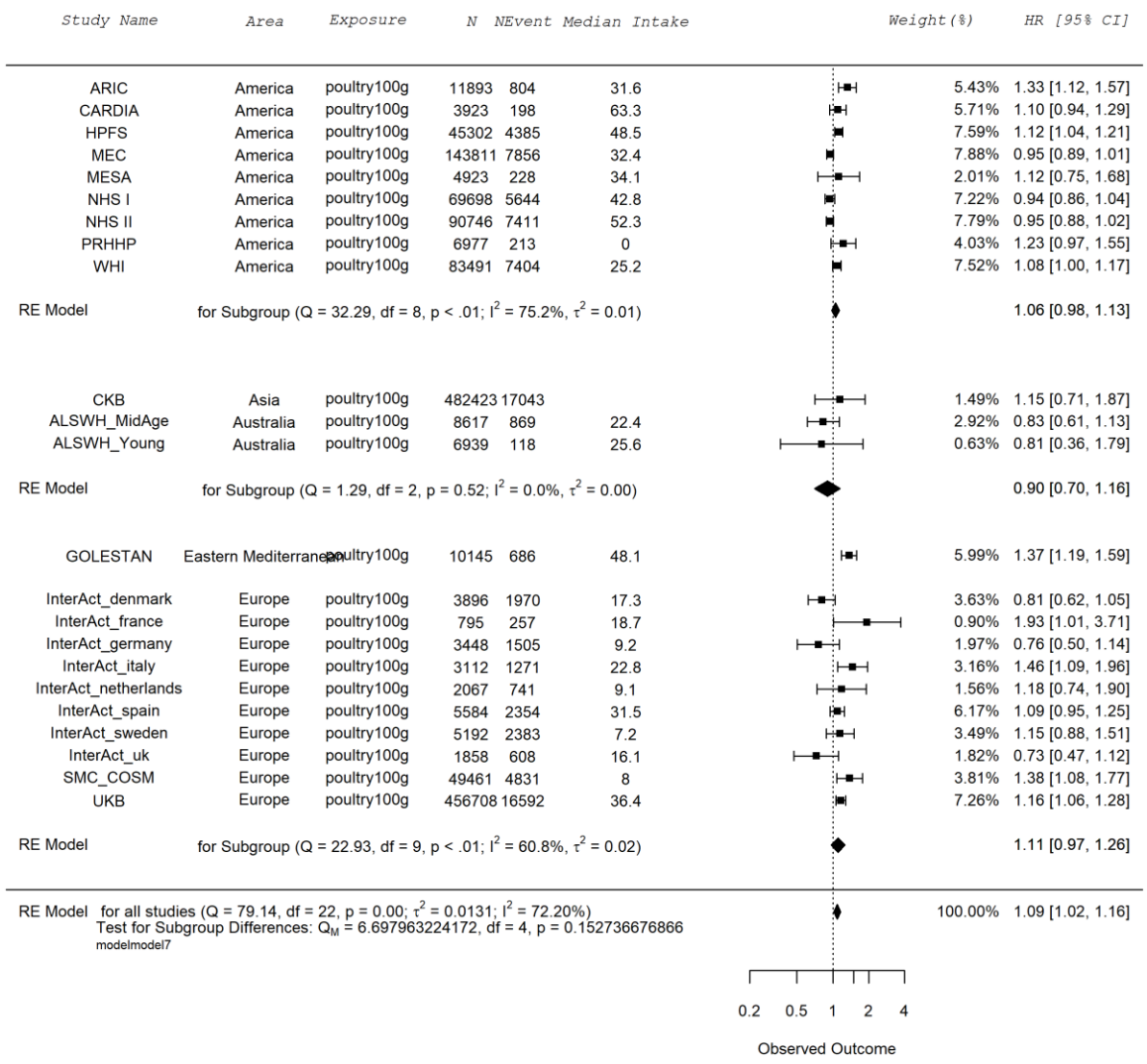


Figure 6.4 Hazard ratios and 95% confidence intervals (CIs) for association between poultry consumption (per 100 g/d) and incident type 2 diabetes (primary outcome) in the InterConnect project. Combined n=1,501,177; total incident type 2 diabetes cases=68,779. Associations are adjusted for age, sex, education, smoking, physical activity, alcohol intake, total energy intake, BMI, and other food intakes.

ALSWH, Australian Longitudinal Study on Women's Health; ARIC, Atherosclerosis Young Adults Study; CARDIA, Coronary Artery Risk Development in Young Adults; CKB, China Kadoorie Biobank; COSM, Cohort of Swedish Men; EPIC, the European Prospective Investigation into Cancer; HPFS, Health Professionals Follow-up Study; MEC, Multiethnic Cohort Study; MESA, Multi-Ethnic Study of Atherosclerosis; NHS, Nurses' Health Study; PRHHP, Puerto Rico Heart Health Program; SMC, Swedish Mammography Cohort; WHI, Women's Health Initiative Study; UKB,UK Biobank. The SMC and COSM used the same protocol and were made available as one combined dataset.

There is no evidence of effect modification by age, sex, or BMI on the meat-T2D association ($P_{interaction}$>0.05). Associations between types of meat consumption and T2D risk were similar when using the secondary definition of T2D outcome (***Supplementary Figures 6.1-6.3***).

In meta-regression to explore potential sources of heterogeneity, the effect sizes of the meat-T2D association between cohorts were not associated with the median meat consumption, follow-up time, or region (P>0.05). An additional analysis with one study removed at a time from the meta-analysis showed that no study substantially affected the heterogeneity nor the overall association.

In the subset of cohorts for which additional covariates were available, individual or pooled effect sizes of meat-T2D associations were not altered with further adjustment for family history of diabetes, comorbidity (stroke, myocardial infarction, hypertension and cancers) or waist circumference (results are not shown).

6.4 Discussion

In this largest federated meta-analysis, including over 1.5 million individuals from 23 cohorts in 12 countries, we reported consistent positive associations of red meat, processed meat and poultry consumption with incident T2D.

Red meat, processed meat consumption and T2D

To our knowledge, this study is the largest to date that revealed positive associations of different types of meat with incident T2D. Our findings of positive associations between red meat, processed meat intake and T2D incidence are consistent with previous evidence from observational studies[35,36,38,40,48,55]. For example, a recent meta-analysis in 2020 by Yang et al. reported a 31% and 46% increased risk of T2D per 100 g/d red meat intake and 50 g/d processed meat intake by including 17 published studies (n=660k). The estimates in this published meta-analysis were more predominent than what we reported, possibly because most of the published studies included by Yang et al. were based on populations from North America and Europe, and the adjustment for confounders in these studies varied. In this federated meta-analysis, we used harmonised data in unified analytical models. Moreover, we not only demonstrated positive associations of red meat and processed meat consumption with T2D risk in the American and European populations with broader coverage of countries and a larger sample size but also provided evidence in previously understudied populations, such as those in the Western Pacific (Australia, China) and Eastern Mediterranean (Iran).

Poultry consumption and T2D risk

We reported a positive association between poultry intake and incident T2D after multivariate adjustment, and the effect size for poultry intake (9% higher risk per 100 g/d intake) was similar to that in the red meat-T2D association (11%) within this study. The association between poultry intake and T2D risk has not been well studied previously, with limited evidence. Most studies have reported a null association between poultry intake and T2D[48,74,152]. The evidence quality of the few published meta-analyses of this topic has been evaluated as low and there is evidence of potential publication bias[40,48]. A recent meta-analysis reported a weak association between poultry and T2D incidence (RR per 100 g/d 1.03, 95% CI 1.00 to 1.07) by involving 9 published studies in a dose-response analysis, although the association in the high vs low intake meta-analysis in the same meta-analysis was not

significant when additional two studies were included (n=411k)[40].

The potential mechanisms linking poultry consumption to T2D risk remain elusive. Poultry is characterised by its high protein content, and some research has indicated a potential association between high intake of animal proteins, including poultry, and increased risk of T2D[234,235]. J Li et al found that substituting animal protein with plant protein was associated with lower T2D risk and the risk reduction was mediated by inflammation factors (e.g., tumor necrosis factor-a receptor 2, interleukin-6), leptin, and endothelial dysfunction biomarkers[236]. Amino acid composition can vary among different sources of animal proteins. The exact amino acids and nutrients contributing to the observed associations have yet to be identified and understood. In comparison to red meat and processed meat, poultry is generally not a source of AGEs. Nevertheless, using high-heat cooking methods, such as frying or grilling, can lead to a substantial increase in the formation of AGEs in poultry[94]. This increase may, in turn, contribute to insulin resistance and the development of T2D[95–97].

Meat-T2D association in the Eastern Mediterranean

We observed a 37% higher risk of T2D per 100 g/d poultry consumption in an Eastern Mediterranean (one cohort in Iran) population, in which the effect estimate for poultry was higher than that in other populations. The association between red meat and T2D incidence in this population was not significant. One feature of this population is that poultry is the primary type of meat for consumption compared with red meat (median intake 48.1 vs 8.6 g/d). The results of meat intake with T2D risk in this study were in line with previous research, which reported on a smaller Iranian population (~6000), indicating a weak poultry-T2D association only in the group with the highest quartile consumption compared with the group with the lowest poultry intake (P_{trend}=0.16)[75]. In that previous report, there was not significant association between red meat and T2D risk, even though the median consumption of red meat in that study was higher than that in the study included in our meta-analysis. The diabetes prevalence in the Middle East and North Africa region is the highest (16.2%) across the world and is estimated to increase by 86% by 2045, according to the latest International Diabetes Federation report[7]. Pursuing additional studies in this area is essential to understand better the association between meat consumption and the development of T2D, given the unique pattern of meat consumption, high prevalence of diabetes, and limited evidence of the meat-T2D association in this area.

Potential mechanisms

The underlying mechanisms between meat intake and T2D incidence are not fully established. Obesity is a major risk factor for diabetes, and it might be a confounder or a mediator of in the association between meat and T2D. We did not observe the effect of modification of obesity on the association between meat intake and T2D risk. Associations for types of meat consumption and incident T2D were visibly attenuated with the adjustment for BMI in this study and previous reports[33,75,80], suggesting that the association with meat may have resulted from adiposity, which was well-captured by the measurement of BMI or waist circumference. Moreover, associations were still significant after adjusting for BMI, or other dietary factors, suggesting meat could affect T2D risk by different pathways independent of adiposity, such as insulin sensitivity and pancreatic beta-cell function[15]. Randomised controlled trials (RCTs) have investigated a mechanistic link between meat intervention and risk factors for T2D, such as HbA1c, fasting glucose, fasting or postprandial insulin, and HOMA-IR, but no significant relationships have been reported[79].

Strengths and limitations

This study is the largest population-based meta-analysis, including individual-level data across populations worldwide, including previously understudied populations. We evaluated associations of different types of meat intake with incident T2D using published and previously unpublished data on this research topic. A federated platform enabled us to analyse harmonised data using standardised methods across studies and reduced heterogeneity relevant to the analysis approach.

This study has several limitations. Despite considerable efforts that have been made for data harmonisation, heterogeneity resulting from the dietary assessment may still exist. Dietary intake was assessed using different self-reported approaches, such as FFQs, dietary records, or mixed methods. These conventional approaches have their position in large-scale population-based studies but are also prone to measurement errors. There was variation in the portion sizes used by each cohort and those assigned at analysis. The accuracy of these portion sizes may vary according to the type of meat in each cohort. There might be residual confounding inherent to observational studies due to unmeasured factors (e.g., healthy dietary habits) or variations in covariates measurements. We have tried to include studies from as many regions as possible but were limited by the lack of studies identified from Africa

and South Asia. These regions have unique meat consumption patterns and suffer the heavy burdens of T2D. These research gaps underline the need for epidemiological studies in these regions.

In conclusion, we observed that higher consumption of different types of meat was associated with an increased risk of T2D. The positive association observed for poultry in the Eastern Mediterranean cohort requires further investigation. Our results suggest that individuals should follow existing dietary guidelines to limit the consumption of red meat and processed meat products to prevent T2D and promote public health. Dietary advice might also need to consider limiting poultry consumption, especially in populations with high consumption levels.

Chapter 7

General Discussion

7.1 Summary of findings

This thesis aimed to understand the roles of different types of meat consumption in the aetiology of T2D. I conducted the following studies during my PhD to fill some research gaps in this topic, in particular, I focused on the effect heterogeneity and potential causal mechanisms underlying the meat-T2D association.

In **Chapter 2**, I refined measures of the exposure (consumption of red meat, processed meat and poultry) and the outcome (incident T2D) in the EPIC-Norfolk study. The dietary information was measured using two self-reported approaches: the food frequency questionnaire (FFQ) and 7-day diet diary (7dDD). I improved T2D case ascertainment in EPIC-Norfolk by linkage of multiple external EHRs data, including diabetic eye screening data and clinical biochemistry data. I then described the association of different types of meat consumption with T2D in EPIC-Norfolk. I observed a stronger red meat-T2D association but a weaker processed meat-T2D association when using 7dDDs compared with using FFQs, suggesting that both dietary assessment approaches might have limitations when measuring subtypes of meat intake.

In **Chapter 3**, I derived metabolite scores to quantify the consumption of red meat, processed meat and poultry based on untargeted metabolite profiling (781 circulating metabolites) and 7dDDs data in 11,432 participants in EPIC-Norfolk. The best performing score was for red meat, comprising 139 metabolites and accounting for 17% of the explained variance of red meat consumption. Eleven top-ranking metabolites that were included in the red meat score were validated in a trial conducted by collaborators in Lyon, France. These metabolites were mainly classified into groups of lipids, amino acids, and xenobiotics, such as plasmalogens, TMAO, and deoxycarnitine. I then showed that this red meat metabolite score was strongly associated with T2D incidence in EPIC-Norfolk.

In **Chapter 4**, I further investigated the potential causal roles of these 11 red meat-related metabolites in T2D incidence by conducting Mendelian randomisation analyses. I observed weak evidence of possible causal associations between meat-related metabolites (e.g., 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4)) and T2D risk, possibly due to limited power and weak genetic instruments.

To understand the heterogeneity of the meat-T2D association, I evaluated whether the

association differed in sub-populations with varying genetic and clinical baseline risk within European population in two large studies (EPIC-InterAct and UK Biobank) (**Chapter 5**). I found that meat intake was associated with incident T2D independently of genetic and clinical predisposition to T2D. This suggests that there are benefits of reducing meat intake on T2D burden in the entire population.

Finally, in **Chapter 6**, I examined associations between types of meat intake and T2D risk based on a federated platform in the InterConnect, which enabled harmonised data analysis of 1.5 million individuals from 23 studies across the world. This individual-participant meta-analysis provided unique evidence of meat-T2D associations in previously understudied populations, such as those in East Asia and East Mediterranean. I included >60k new-onset T2D cases with a median of 13 years of follow-up showing that consumption of red meat, processed meat and poultry were each individually associated with an increased risk of T2D.

Taken together, my work provides strong evidence on the consistency of the association of meat consumption with T2D risk in sub-groups within European populations and also across heterogeneous populations worldwide. In this last chapter, I discuss the strengths and limitations in the context of the overall research, and propose the implications of my findings for future research in public health.

7.2 Strengths

To my knowledge, the work of this thesis is the most comprehensive research to date about the role of meat consumption in the aetiology of T2D.

A strong advantage of this thesis is the innovation in dietary assessment for meat consumption. In this thesis, I not only examined the meat-T2D association using two commonly applied and validated dietary assessment approaches (7dDDs and FFQs) simultaneously, but also developed and validated scores for meat consumption based on large-scale plasma metabolite profiling data. These analyses were enabled by the recent availability of the largest in-depth nutritional and untargeted metabolomics data to date in the EPIC-Norfolk study, and the availably of metabolomics data in a meat intervention study. The use of a machine learning approach (elastic net regression) supported the identification of candidate metabolites and the construction of predictive scores for meat consumption while accounting for between-metabolite correlations. The combination of observational and interventional data improved

the confidence of using identified metabolites to assess meat consumption. The use of different dietary assessment approaches helped to tackle the challenges in measurement errors of dietary exposures and facilitated the evaluation of meat-T2D association from different angles.

The derived predictive metabolites for meat consumption also opened windows for understanding the potential biological pathways in the aetiology of T2D. I have taken the advantage of leveraging the largest GWAS summary statistics to date for expansive metabolites and the T2D outcome to extend beyond the earlier understanding of the causal roles of candidate metabolites in the development of T2D.

A further strength of this thesis is the opportunity to examine the association between meat consumption and T2D risk within InterConnect, which enabled the combination of global data for T2D research using a federated meta-analysis approach. This approach is compatible with a meta-analysis using individual participant data that allows harmonisation of exposure and outcome variables, analysis methods and results interpretation. To my knowledge, this is the largest study on this research question to date with the widest coverage of regions and in particular, it involved data from previously understudied populations, such as those in East Asia and East Mediterranean. It also used harmonised data in the exposure and outcome assessment and analytical models. Taken together, these enabled the investigation of effect heterogeneity for the meat-T2D association using global data with minimised methodological variation.

7.3 Limitations

7.3.1 Measurement of dietary information

In free-living populations, measuring diet is challenging because diet is a complex and time-varying exposure. The work of this thesis (**Chapters 2, 5** and **6**) included large-scale population-based studies which adopted traditional self-reported dietary measurement approaches. One key challenge inherent in these nutritional epidemiological studies is that the self-reported dietary information are prone to measurement error and are unable to measure diet accurately, which might distort the association results. In this thesis, the utilisation of different dietary measurement methods (FFQs and 7dDDs) and the leverage of

nutrition data from diverse populations with large variations could mitigate the potential impacts of measurement error[117]. Moreover, in **Chapter 3**, I assessed the association with T2D risk using a derived metabolite score for red meat intake which showed comparable associations to that using the self-reported approaches. The consistent evidence from different approaches strengthened the associations between meat intake and the risk of T2D. Comparisons between different dietary assessment methods are of future interest, such as utility to understand diet-disease associations and cost-effectiveness in clinical practice.

In some cases, well-conducted dietary intervention studies can provide complementary evidence to self-reported methods in observational studies. However, intervention studies usually have small sample sizes and are not feasible as a method to study the long-term effects of specific foods or nutrients. Therefore, nutritional epidemiological studies based on large cohorts have their roles in providing evidence of diet-disease associations and informing public health policy and guidelines[23]>.

7.3.2 Metabolomics in observational vs. interventional studies

The meat-metabolite analyses (**Chapter 3**) were based on a cross-sectional design. Although I applied advanced mothods to derive a group of candidate metabolites for meat consumption, causal relationships between meat intake and most of these candidates are largely unestablished. The evaluation of the association between the change of meat consumption and the change of metabolites during follow-up surveys could provide further evidence for the meat-metabolite association in future studies.

Dietary intervention studies can provide supporting evidence for the meat-metabolite association by measuring intrapersonal changes in response to meat intake (the intervention diet). In the work of this thesis, I combined data from a cross-over intervention study (n=12), in which 11 candidate metabolites increased after red meat intake compared with tofu. However, intervention studies are usually with small sample sizes, relatively short intervention duration (from a few days, weeks to months), are limited to specific predefined food interventions, all of which limit the power to identify those metabolites with small effects and those with shorter or longer half-lives than the intervention period. Another barrier to the validation of candidate metabolites is the need to confirm the metabolites measured by

untargeted analysis platforms in different studies. Untargeted metabolomics by high-resolution mass spectrometry is an important discovery tool but the standardisation and validation of metabolites remains a considerable challenge due to the lack of cost-effective benchmark materials[23]>. A comprehensive and systematic validation of all candidate metabolites for different types of meat consumption is warranted in the future with available data and advanced technologies.

7.3.3 Causal inference of the metabolite-T2D association

Mendelian randomisation has been developed as a promising tool to investigate possible causal associations in epidemiological studies with the development of GWASs for a variety of traits. The availability of the largest GWASs summary statistics for small molecules enabled the rigorous detection of instruments for meat related metabolites for the work of this thesis. However, due to the cost of genetics and metabolomics assays in large-scale studies, currently available GWASs for metabolites of interest have moderate sample sizes which limits power to derive valid instruments. Moreover, the small number of instruments (n≤3) detected for some metabolites in this study hindered the application of analyses approaches to examine the key assumptions of Mendelian randomisation, such as horizontal pleiotropy. Therefore, the results should be interpreted with caution. The advent of larger GWASs for wide-scale of metabolite profiling data will facilitate the detection of valid instruments for metabolites and the inference of causal associations with T2D and other disease endpoints.

7.3.4 Generalisability of findings

The metabolomics and genetics analyses of this PhD were based on populations of British (**Chapters 3** and **4**) and European ancestry (**Chapters 4** and **5**). Therefore, the validity of the metabolite scores for meat consumption, the findings about the potential causal effects of metabolites on T2D risk and the impacts of meat consumption on T2D incidence within subgroups are limited in the generalisability to other ethnic groups. Further work would be warranted to demonstrate the generalisability of the findings of this thesis to other ancestries.

In **Chapter 6**, I conducted a meta-analysis of associations between types of meat intake and T2D risk. Despite the efforts to include the global data with the widest coverage of populations

(e.g., populations in East Asia and East Mediterranean) in an analysis that is critically not limited by whether studies are published or not, data from populations from Africa, South Asia, and South America are still limited. The generalisability of the findings of this thesis to these underrepresented populations are challenging. The lack of data from these underrepresented populations hinders the understanding of how risk factors affect the aetiology of diseases and the formulation of prevention strategies for these people. Future efforts are warranted to develop prospective studies in these understudied populations to help tackle public health questions in these areas.

7.4 Implications and future perspectives

This section discusses some implications for future research from the findings of this thesis regarding the context of the aetiology and prevention of T2D.

7.4.1 Utilisation of different dietary assessment approaches to facilitate nutritional research

Accurate assessment of the complex and dynamic dietary exposures is the key element to understand the impacts of diet on human health. There is no perfect method to measure all aspects of dietary intakes. FFQs measure usual frequency of intake of a structured food list. Well designed and validated FFQs are the most commonly applied approach for measuring diet in large-scale epidemiological studies. Dietary records (e.g., 7dDDs) can quantify detailed dietary information by recording everything that participants consumed over a period. However, the high burden on participants and high costs of data processing have limited their application in large population-based studies, but they have been widely used in validation studies of other diet assessment methods[23]>.

Metabolites as biomarkers could provide objective measurements of dietary consumption and improve comparability across populations. The work of this thesis (**Chapter 3**) established a feasible template to develop and validate an objective metabolite score for a food group (e.g., red meat) using large-scale untargeted metabolomics data. The score derived from this approach shows comparable validity to conventional dietary assessment approaches in different aspects, including good predictive performance for meat consumption in both observational and interventional studies, and comparable associations with T2D incidence

compared to existing methods (**Chapters 2** and **3**). These findings indicate that this new approach may be complementary to traditional self-report approaches.

Our current work is a small step on the road to the development of complementary approaches for dietary assessment in nutritional studies. In future clinical applications, metabolite profiling may not need to adopt the untargeted metabolomics and may better target specific metabolites predicting meat consumption and also other dietary or modifiable risk factors. Additionally, future work applying this approach to discover and validate metabolite scores for other foods, and the combination of conventional self-reported approaches with newly derived metabolite biomarkers could help advance this process.

7.4.2 EHRs linkage for the ascertainment of T2D

Accurate ascertainment of diabetes cases is vital to ensure valid estimation of disease risk in epidemiological studies. One element of this thesis has been the extension of ascertainment of diabetes cases in the EPIC-Norfolk study. The EPIC-Norfolk study is a large cohort of a British population with wide-ranging variables collected, including genetics, proteomics, metabolomics and expansive phenotypes data. The previous case ascertainment in EPIC-Norfolk was conducted in 2005-2006, and that work contributed to a designed case-cohort for T2D (n=1,503, with ~800 T2D cases) within EPIC-Norfolk. The recent availability of linkage to multiple external EHRs data enabled the improvement of T2D cases ascertainment to the whole cohort during a longer follow-up. More than 2000 new-onset diabetes cases were ascertained, which will support research community to perform future studies in the aetiology of T2D with enlarged power (**Chapter 2**). Most of these new identified T2D cases were contributed by the linkage with the diabetes eye screening data and clinical biochemistry measurements, which are two novel sources not commonly used in large cohort studies. Although these two data sources are regionally specific for the EPIC-Norfolk study, the eye screening programme has been a national screening programme in the UK since early 2000s[23]>. Therefore, these records could be a potential source to help the surveillance of diabetes in large population-based studies in the UK and other regions where data are available. For future research, the use of other national-wide registry data, such as the UK Diabetes Audit Data[240], could be included to improve the work of diabetes ascertainment and

support the evaluation of validity of using different data sources and approaches. We did intend to include such data in this thesis but the slowness of NHS Digital has precluded this approach.

7.4.3 Potential mechanisms of the meat-metabolite-T2D association

The Mendelian randomisation analyses between meat related metabolites and T2D risk in **Chapter 4** suggest that genetically predicted higher concentrations of metabolites in the plasmalogen pathway (e.g., 1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4), 1-palmityl-2-arachidonoyl-GPC (O-16:0/20:4)) were associated with increased risk of T2D. These observations indicate that these two metabolites might play roles in the association between meat consumption and T2D risk. GWASs for these candidate metabolites in larger and diverse populations are required to enable the detection of more genetic instruments and the determination of their possible causal roles in diseases. Moreover, future work is warranted to investigate how these metabolites may perform as biomarkers in the diagnosis, prevention and management of T2D.

7.4.4 Dietary recommendations for the general population

In this thesis work (**Chapter 5**), I did not find evidence to support the hypothesis that meat consumption has different effects in subgroups of European populations who have different levels of genetic predisposition or clinical risk to T2D. I did not observe any multiplicative interaction between meat consumption and genetic or clinical risks of T2D. Based on these findings, personalised or stratified recommendations on dietary change based on genetic or clinical risk for T2D prevention should not be prioritised over public health approaches that address meat consumption at the population level.

In the meta-analysis in **Chapter 6**, consistent evidence of the meat-T2D association from heterogeneous populations was observed, suggesting that all individuals globally should follow the existing clinical or public health advice regarding the benefits of lower meat consumption on T2D risk. Health professionals, including clinicians, dietitians and public health officers should also consider limiting poultry consumption, especially in regions with

high intake levels.

Conclusions

In summary, the work during this PhD provides a comprehensive investigation of the association between meat consumption and the risk of T2D. The use of diverse types of data and cutting-edge data-driven approaches has contributed to the advancement of assessment of dietary meat intake, the T2D outcome and their associations. The combination of genetics and the metabolic profile of meat consumption also presents an opportunity to investigate the causal nature of the association and the mechanisms that may underlie it. Moreover, my PhD work demonstrates the association between meat consumption and T2D risk in sub-groups within European populations and also across heterogeneous populations worldwide. The combination of consistent evidence from a collection of epidemiological analyses in this thesis supports the current public health policy and nutrition advice to limit meat consumption in all individuals in all populations. Future research building on the work of this thesis may provide an enhanced understanding of the aetiology of T2D and contribute towards the future improvement in T2D prevention and management.

References

1 Westhoek H, Rood T, van den Berg M, *et al.* The protein puzzle: the consumption of meat, dairy and fish in the European Union. *PBL (Netherlands Environ Assess Agency)* 2011. http://www.pbl.nl/sites/default/files/cms/publicaties/Protein_Puzzle_web_1.pdf.

2 Godfray HCJ, Aveyard P, Garnett T, *et al.* Meat consumption, health, and the environment. *Science* 2018; **361**: 6399:eaam5324.

3 Xu X, Sharma P, Shu S, *et al.* Global greenhouse gas emissions from animal-based foods are twice those of plant-based foods. *Nat Food* 2021; **2**: 724–32.

4 Mekonnen MM, Hoekstra AY. A global assessment of the water footprint of farm animal products. *Ecosystems* 2012; **15**: 401–15.

5 Willett W, Rockström J, Loken B, *et al.* Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. *Lancet* 2019; **393**: 447–92.

6 World Health Organization. Global report on diabetes. Available from https://apps.who.int/iris/bitstream/handle/10665/204871/9789241565257_eng.pdf?sequence=1 (2016).

7 International Diabetes Federation Committee. IDF Diabetes Atlas 10th edition 2021.

8 DeFronzo RA, Ferrannini E, Groop L, *et al.* Type 2 diabetes mellitus. *Nat Rev Dis Prim* 2015; **1**: 1–22.

9 Weyer C, Bogardus C, Mott DM, Pratley RE. The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *J Clin Invest* 1999; **104**: 787–94.

10 Karakelides H, Irving BA, Short KR, O'Brien P, Sreekumaran Nair K. Age, obesity, and sex effects on insulin sensitivity and skeletal muscle mitochondrial function. *Diabetes* 2010; **59**: 89–97.

11 Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature* 2006; **444**: 840–6.

12 Ebbeling CB, Knapp A, Johnson A, *et al.* Effects of a low-carbohydrate diet on insulin-

resistant dyslipoproteinemia - a randomized controlled feeding trial. *Am J Clin Nutr* 2022; **115**: 154–62.

13 Gołąbek KD, Regulska-Ilow B. Dietary support in insulin resistance: an overview of current scientific reports. *Adv Clin Exp Med* 2019; **28**: 1577–85.

14 Ekelund U, Brage S, Griffin SJ, Wareham NJ. Objectively measured moderate- and vigorous-intensity physical activity but not sedentary time predicts insulin resistance in high-risk individuals. *Diabetes Care* 2009; **32**: 1081–6.

15 Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol* 2018; **14**: 88–98.

16 Czech A. Long-term complications of diabetes mellitus. *N Engl J Med* 1993; **49**: 16–22.

17 McCarthy MI. Genomics, type 2 diabetes, and obesity. *N Engl J Med* 2010; **363**: 2339–50.

18 Almgren P, Lehtovirta M, Isomaa B, *et al.* Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia* 2011; **54**: 2811–9.

19 Willemsen G, Ward KJ, Bell CG, *et al.* The concordance and heritability of type 2 diabetes in 34,166 twin pairs from international twin registers: the Discordant Twin (DISCOTWIN) Consortium. *Twin Res Hum Genet* 2015; **18**: 762–71.

20 Pan XR, Li GW, Hu YH, *et al.* Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance: the Da Qing IGT and diabetes study. *Diabetes Care* 1997; **20**: 537–44.

21 Labetes D, Revention P, Rogram P, Esearch R, Roup G. Reduction of the incidence of type 2 diabetes with lifestyle intervention or metformin. *Int Urol Nephrol* 2002; **34**: 162–3.

22 Ramachandran A, Snehalatha C, Mary S, Mukesh B, Bhaskar AD, Vijay V. The Indian Diabetes Prevention Programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). *Diabetologia* 2006; **49**: 289–97.

23 Lindström J, Ilanne-Parikka P, Peltonen M, *et al.* Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. *Lancet* 2006; **368**: 1673–9.

24 Saito T, Watanabe M, Nishida J, *et al.* Lifestyle modification and prevention of type 2 diabetes in overweight Japanese with impaired fasting glucose levels: a randomized controlled trial. *Arch Intern Med* 2011; **171**: 1352–60.

25 Afshin A, Sur PJ, Fay KA, *et al.* Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2019; **393**: 1958–72.

26 Organisation for Economic Co-operation and Development (OECD). FAO Agric. Outlook 2022. https://data.oecd.org/agroutput/meat-consumption.htm.

27 Ekmekcioglu C, Wallner P, Kundi M, Weisz U, Haas W, Hutter HP. Red meat, diseases, and healthy alternatives: a critical review. *Crit Rev Food Sci Nutr* 2018; **58**: 247–61.

28 Wolk A. Potential health hazards of eating red meat. *J Intern Med* 2017; **281**: 106–22.

29 Hawkes N. Cutting Europe's meat and dairy consumption would benefit health and environment, says report. *BMJ* 2014; **348**: g2949.

30 Bouvard V, Loomis D, Guyton KZ, *et al.* Carcinogenicity of consumption of red and processed meat. *Lancet Oncol* 2015; **16**: 1599–600.

31 Schwingshackl L, Hoffmann G, Lampousi AM, *et al.* Food groups and risk of type 2 diabetes mellitus: a systematic review and meta-analysis of prospective studies. *Eur J Epidemiol* 2017; **32**: 363–75.

32 Tian S, Xu Q, Jiang R, Han T, Sun C, Na L. Dietary protein consumption and the risk of type 2 diabetes: a systematic review and meta-analysis of cohort studies. *Nutrients* 2017; **9**: 1–17.

33 Bendinelli B, Palli D, Masala G, *et al.* Association between dietary meat consumption and incident type 2 diabetes: the EPIC-InterAct Study. *Diabetologia* 2013; **56**: 47–59.

34 Pan A, Sun Q, Bernstein AM, *et al.* Red meat consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis. *Am J Clin Nutr* 2011; **94**: 1088–96.

35 Micha R, Michas G, Mozaffarian D. Unprocessed red and processed meats and risk of coronary artery disease and type 2 diabetes - an updated review of the evidence. *Curr Atheroscler Rep* 2012; **14**: 515–24.

36 Feskens EJM, Sluik D, Van Woudenbergh GJ. Meat consumption, diabetes, and its complications. *Curr Diab Rep* 2013; **13**: 298–306.

37 Micha R, Wallace SK, Mozaffarian D. Red and processed meat consumption and risk of incident coronary heart disease, stroke, and diabetes mellitus: a systematic review and meta-analysis. *Circulation* 2010; **121**: 2271–83.

38 Aune D, Ursin G, Veierød MB. Meat consumption and the risk of type 2 diabetes: a systematic review and meta-analysis of cohort studies. *Diabetologia* 2009; **52**: 2277–87.

39 Zhang R, Fu J, Moore JB, Stoner L, Li R. Processed and unprocessed red meat consumption and risk for type 2 diabetes mellitus: an updated meta-analysis of cohort studies. *Int J Environ Res Public Health* 2021; **18**: 10788.

40 Yang X, Li Y, Wang C, *et al.* Meat and fish intake and type 2 diabetes: dose-response meta-analysis of prospective cohort studies. *Diabetes Metab* 2020; **46**: 345–52.

41 Montonen J, Järvinen R, Heliövaara M, Reunanen A, Aromaa A, Knekt P. Food consumption and the incidence of type II diabetes mellitus. *Eur J Clin Nutr* 2005; **59**: 441–8.

42 Hodge AM, English DR, O'Dea K, Giles GG. Dietary patterns and diabetes incidence in the Melbourne Collaborative Cohort Study. *Am J Epidemiol* 2007; **165**: 603–10.

43 Männistö S, Kontto J, Kataja-Tuomola M, Albanes D, Virtamo J. High processed meat consumption is a risk factor of type 2 diabetes in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention study. *Br J Nutr* 2010; **103**: 1817–22.

44 Steinbrecher A, Erber E, Grandinetti A, Kolonel LN, Maskarinec G. Meat consumption and risk of type 2 diabetes: the Multiethnic Cohort. *Public Health Nutr* 2011; **14**: 568–74.

45 Van Woudenbergh GJ, Kuijsten A, Tigcheler B, *et al.* Meat consumption and its association with C-reactive protein and incident type 2 diabetes: the Rotterdam Study. *Diabetes Care* 2012; **35**: 1499–505.

46 Würtz AML, Jakobsen MU, Bertoia ML, *et al.* Replacing the consumption of red meat with other major dietary protein sources and risk of type 2 diabetes mellitus: a

prospective cohort study. *Am J Clin Nutr* 2021; **113**: 612–21.

47 Ibsen DB, Steur M, Imamura F, *et al.* Replacement of red and processed meat with other food sources of protein and the risk of type 2 diabetes in European populations: the EPIC-InterAct Study. *Diabetes Care* 2020; **43**: 2660–7.

48 Neuenschwander M, Ballon A, Weber KS, *et al.* Role of diet in type 2 diabetes incidence: umbrella review of meta-analyses of prospective observational studies. *BMJ* 2019; **366**: l2368.

49 Cocking C, Walton J, Kehoe L, Cashman KD, Flynn A. The role of meat in the European diet: current state of knowledge on dietary recommendations, intakes and contribution to energy and nutrient intakes and status. *Nutr Res Rev* 2020; **33**: 181–9.

50 Scientific Advisory Committee on Nutrition. Iron and Health Report. 2011. https://www.gov.uk/government/publications/sacn-iron-and-health-report (accessed Aug 26, 2022).

51 World Cancer Research Fund/American Institute for Cancer Research. Diet, nutrition, physical activity and cancer: a global perspective. 2018. Available at dietandcancerreport.org (accessed Aug 26, 2022).

52 Johnston BC, Zeraatkar D, Han MA, *et al.* Unprocessed red meat and processed meat consumption: dietary guideline recommendations from the nutritional recommendations (NUTRIRECS) consortium. *Ann Intern Med* 2019; **171**: 756–64.

53 Erratum: Nutritional recommendations (NutriRECS) on consumption of red and processed meat (Annals of Internal Medicine). *Ann Intern Med* 2020; **172**: 228.

54 Red meat study may have conflicts of interest - The Washington Post. https://www.washingtonpost.com/business/2019/10/14/research-group-that-discounted-risks-red-meat-has-ties-program-partly-backed-by-beef-industry/ (accessed Oct 23, 2022).

55 Pan A, Sun Q, Bernstein AM, *et al.* Red meat consumption and risk of type 2 diabetes: 3 cohorts of US adults and an updated meta-analysis. *Am J Clin Nutr* 2011; **94**: 1088–96.

56 Meyer KA, Kushi LH, Jacobs DR, Folsom AR. Dietary fat and incidence of type 2 diabetes

in older Iowa women. *Diabetes Care* 2001; **24**: 1528–35.

57 Van Dam RM, Willett WC, Rimm EB, Stampfer MJ, Hu FB. Dietary fat and meat intake in relation to risk of type 2 diabetes in men. *Diabetes Care* 2002; **25**: 417–24.

58 Schulze MB, Manson JE, Willett WC, Hu FB. Processed meat intake and incidence of type 2 diabetes in younger and middle-aged women. *Diabetologia* 2003; **46**: 1465–73.

59 Lee DH, Folsom AR, Jacobs DR. Dietary iron intake and type 2 diabetes incidence in postmenopausal women: the Iowa Women's Health Study. *Diabetologia* 2004; **47**: 185–94.

60 Fung TT, Schulze M, Manson JAE, Willett WC, Hu FB. Dietary patterns, meat intake, and the risk of type 2 diabetes in women. *Arch Intern Med* 2004; **164**: 2235–40.

61 Song Y, Manson JE, Buring JE, Liu S. A prospective study of red meat consumption and type 2 diabetes in middle-aged and elderly women: the Women's Health Study. *Diabetes Care* 2004; **27**: 2108–15.

62 Vang A, Singh PN, Lee JW, Haddad EH, Brinegar CH. Meats, processed meats, obesity, weight gain and occurrence of diabetes among adults: findings from Adventist Health Studies. *Ann Nutr Metab* 2008; **52**: 96–104.

63 Fretts AM, Howard B V., McKnight B, *et al.* Associations of processed meat and unprocessed red meat intake with incident diabetes: the Strong Heart Family Study. *Am J Clin Nutr* 2012; **95**: 752–8.

64 Isanejad M, Lacroix AZ, Thomson CA, *et al.* Branched-chain amino acid, meat intake and risk of type 2 diabetes in the Women's Health Initiative. *Br J Nutr* 2017; **117**: 1523–30.

65 Schulze MB, Häring HU, Pfeiffer AFH, Joost HG. An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. *Diabetes Care* 2007; **30**: 510–5.

66 Simmons RK, Harding AH, Wareham NJ, Griffin SJ. Do simple questions about diet and physical activity help to identify those at risk of type 2 diabetes? *Diabet Med* 2007; **24**: 830–5.

67 Lajous M, Tondeur L, Fagherazzi G, De Lauzon-Guillain B, Boutron-Ruaualt MC, Clavel-Chapelon F. Processed and unprocessed red meat consumption and incident type 2

diabetes among French women. *Diabetes Care* 2012; **35**: 128–30.

68 Ericson U, Hellstrand S, Brunkwall L, *et al.* Food sources of fat may clarify the inconsistent role of dietary fat intake for incidence of type 2 diabetes. *Am J Clin Nutr* 2015; **101**: 1065–80.

69 Mari-Sanchis A, Gea A, Basterra-Gortari FJ, Martinez-Gonzalez MA, Beunza JJ, Bes-Rastrollo MBR. Meat consumption and risk of developing type 2 diabetes in the SUN project: a highly educated middle-class population. *PLoS One* 2016; **11**: e0157990.

70 Virtanen HEK, Koskinen TT, Voutilainen S, *et al.* Intake of different dietary proteins and risk of type 2 diabetes in men: the Kuopio Ischaemic Heart Disease Risk Factor Study. *Br J Nutr* 2017; **117**: 882–93.

71 Hirayama T. Life-style and cancer: from epidemiological evidence to public behavior change to mortality reduction of target cancers. *J Natl Cancer Inst Monogr* 1992; : 65–74.

72 Villegas R, Xiao OS, Gao YT, *et al.* The association of meat intake and the risk of type 2 diabetes may be modified by body weight. *Int J Med Sci* 2006; **3**: 152–9.

73 Kurotani K, Nanri A, Goto A, *et al.* Red meat consumption is associated with the risk of type 2 diabetes in men but not in women: a Japan Public Health Center-based Prospective Study. *Br J Nutr* 2013; **110**: 1910–8.

74 Talaei M, Wang YL, Yuan JM, Pan A, Koh WP. Meat, dietary heme iron, and risk of type 2 diabetes mellitus. *Am J Epidemiol* 2017; **186**: 824–33.

75 Hassanzadeh-Rostami Z, Hemmatdar Z, Pishdad GR, Faghih S. Moderate consumption of red meat, compared to soy or non-soy legume, has no adverse effect on cardio-metabolic gactors in patients with type 2 diabetes. *Exp Clin Endocrinol Diabetes* 2021; **129**: 429–37.

76 Turner KM, Keogh JB, Clifton PM. Red meat, dairy, and insulin sensitivity: a randomized crossover intervention study. *Am J Clin Nutr* 2015; **101**: 1173–9.

77 O'Connor LE, Paddon-Jones D, Wright AJ, Campbell WW. A Mediterranean-style eating pattern with lean, unprocessed red meat has cardiometabolic benefits for adults who are overweight or obese in a randomized, crossover, controlled feeding trial. *Am J Clin*

Nutr 2018; **108**: 33–40.

78 Maki KC, Wilcox ML, Dicklin MR, *et al.* Substituting lean beef for carbohydrate in a healthy dietary pattern does not adversely affect the cardiometabolic risk factor profile in men and women at risk for type 2 diabetes. *J Nutr* 2020; **150**: 1824–33.

79 Sanders LM, Wilcox ML, Maki KC. Red meat consumption and risk factors for type 2 diabetes: a systematic review and meta-analysis of randomized controlled trials. *Eur J Clin Nutr* 2023; **77**: 156–65.

80 Papier K, Fensom GK, Knuppel A, *et al.* Meat consumption and risk of 25 common conditions: outcome-wide analyses in 475,000 men and women in the UK Biobank study. *BMC Med* 2021; **19**: 53.

81 Kim Y, Keogh J, Clifton P. A review of potential metabolic etiologies of the observed association between red meat consumption and development of type 2 diabetes mellitus. *Metabolism* 2015; **64**: 768–79.

82 Heianza Y, Ma W, Manson JAE, Rexrode KM, Qi L. Gut microbiota metabolites and risk of major adverse cardiovascular disease events and death: a systematic review and meta-analysis of prospective studies. *J Am Heart Assoc* 2017; **6**: e004947.

83 Abbasi J. TMAO and heart disease: the new red meat risk? *JAMA* 2019; **321**: 2149–51.

84 Farhangi MA, Vajdi M. Gut microbiota–associated trimethylamine N-oxide and increased cardiometabolic risk in adults: a systematic review and dose-response meta-analysis. *Nutr Rev* 2021; **79**: 1022–42.

85 Zhuang R, Ge X, Han L, *et al.* Gut microbe–generated metabolite trimethylamine N-oxide and the risk of diabetes: a systematic review and dose-response meta-analysis. *Obes Rev* 2019; **20**: 883–94.

86 Liu S, van der Schouw YT, Soedamah-Muthu SS, Spijkerman AMW, Sluijs I. Intake of dietary saturated fatty acids and risk of type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition-Netherlands cohort: associations by types, sources of fatty acids and substitution by macronutrients. *Eur J Nutr* 2019; **58**: 1125–36.

87 De Souza RJ, Mente A, Maroleanu A, *et al.* Intake of saturated and trans unsaturated

fatty acids and risk of all cause mortality, cardiovascular disease, and type 2 diabetes: systematic review and meta-analysis of observational studies. *BMJ* 2015; **351**: h3978.

88 Vessby B, Uusitupa M, Hermansen K, *et al.* Substituting dietary saturated for monounsaturated fat impairs insulin sensitivity in healthy men and women: the KANWU Study. *Diabetologia* 2001; **44**: 312–9.

89 Chiu S, Williams PT, Dawson T, *et al.* Diets high in protein or saturated fat do not affect insulin sensitivity or plasma concentrations of lipids and lipoproteins in overweight and obese adults. *J Nutr* 2014; **144**: 1753–9.

90 Imamura F, Micha R, Wu JHY, *et al.* Effects of saturated fat, polyunsaturated fat, monounsaturated fat, and carbohydrate on glucose-insulin homeostasis: a systematic review and meta-analysis of randomised controlled feeding trials. *PLoS Med* 2016; **13**: e1002087.

91 Shahinfar H, Jayedi A, Shab-Bidar S. Dietary iron intake and the risk of type 2 diabetes: a systematic review and dose–response meta-analysis of prospective cohort studies. *Eur J Nutr* 2022; **61**: 2279–96.

92 Talaei M, Wang YL, Yuan JM, Pan A, Koh WP. Meat, dietary heme iron, and risk of type 2 diabetes mellitus. *Am J Epidemiol* 2017; **186**: 824–33.

93 Wittenbecher C, Mühlenbruch K, Kröger J, *et al.* Amino acids, lipid metabolites, and ferritin as potential mediators linking red meat consumption to type 2 diabetes. *Am J Clin Nutr* 2015; **101**: 1241–50.

94 Uribarri J, Woodruff S, Goodman S, *et al.* Advanced glycation end products in foods and a practical guide to their reduction in the diet. *J Am Diet Assoc* 2010; **110**: 911.

95 Sohouli MH, Fatahi S, Sharifi-Zahabi E, *et al.* The impact of low advanced glycation end products diet on metabolic risk factors: a systematic review and meta-analysis of randomized controlled trials. *Adv Nutr* 2021; **12**: 766–76.

96 Kellow NJ, Savige GS. Dietary advanced glycation end-product restriction for the attenuation of insulin resistance, oxidative stress and endothelial dysfunction: a systematic review. *Eur J Clin Nutr* 2013; **67**: 239–48.

97 Liu G, Zong G, Wu K, *et al.* Meat cooking methods and risk of type 2 diabetes: results

from three prospective cohort studies. *Diabetes Care* 2018; **41**: 1049–60.

98 Day N, Oakes S, Luben R, *et al.* EPIC-Norfolk: study design and characteristics of the cohort. *Br J Cancer* 1999; **80**: 95–103.

99 Mulligan AA, Luben RN, Bhaniani A, *et al.* A new tool for converting food frequency questionnaire data into nutrient and food group values: FETA research methods and availability. *BMJ Open* 2014; **4**: e004503.

100 Lentjes MAH, McTaggart A, Mulligan AA, *et al.* Dietary intake measurement using 7-d diet diaries in British men and women in the European prospective investigation into cancer-norfolk study: a focus on methodological issues. *Br J Nutr* 2014; **111**: 516–26.

101 Welch A, McTaggart A, Mulligan A, *et al.* DINER (Data Into Nutrients for Epidemiological Research) – a new data-entry program for nutritional analysis in the EPIC–Norfolk cohort and the 7-day diary method. *Public Health Nutr* 2001; **4**: 1253–65.

102 Welch AA, Luben R, Khaw KT, Bingham SA. The CAFE computer program for nutritional analysis of the EPIC-Norfolk food frequency questionnaire and identification of extreme nutrient values. *J Hum Nutr Diet* 2005; **18**: 99–116.

103 Bingham SA, Luben R, Welch A, Wareham N, Khaw KT, Day N. Are imprecise methods obscuring a relation between fat and breast cancer? *Lancet* 2003; **362**: 212–4.

104 Prynne CJ, Wagemakers JJMF, Stephen AM, Wadsworth MEJ. Meat consumption after disaggregation of meat dishes in a cohort of British adults in 1989 and 1999 in relation to diet quality. *Eur J Clin Nutr* 2009; **63**: 660–6.

105 Bingham S, Luben R, Welch A, *et al.* Associations between dietary methods and biomarkers, and between fruits and vegetables and risk of ischaemic heart disease, in the EPIC Norfolk Cohort Study. *Int J Epidemiol* 2008; **37**: 978–87.

106 Cosgrove M, Flynn A, Kiely M. Impact of disaggregation of composite foods on estimates of intakes of meat and meat products in Irish adults. *Public Health Nutr* 2005; **8**: 327–37.

107 Bingham SA, Gill C, Welch A, *et al.* Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *Int J Epidemiol* 1997; **26**: S137–51.

108 McKeown NM, Day NE, Welch AA, *et al.* Use of biological markers to validate self-reported dietary intake in a random sample of the European Prospective Investigation into Cancer United Kingdom Norfolk cohort. *Am J Clin Nutr* 2001; **74**: 188–96.

109 Day NE, McKeown N, Wong MY, Welch A, Bingham S. Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *Int J Epidemiol* 2001; **30**: 309–17.

110 Bingham SA, Day NE. Using biochemical markers to assess the validity of prospective dietary assessment methods and the effect of energy adjustment. *Am J Clin Nutr* 1997; **65**: 1130S-1137S.

111 Bingham SA, Gill C, Welch A, *et al.* Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. *Int J Epidemiol* 1997; **26**: S137–51.

112 Bingham SA, Cassidy A, Cole TJ, *et al.* Validation of weighed records and other methods of dietary assessment using the 24 h urine nitrogen technique and other biological markers. *Br J Nutr* 1995; **73**: 531–50.

113 Forouhi NG, Wareham NJ. The EPIC-InterAct Study: a study of the interplay between genetic and lifestyle behavioral factors on the risk of type 2 diabetes in European populations. *Curr Nutr Rep* 2014; **3**: 355–63.

114 Langenberg C, Sharp S, Forouhi NG, *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* 2011; **54**: 2272–82.

115 Riboli E, Hunt K, Slimani N, *et al.* European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002; **5**: 1113–24.

116 Slimani N, Deharveng G, Unwin I, *et al.* The EPIC nutrient database project (ENDB): a first attempt to standardize nutrient databases across the 10 European countries participating in the EPIC study. *Eur J Clin Nutr* 2007; **61**: 1037–56.

117 Slimani N, Kaaks R, Ferrari P, *et al.* European Prospective Investigation into Cancer and Nutrition (EPIC) calibration study: rationale, design and population characteristics.

Public Health Nutr 2002; **5**: 1125–45.

118 Slimani N, Deharveng G, Charrondière RU, *et al.* Structure of the standardized computerized 24-h diet recall interview used as reference method in the 22 centers participating in the EPIC project. *Comput Methods Programs Biomed* 1999; **58**: 251–66.

119 Fry A, Littlejohns TJ, Sudlow C, *et al.* Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 2017; **186**: 1026–34.

120 Sudlow C, Gallacher J, Allen N, *et al.* UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; **12**: e1001779.

121 Bycroft C, Freeman C, Petkova D, *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562**: 203–9.

122 Bradbury KE, Young HJ, Guo W, Key TJ. Dietary assessment in UK Biobank: an evaluation of the performance of the touchscreen dietary questionnaire. *J Nutr Sci* 2018; **7**: 1–11.

123 Elliott P, Peakman TC. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. *Int J Epidemiol* 2008; **37**: 234–44.

124 Liu B, Young H, Crowe FL, *et al.* Development and evaluation of the oxford WebQ, a low-cost, web-based method for assessment of previous 24 h dietary intakes in large-scale prospective studies. *Public Health Nutr* 2011; **14**: 1998–2005.

125 Nicholson JK, Holmes E, Kinross JM, Darzi AW, Takats Z, Lindon JC. Metabolic phenotyping in clinical and surgical environments. *Nature* 2012; **491**: 384–92.

126 Pietzner M, Stewart ID, Raffler J, *et al.* Plasma metabolites to profile pathways in noncommunicable disease multimorbidity. *Nat Med* 2021; **27**: 471–9.

127 Shim J-S, Oh K, Kim HC. Dietary assessment methods in epidemiologic studies. *Epidemiol Health* 2014; **36**: e2014009.

128 Brennan L. Metabolomics in nutrition research: current status and perspectives. *Biochem Soc Trans* 2013; **41**: 670–3.

129 Guertin KA, Moore SC, Sampson JN, *et al.* Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. *Am J Clin Nutr* 2014; **100**: 208–17.

130 Brouwer-Brolsma EM, Brennan L, Drevon CA, *et al.* Combining traditional dietary assessment methods with novel metabolomics techniques: present efforts by the food biomarker alliance. *Proc Nutr Soc* 2017; **76**: 619–27.

131 O'Gorman A, Brennan L. The role of metabolomics in determination of new dietary biomarkers. *Proc Nutr Soc* 2017; **76**: 295–302.

132 Collins C, McNamara AE, Brennan L. Role of metabolomics in identification of biomarkers related to food intake. *Proc Nutr Soc* 2019; **78**: 189–96.

133 Scalbert A, Brennan L, Manach C, *et al.* The food metabolome: a window over dietary exposure. *Am J Clin Nutr* 2014; **99**: 1286–308.

134 Guasch-Ferre M, Bhupathiraju SN, Hu FB. Use of metabolomics in improving assessment of dietary intake. *Clin Chem* 2018; **64**: 82–98.

135 Posma JM, Garcia-Perez I, Frost G, *et al.* Nutriome–metabolome relationships provide insights into dietary intake and metabolism. *Nat Food* 2020; **1**: 426–36.

136 Maruvada P, Lampe JW, Wishart DS, *et al.* Perspective: dietary biomarkers of intake and exposure - exploration with omics approaches. *Adv Nutr* 2020; **11**: 200–15.

137 Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; **32**: 1–22.

138 Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Smith GD. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008; **27**: 1133–63.

139 Smith GD, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014; **23**: R89-98.

140 Emdin CA, Khera A V., Kathiresan S. Mendelian randomization. *JAMA* 2017; **318**: 1925–6.

141 Sanderson E, Glymour MM, Holmes M V., *et al.* Mendelian randomization. *Nat Rev Methods Prim* 2022; **2**: 1–21.

142 Swerdlow DI, Kuchenbaecker KB, Shah S, *et al.* Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int J Epidemiol* 2016; **45**: 1600–16.

143 Burgess S, Davey Smith G, Davies NM, *et al.* Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res* 2019; **4**: 186.

144 Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013; **37**: 658–65.

145 Bowden J, Smith GD, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015; **44**: 512–25.

146 Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016; **40**: 304–14.

147 Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med* 2016; **35**: 1880–906.

148 Hartwig FP, Smith GD, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017; **46**: 1985–98.

149 Teumer A. Common methods for performing Mendelian randomization. *Front Cardiovasc Med* 2018; **5**: 51.

150 Boehm FJ, Zhou X. Statistical methods for Mendelian randomization in genome-wide association studies: a review. *Comput Struct Biotechnol J* 2022; **20**: 2338–51.

151 DataSHIELD. https://www.datashield.org/.

152 Du H, Guo Y, Bennett DA, *et al.* Red meat, poultry and fish consumption and risk of diabetes: a 9 year prospective cohort study of the China Kadoorie Biobank. *Diabetologia* 2020; **63**: 767–79.

153 Beaney AJ, Banim PJR, Luben R, Lentjes MAH, Khaw KT, Hart AR. Higher meat intake is positively associated with higher risk of developing pancreatic cancer in an age-dependent manner and are modified by plasma antioxidants: a prospective cohort study (EPIC-Norfolk) using data from food diaries. *Pancreas* 2017; **46**: 672–8.

154 Cosgrove M, Flynn A, Kiely M. Impact of disaggregation of composite foods on estimates of intakes of meat and meat products in Irish adults. *Public Health Nutr* 2005; **8**: 327–37.

155 Bingham SA, Welch AA, McTaggart A, *et al.* Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. *Public Health Nutr* 2001; **4**: 847–58.

156 Cox DR. Regression models and life-tables. *J R Stat Soc Ser B* 1972; **34**: 187–202.

157 Hosseini-Esfahani F, Beheshti N, Koochakpoor G, Mirmiran P, Azizi F. Meat food group intakes and the risk of type 2 diabetes incidence. *Front Nutr* 2022; **9**: 891111.

158 Wood A, Denholm R, Hollings S, *et al.* Linked electronic health records for research on a nationwide cohort of more than 54 million people in England: data resource. *BMJ* 2021; **373**: n826.

159 Eastwood S V., Mathur R, Atkinson M, *et al.* Algorithms for the capture and adjudication of prevalent and incident diabetes in UK Biobank. *PLoS One* 2016; **11**: e0162388.

160 Hutcheon JA, Chiolero A, Hanley JA. Random measurement error and regression dilution bias. *BMJ* 2010; **340**: 1402–6.

161 Hoogendijk EO, Afilalo J, Ensrud KE, Kowal P, Onder G, Fried LP. Frailty: implications for clinical practice and public health. Lancet. 2019; **394**: 1365–75.

162 Bouvard V, Loomis D, Guyton KZ, *et al.* Carcinogenicity of consumption of red and processed meat. *Lancet Oncol* 2015; **16**: 1599–600.

163 Bar N, Korem T, Weissbrod O, *et al.* A reference map of potential determinants for the human serum metabolome. *Nature* 2020; **588**: 135–40.

164 Cheung W, Keski-Rahkonen P, Assi N, *et al.* A metabolomic study of biomarkers of meat and fish intake. *Am J Clin Nutr* 2017; **105**: 600–8.

165 Wedekind R, Kiss A, Keski-Rahkonen P, *et al.* A metabolomic study of red and processed

meat intake and acylcarnitine concentrations in human urine and blood. *Am J Clin Nutr* 2020; **112**: 381–8.

166 Cuparencu C, Rinnan Å, Dragsted LO. Combined markers to assess meat intake-human metabolomic studies of discovery and validation. *Mol Nutr Food Res* 2019; **63**: e1900106.

167 Mitry P, Wawro N, Rohrmann S, Giesbertz P, Daniel H, Linseisen J. Plasma concentrations of anserine, carnosine and pi-methylhistidine as biomarkers of habitual meat consumption. *Eur J Clin Nutr* 2019; **73**: 692–702.

168 Forouhi NG, Ye Z, Rickard AP, *et al.* Circulating 25-hydroxyvitamin D concentration and the risk of type 2 diabetes: results from the European Prospective Investigation into Cancer (EPIC)-Norfolk cohort and updated meta-analysis of Prospective studies. *Diabetologia* 2012; **55**: 2173–82.

169 Wang Y, Gapstur SM, Carter BD, *et al.* Untargeted metabolomics identifies novel potential biomarkers of habitual food intake in a cross-sectional study of postmenopausal women. *J Nutr* 2018; **148**: 932–43.

170 Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol* 2005; **67**: 301–20.

171 Abram S V., Helwig NE, Moodie CA, DeYoung CG, MacDonald AW, Waller NG. Bootstrap enhanced penalized regression for variable selection with neuroimaging data. *Front Neurosci* 2016; **10**: 344.

172 Bunea F, She Y, Ombao H, Gongvatana A, Devlin K, Cohen R. Penalized least squares regression methods and applications to neuroimaging. *Neuroimage* 2011; **55**: 1519–27.

173 Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 1970; **12**: 55–67.

174 Wedekind R, Keski-Rahkonen P, Robinot N, *et al.* Syringol metabolites as new biomarkers for smoked meat intake. *Am J Clin Nutr* 2019; **110**: 1424–33.

175 Sumner LW, Amberg A, Barrett D, *et al.* Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 2007; **3**: 211–21.

176 Onland-Moret NC, van der A DL, van der Schouw YT, *et al.* Analysis of case-cohort data: a comparison of different methods. *J Clin Epidemiol* 2007; **60**: 350–5.

177 Cuparencu C, Praticó G, Hemeryck LY, *et al.* Biomarkers of meat and seafood intake: an extensive literature review. *Genes Nutr* 2019; **14**: 35.

178 Farhangi MA. Gut microbiota-dependent trimethylamine N-oxide and all-cause mortality: findings from an updated systematic review and meta-analysis. *Nutrition* 2020; **78**: 110856.

179 Shortt C, Hasselwander O, Meynier A, *et al.* Systematic review of the effects of the intestinal microbiota on selected nutrients and non-nutrients. *Eur J Nutr* 2018; **57**: 25–49.

180 Khodorova N V., Rutledge DN, Oberli M, *et al.* Urinary metabolomics profiles associated to bovine meat ingestion in humans. *Mol Nutr Food Res* 2019; **63**: e1700834.

181 Braverman NE, Moser AB. Functions of plasmalogen lipids in health and disease. *Biochim Biophys Acta* 2012; **1822**: 1442–52.

182 Mazzilli KM, McClain KM, Lipworth L, *et al.* Identification of 102 correlations between serum metabolites and habitual diet in a metabolomics study of the Prostate, Lung, Colorectal, and Ovarian Cancer Trial. *J Nutr* 2020; **150**: 694–703.

183 Thürmann PA, Steffen J, Zwernemann C, *et al.* Plasma concentration response to drinks containing β-carotene as carrot juice or formulated as a water dispersible powder. *Eur J Nutr* 2002; **41**: 228–35.

184 Skeaff CM, Hodson L, McKenzie JE. Dietary-induced changes in fatty acid composition of human plasma, platelet, and erythrocyte lipids follow a similar time course. *J Nutr* 2006; **136**: 565–9.

185 Welch AA, Bingham SA, Ive J, *et al.* Dietary fish intake and plasma phospholipid n-3 polyunsaturated fatty acid concentrations in men and women in the European Prospective Investigation into Cancer-Norfolk United Kingdom cohort. *Am J Clin Nutr* 2006; **84**: 1330–9.

186 Li J, Guasch-Ferré M, Chung W, *et al.* The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. *Eur Heart J* 2020; **41**: 2645–56.

187 Delpino FM, Figueiredo LM. Does creatine supplementation improve glycemic control and insulin resistance in healthy and diabetic patients? A systematic review and meta-analysis. *Clin Nutr ESPEN* 2022; **47**: 128–34.

188 Gualano B, De Salles Painneli V, Roschel H, *et al.* Creatine in type 2 diabetes: a randomized, double-blind, placebo-controlled trial. *Med Sci Sports Exerc* 2011; **43**: 770–8.

189 Post A, Groothof D, Schutten JC, *et al.* Plasma creatine and incident type 2 diabetes in a general population-based cohort: the PREVEND study. *Clin Endocrinol* 2021; **94**: 563–74.

190 Mynatt RL. Carnitine and type 2 diabetes. *Diabetes Metab Res Rev* 2009; **25**: S45.

191 Papandreou C, Bulló M, Zheng Y, *et al.* Plasma trimethylamine-N-oxide and related metabolites are associated with type 2 diabetes risk in the Prevención con Dieta Mediterránea (PREDIMED) trial. *Am J Clin Nutr* 2018; **108**: 163–73.

192 Li S, Chen S, Lu X, Fang A. Serum trimethylamine-N-oxide is associated with incident type 2 diabetes in middle-aged and older adults: a prospective cohort study. *J Transl Med* 2022; **20**: 1–10.

193 Dambrova M, Latkovskis G, Kuka J, *et al.* Diabetes is associated with higher trimethylamine N-oxide plasma levels. *Exp Clin Endocrinol Diabetes* 2016; **124**: 251–6.

194 Li P, Zhong C, Li S, *et al.* Plasma concentration of trimethylamine-N-oxide and risk of gestational diabetes mellitus. *Am J Clin Nutr* 2018; **108**: 603–10.

195 Wang TJ, Larson MG, Vasan RS, *et al.* Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011; **17**: 448–53.

196 Meikle PJ, Wong G, Barlow CK, *et al.* Plasma lipid profiling shows similar associations with prediabetes and type 2 diabetes. *PLoS One* 2013; **8**: e74341.

197 Davies NM, Holmes M V., Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* 2018; **362**: 601.

198 McCarthy S, Das S, Kretzschmar W, *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; **48**: 1279–83.

199 Walter K, Min JL, Huang J, *et al.* The UK10K project identifies rare variants in health and disease. *Nature* 2015; **526**: 82–9.

200 Auton A, Abecasis GR, Altshuler DM, *et al.* A global reference for human genetic variation. *Nature* 2015; **526**: 68–74.

201 Surendran P, Stewart ID, Au Yeung VPW, *et al.* Rare and common genetic determinants of metabolic individuality and their effects on human health. *Nat Med* 2022; **28**: 2321–32.

202 Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* 2010; **11**: 499–511.

203 Mahajan A, Taliun D, Thurner M, *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018; **50**: 1505–13.

204 Mathias RA, Vergara C, Gao L, *et al.* FADS genetic variants and ω-6 polyunsaturated fatty acid metabolism in a homogeneous island population. *J Lipid Res* 2010; **51**: 2766–74.

205 Reynolds LM, Dutta R, Seeds MC, *et al.* FADS genetic and metabolomic analyses identify the Δ5 desaturase (FADS1) step as a critical control point in the formation of biologically important lipids. *Sci Rep* 2020; **10**: 1–12.

206 Moseti D, Regassa A, Kim WK. Molecular regulation of adipogenesis and potential anti-adipogenic bioactive molecules. *Int J Mol Sci* 2016; **17**: 124.

207 Merino DM, Johnston H, Clarke S, *et al.* Polymorphisms in FADS1 and FADS2 alter desaturase activity in young Caucasian and Asian adults. *Mol Genet Metab* 2011; **103**: 171–8.

208 He Z, Zhang R, Jiang F, *et al.* FADS1-FADS2 genetic polymorphisms are associated with fatty acid metabolism through changes in DNA methylation and gene expression. *Clin Epigenetics* 2018; **10**: 1–13.

209 Shetty SS, Suchetha KN, Devi H, Sharmila KP, Rai S. Association of FADS2 rs174575 gene polymorphism and insulin resistance in type 2 diabetes mellitus. *Afr Health Sci* 2020; **20**: 1770–6.

210 Shetty SS, Kumari NS. Fatty acid desaturase 2 (FADS 2) rs174575 (C/G) polymorphism,

circulating lipid levels and susceptibility to type-2 diabetes mellitus. *Sci Rep* 2021; **11**: 1–8.

211 Mazoochian L, Mohammad Sadeghi HM, Pourfarzam M. The effect of FADS2 gene rs174583 polymorphism on desaturase activities, fatty acid profile, insulin resistance, biochemical indices, and incidence of type 2 diabetes. *J Res Med Sci* 2018; **23**: 47.

212 Paul S, Lancaster GI, Meikle PJ. Plasmalogens: a potential therapeutic target for neurodegenerative and cardiometabolic disease. *Prog Lipid Res* 2019; **74**: 186–95.

213 Huynh K, Barlow CK, Jayawardana KS, *et al.* High-throughput plasma lipidomics: detailed mapping of the associations with cardiometabolic risk factors. *Cell Chem Biol* 2019; **26**: 71-84.e4.

214 Yousri NA, Suhre K, Yassin E, *et al.* Metabolic and metabo-clinical signatures of type 2 diabetes, obesity, retinopathy, and dyslipidemia. *Diabetes* 2022; **71**: 184–205.

215 Jia J, Dou P, Gao M, *et al.* Assessment of causal direction between gut microbiota-dependent metabolites and cardiometabolic health: a bidirectional Mendelian randomization analysis. *Diabetes* 2019; **68**: 1747–55.

216 Rhee EP, Ho JE, Chen MH, *et al.* A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metab* 2013; **18**: 130–43.

217 Yang X, Li Y, Wang C, *et al.* Meat and fish intake and type 2 diabetes: dose-response meta-analysis of prospective cohort studies. *Diabetes Metab* 2020; **46**: 345–52.

218 Celis-Morales C, Livingstone KM, Marsaux CFM, *et al.* Effect of personalized nutrition on health-related behaviour change: evidence from the Food4Me European randomized controlled trial. *Int J Epidemiol* 2017; **46**: 578–88.

219 Ordovas JM, Ferguson LR, Tai ES, Mathers JC. Personalised nutrition and health. *BMJ* 2018; **361**: k2173.

220 Zhuang P, Liu X, Li Y, *et al.* Effect of diet quality and genetic predisposition on hemoglobin A1c and type 2 diabetes risk: gene-diet interaction analysis of 357,419 individuals. *Diabetes Care* 2021; **44**: 2470–9.

221 Qi L, Cornelis MC, Zhang C, Van Dam RM, Hu FB. Genetic predisposition, Western dietary pattern, and the risk of type 2 diabetes in men. *Am J Clin Nutr* 2009; **89**: 1453–

8.

222 Cai L, Wheeler E, Kerrison ND, *et al.* Genome-wide association analysis of type 2 diabetes in the EPIC-InterAct Study. *Sci Data* 2020; **7**: 1–6.

223 Lotta LA, Gulati P, Day FR, *et al.* Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat Genet* 2017; **49**: 17–26.

224 Locke AE, Kahali B, Berndt SI, *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518**: 197–206.

225 Wand H, Lambert SA, Tamburro C, *et al.* Improving reporting standards for polygenic scores in risk prediction studies. *Nature* 2021; **591**: 211–9.

226 Wheeler E, Leong A, Liu CT, *et al.* Impact of common genetic determinants of hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: a transethnic genome-wide meta-analysis. *PLoS Med* 2017; **14**: e1002383.

227 UK Biobank Biomarker Enhancement Project. https://biobank.ctsu.ox.ac.uk/crystal/ukb/docs/serum_hb1ac.pdf.

228 Rahman M, Simmons RK, Harding AH, Wareham NJ, Griffin SJ. A simple risk score identifies individuals at high risk of developing type 2 diabetes: a prospective cohort study. *Fam Pract* 2008; **25**: 191–6.

229 Langenberg C, Sharp SJ, Franks PW, *et al.* Gene-lifestyle interaction and type 2 diabetes: the EPIC InterAct case-cohort Study. *PLoS Med* 2014; **11**: 29.

230 Ericson U, Hindy G, Drake I, *et al.* Dietary and genetic risk scores and incidence of type 2 diabetes. *Genes Nutr* 2018; **13**: 1–14.

231 Fretts AM, Follis JL, Nettleton JA, *et al.* Consumption of meat is associated with higher fasting glucose and insulin concentrations regardless of glucose and insulin genetic risk scores: a meta-analysis of 50,345 Caucasians. *Am J Clin Nutr* 2015; **102**: 1266–78.

232 Pearce M, Fanidi A, Bishop TRP, *et al.* Associations of total legume, pulse, and soy consumption with incident type 2 diabetes: federated meta-analysis of 27 studies from diverse world regions. *J Nutr* 2021; **151**: 1231–40.

233 Pastorino S, Bishop T, Sharp SJ, *et al.* Heterogeneity of associations between total and types of fish intake and the incidence of type 2 diabetes: federated meta-analysis of 28 prospective studies including 956,122 participants. *Nutrients* 2021; **13**: 1223.

234 Malik VS, Li Y, Tobias DK, Pan A, Hu FB. Dietary protein intake and risk of type 2 diabetes in US men and women. *Am J Epidemiol* 2016; **183**: 715–28.

235 Fan M, Li Y, Wang C, *et al.* Dietary protein consumption and the risk of type 2 diabetes: a dose-response meta-analysis of prospective studies. *Nutrients* 2019; **11**: 2783.

236 Li J, Glenn AJ, Yang Q, *et al.* Dietary protein sources, mediating biomarkers, and incidence of type 2 diabetes: findings from the Women's Health Initiative and the UK Biobank. *Diabetes Care* 2022; **45**: 1742–53.

237 Satija A, Yu E, Willett WC, Hu FB. Understanding nutritional epidemiology and its role in policy. *Adv Nutr* 2015; **6**: 5–18.

238 Rampler E, Hermann G, Grabmann G, *et al.* Benchmarking non-targeted metabolomics using yeast-derived libraries. *Metabolites* 2021; **11**: 160.

239 Scanlon PH. The contribution of the English NHS Diabetic Eye Screening Programme to reductions in diabetes-related blindness, comparisons within Europe, and future challenges. *Acta Diabetol* 2021; **58**: 521–30.

240 Wareham NJ. Personalised prevention of type 2 diabetes. *Diabetologia* 2022; **65**: 1796–803.

241 Nathan DM, Balkau B, Bonora E, *et al.* International expert committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care* 2009; **32**: 1327–34.

242 Cai L, Gonzales T, Wheeler E, Kerrison N, Day FR, Langenberg C et al. Causal associations between cardiorespiratory fitness and type 2 diabetes. *Submiss*.

Supplementary Information

Data sources for T2D ascertainment in EPIC-Norfolk

In EPIC-Norfolk, the following data sources were used for the ascertainment of diabetes cases: 1) self-report of diabetes diagnosis and self-report of diabetic medication use by questionnaires during follow-up surveys; 2) biochemistry measurements of glycated haemoglobin (HbA1c) in the second health examination; 3) record linkage to external databases, including a) medical records from NHS Hospital Episode Statistics(HES), b) death certification from the Office of National Statistics, c) diabetic retinopathy screening data and relevant data from eye clinic in Norfolk and Norwich University Hospital (NNUH), and 4) clinical biochemistry measurements of HbA1c in NNUH.

Record linkage with HES

We obtained medical records from HES, which is a national-based database containing information of hospital admission, outpatients, and emergency care attendance at NHS hospitals in England. The current available dataset in this study included medical records from April 1997 to March 2020.

Record linkage with Death Registry

Death records were obtained from the Office of National Statistics, which provided the cause of death (including diabetes). Data were available until to March 2020.

Diabetic eye screening programme (DESP), Medisoft and Laser records

In EPIC-Norfolk, we also linked with the DESP data and relevant databases as a unique complementary source to ascertain T2D cases. The DESP is a community-wide programme to detect diabetic retinopathy, a complication of diabetes caused by damaged blood vessels in the back of eyes (retina). Eye screening is a crucial part of diabetes care as it is important to find eye problems early before changes of vision occur. Early detection and prompt treatment can prevent sight loss and blindness[23]>.

The DESP in the Norfolk area has been in place since 1990 at the Norfolk and Norwich University Hospital (NNUH) and was taken over by the National DESP in 2006. All residents

who live in the Norfolk area (North Norfolk, South Norfolk, Norwich) are eligible for eye screening if they had a diagnosis of diabetes and are over 13 years old. Information about the eligible population is shared from General Practices to DESP in Norfolk automatically and then an appointment is offered to participants. Therefore, we linked participants in the EPIC-Norfolk study with the DESP database to help identify diabetes cases. One reason for not being offered diabetic retinopathy screening is if patients are already under follow-up in the eye clinic. Therefore, we also linked EPIC-Norfolk participants with Medisoft ophthalmic records from NNUH and laser treatment records in eye clinics. The linkage data were available from January 1990 to March 2020.

Matching EPIC-Norfolk participants to participants in DESP was based on NHS number or NNUH hospital number if available. Participants with older DESP data does not have NHS or hospital numbers available and hence these people were matched using other identifiers, such as postcode, date of birth, sex, and surname/forename. There were 3421 individuals linked with DESP data, including 33 individuals who were identified by linkage with the Medisoft database and 44 with the laser treatment records.

Biochemistry measurements in the NNUH Lab

HbA1c is widely used to monitor glycemic control of people with diabetes. It is also now increasingly used to help the diagnosis of T2D and screening for individuals with or at high risk of diabetes. HbA1c has been recommended as a diagnostic test because it reflects average blood glucose over the preceding two to three months, and does not need to be measured when fasting[24]>. Many diabetes cases are unsymptomatic and undetected for years. The use of HbA1c can identify undetected cases and reduce the delay in diagnosis. In the EPIC-Norfolk study, we were able to link participants with clinical measured HbA1c in NNUH recorded from May 2009 to Jan 2022, which provide additional information to improve the ascertainment of T2D cases.

We were able to link the whole cohort with the HbA1c measurements in NNUH and 15,073 people had at least one HbA1c measurement, with an average of 6 records per person. There were 3,333 individuals with at least one measurements of HbA1c≥48 mmol/mol.

Information of T2D in EPIC-Norfolk surveys

In EPIC-Norfolk baseline and follow-up surveys (1993-2018), participants were asked the following questions relevant to diabetes through questionnaires: 1) whether they were told by the doctor that they have diabetes previously (yes or no) and 2) information of drugs that were taken in the past week. This information has been entered into the EPIC-Norfolk database as free-text records in unstructured formats of synonymous terms, such as generic names or brand names. Generic names are names of active ingredients of the drugs and brand names are proprietary names generated by pharmacological companies. Some drugs were reported using names of mixture products. As they are entered in free text, many drug names have typos. For example, metformin or metformin hydrochloride, is a commonly prescribed glucose-lowering therapy. In the EPIC Norfolk database, the use of metformin was reported in over 100 different formats, such as "glucophage", "pioglitazone/metformin", "rosiglitazone & metformin tabs 2+500mg", “metfarmin", "metfobmin", or “metformir".

During the 2nd health examination (1998-2000), which included 15, 000 participants, HbA1c levels were measured using a Biorad Diomat high pressure liquid chromatography analyser from ethylenediaminetetraacetic acid (EDTA) anticoagulated blood samples examined at the Department of Clinical Biochemistry, University of Cambridge.

Matching self-reported drug names with diabetic drugs

We identified glucose-lowering drug usage from self-reported information using string matching approaches. The identified anti-diabetic drugs were manually reviewed subsequently.

Firstly, I compiled a diabetes drug list to include standard names of drugs and products that were commonly prescribed for the treatment of type 2 diabetes using the following data sources: 1) the British National Formulary (BNF), which lists oral glucose-lowering drugs and insulin in their generic names; 2) UK Biobank BNF data dictionary, which provides brand names of BNF drugs; 3) Other online drug databases, including DrugBank and PHARMGKB, which enable searches for additional synonyms of diabetes drugs.

Then diabetes drug names in the compiled list were used to match with strings of self-reported drug context in the EPIC-Norfolk study using Pattern Matching approaches. A pattern represents a specific string character of a drug name. Matching two strings (a self-

reported name vs. a name for reference) using a pattern can traverse complex string structures, account for typo errors, and bring about comprehensive results of possible diabetes drugs. I applied multiple patterns for each diabetes drug and selected those which can accurately match with drug names via a manual check. Specifically, I started from a pattern that was made up of the first five letters of a drug name, applied to the whole data and then modified it by adding or reducing string characters.

Data processing and quality control of the untargeted metabolomics data

The detail methods have been reported previously[126]. Several types of control assays were performed in concert with the experimental samples to guarantee the instrument and process variability by Metabolon when metabolites were detected. Raw data were extracted, peak-identified, quality control processed and curated using the hardware and software of Metabolon to ensure the data quality. Then metabolomics data were quantified and processed through normalisation setting median to equal one to account for day-to-day instrumental variation. Then we obtained the data for analyses.

Bootstrapping enhanced elastic net regression

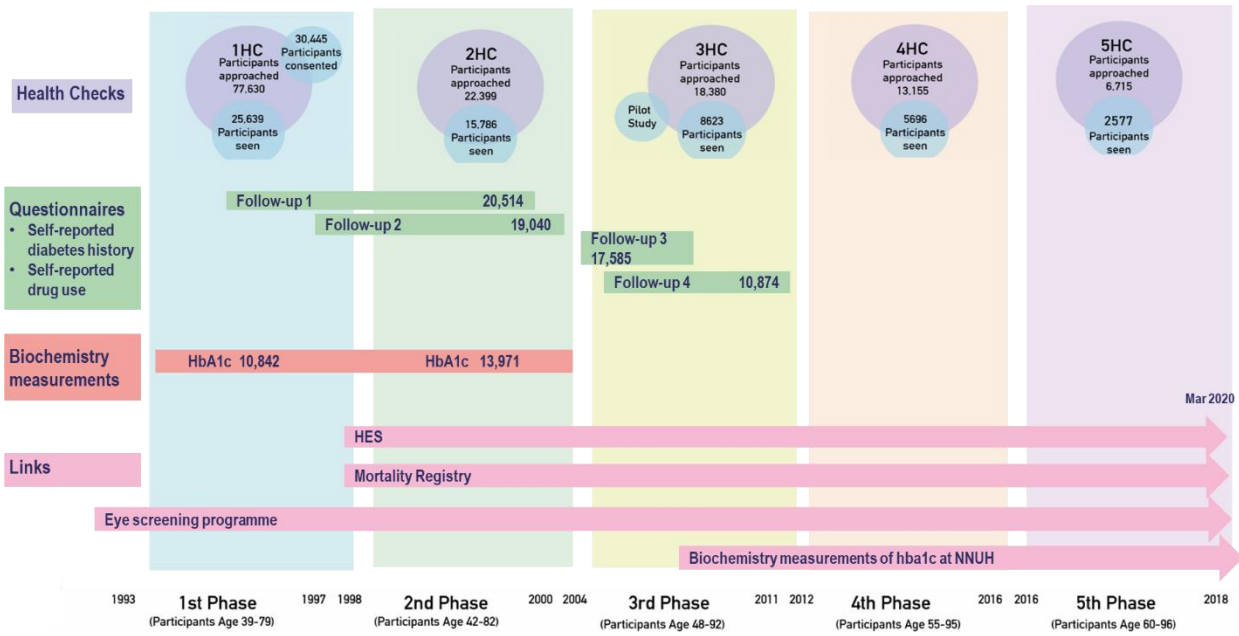
We applied elastic net penalised regression[170] in a two-stage design in the exploratory set (n=11,432, Figure 3.1) to derive a list of metabolites predictive of meat consumption. The inner stage ran one iteration of the elastic net using 10-fold cross validation for parameter optimisation and the outer stage generated random subsets of the validation set. For parameter optimisation, two penalisation parameters were applied (α and λ) to regularise the correlated metabolites by which we identified metabolites predictive of meat consumption. We pre-selected α values from 0.1 to 0.9 grid by 0.05 and then we selected the optimal λ and α values for the final model by 10-fold cross-validation, which optimised predictive performance by minimizing the cross-validated mean-squared error. Using the regularisation parameters obtained, we fitted the elastic net regression using the *glmnet* package in R. We used bootstrapping strategy to generating random subsets of the validation set using 80% of the size of the dataset. Metabolites were defined as candidates to construct metabolite scores if they had non-zero coefficient estimates over 90 times or more out of 100 bootstrapping resampling[171–173].

Diabetes ascertainment in EPIC-InterAct and UK Biobank

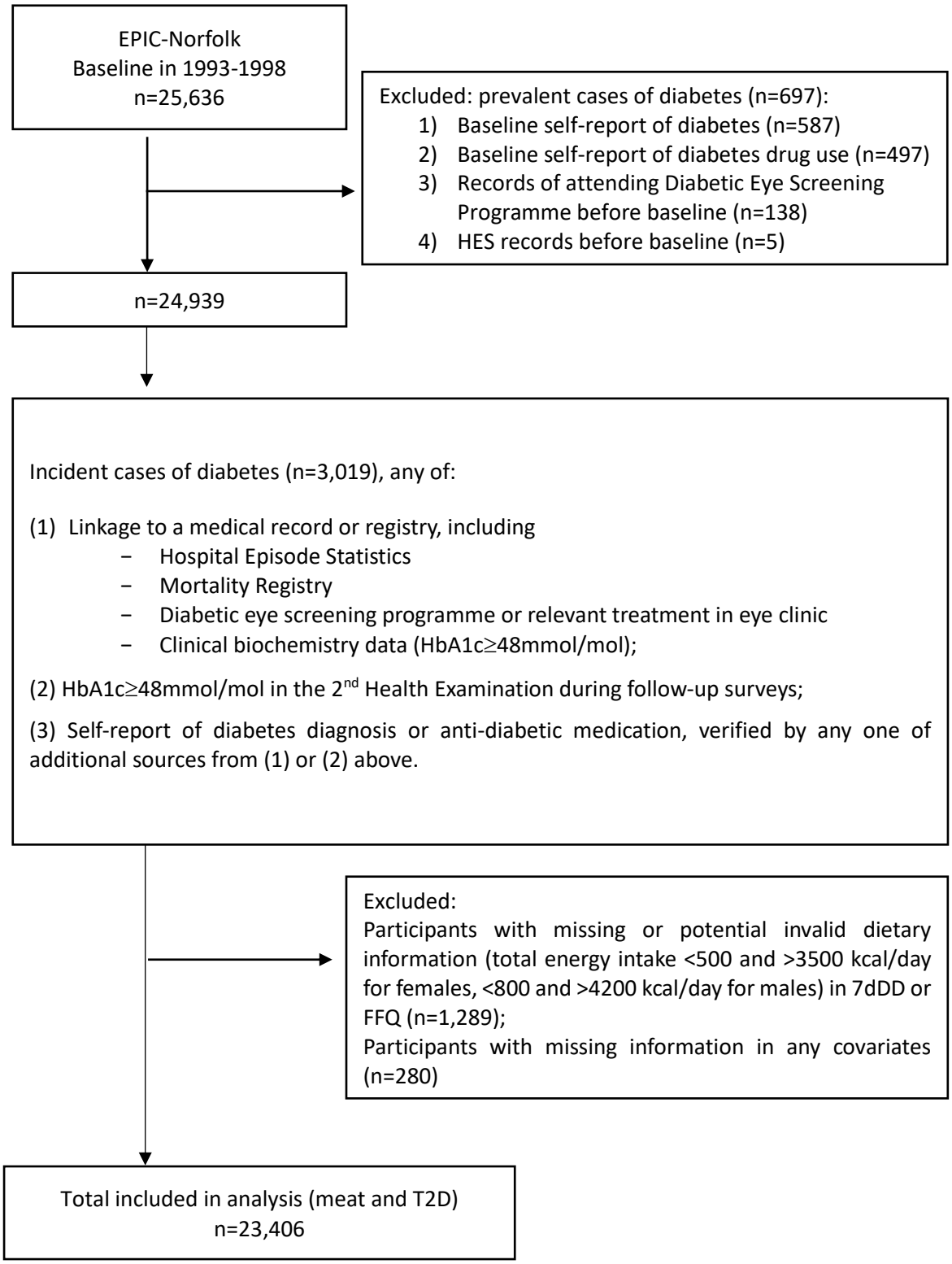
In EPIC-InterAct, T2D were ascertained and verified in each research centre using multiple sources, including the self-reported history of T2D, diabetes drug use, doctor-diagnosed T2D, linkage to primary care registers, secondary care registers, pharmacy registers, hospital admissions data, or mortality data. Cases in Denmark and Sweden were not ascertained by self-report, but identified via local and national diabetes and pharmaceutical registers.

Algorithms to identify prevalent and incident diabetes in UK Biobank were developed by colleagues at MRC Epidemiology Unit, University of Cambridge. Prevalent diabetes were defined as individuals who either self-reported any diabetes other than gestational diabetes, or self-reported diabetes drug usage (acarbose, glitazones, meglitinides, sulfonylureas, or insulin) at baseline, or had a HES event (ICD-10 codes E10-E14) prior to accelerometry[24]>. Incident T2D was ascertained using HES and mortality data with ICD code E11 without E10, or E14 without E10-E13. Diagnosis date was defined using the mid-point between the last record without T2D and the date of the first record with T2D.

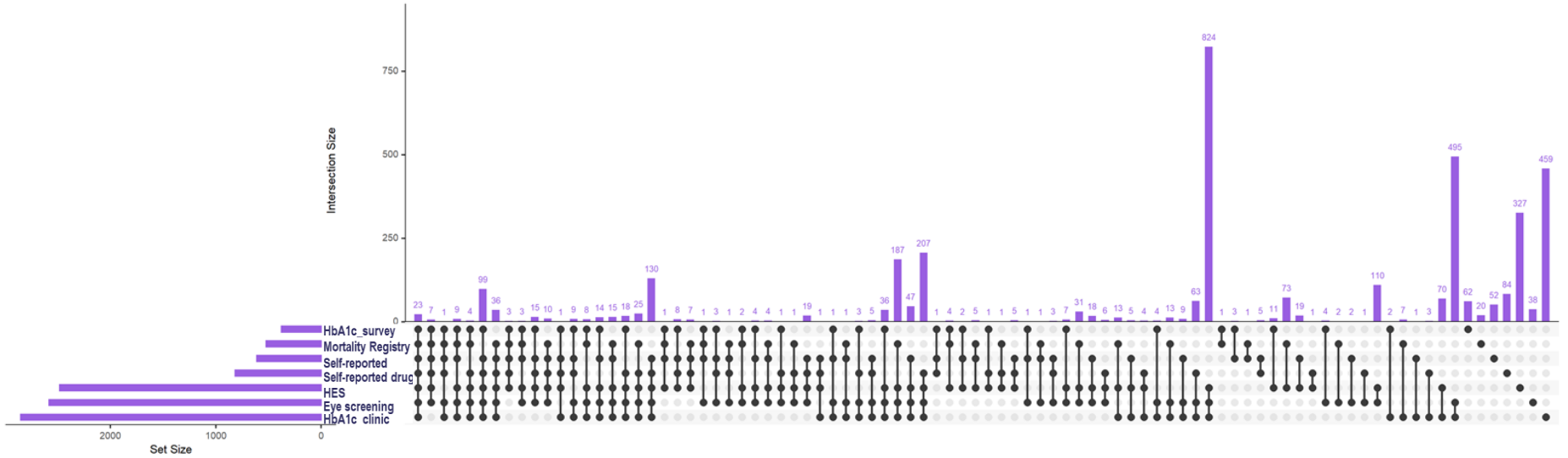
Supplementary Figures



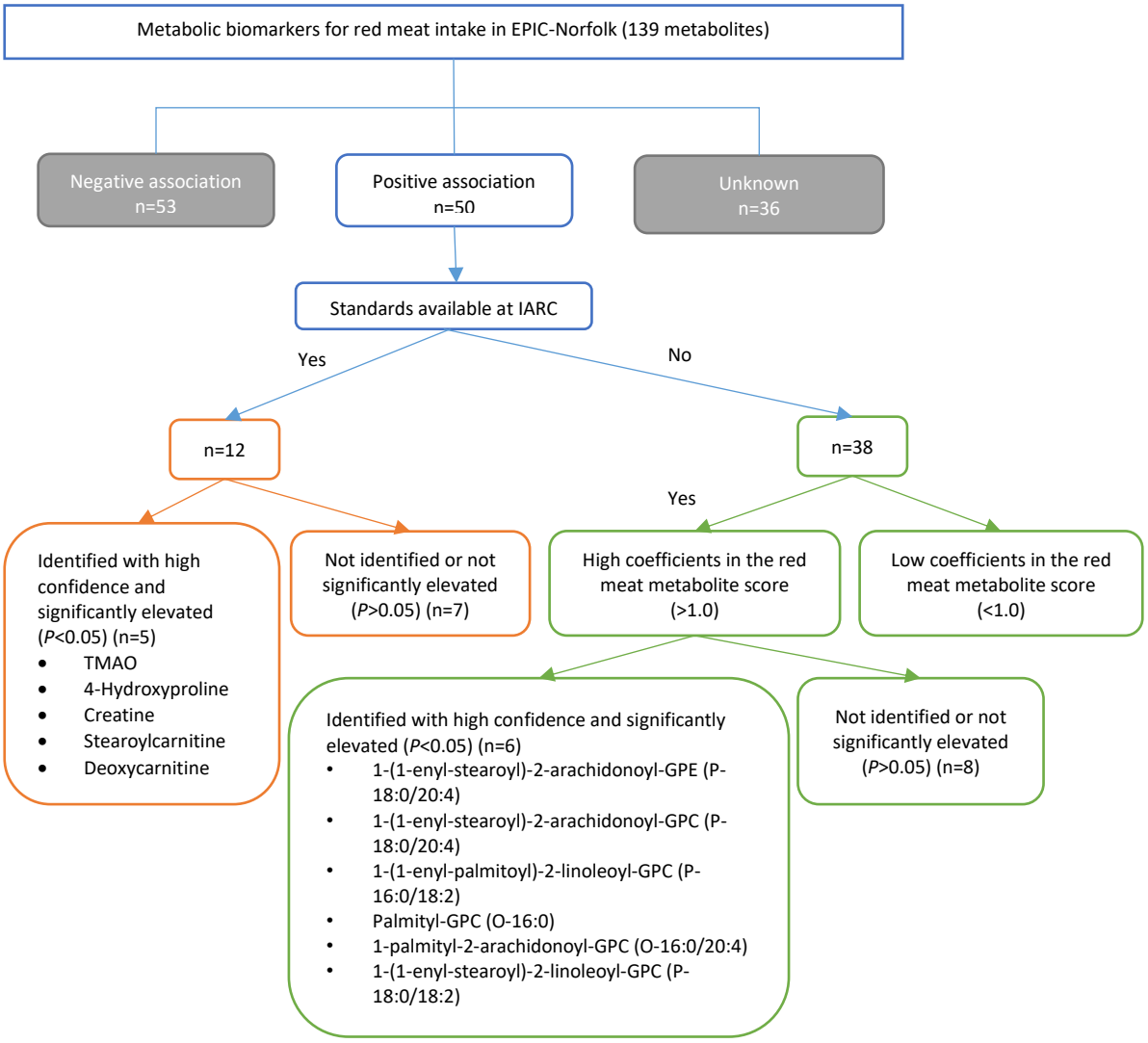
Supplementary Figure 2.1 The time line of multiple data sources used for T2D case ascertainment in the EPIC-Norfolk study



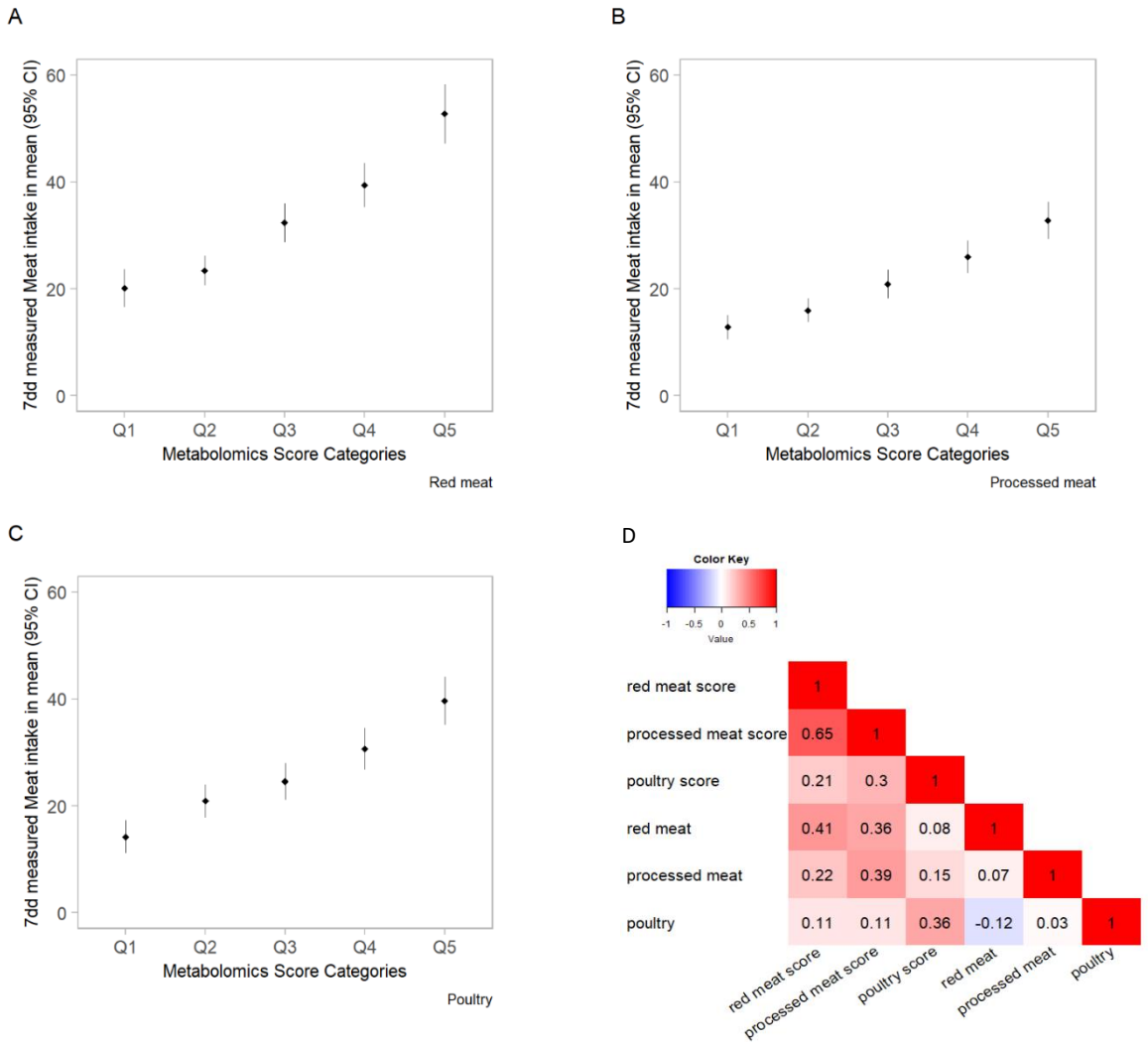
Supplementary Figure 2.2 Flow chart of inclusion of participants through the study



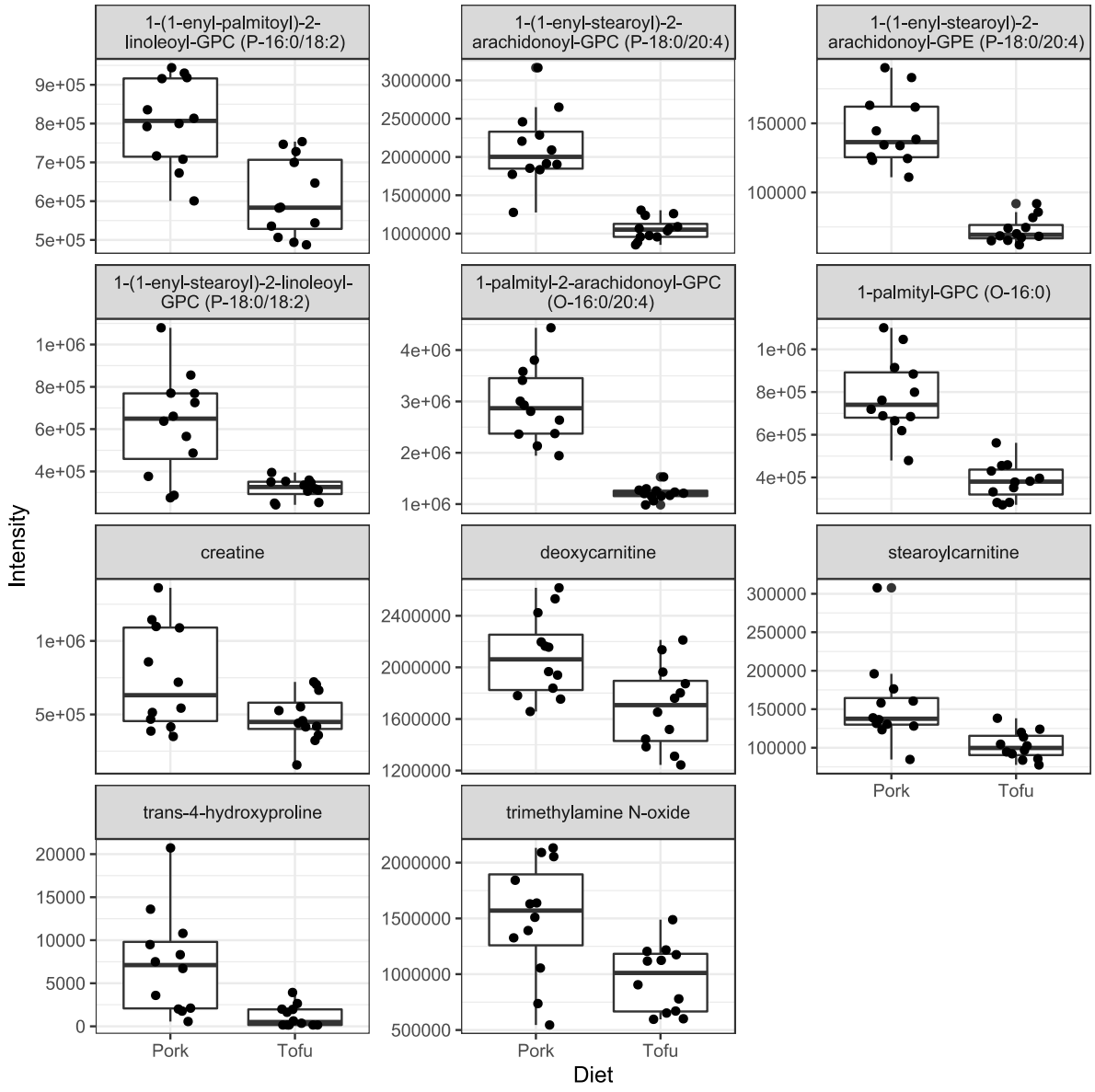
Supplementary Figure 2.3 The intersection of different data sources to ascertain incident type 2 diabetes cases in EPIC-Norfolk after exclusion of prevalent diabetes.



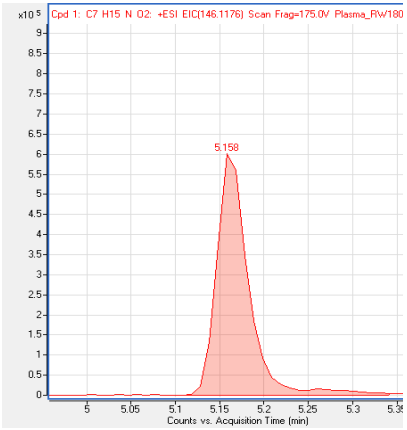
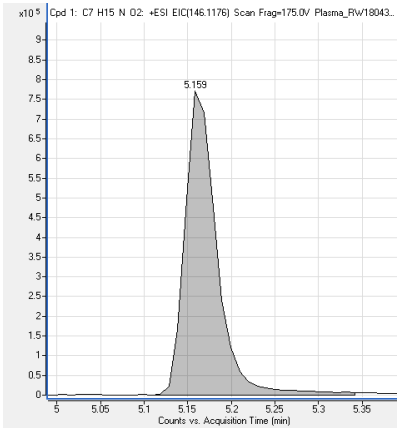
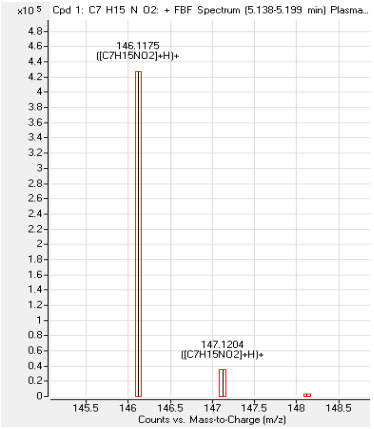
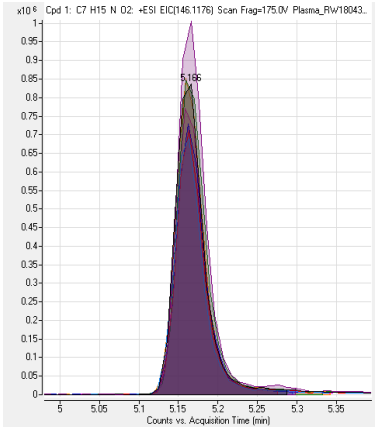
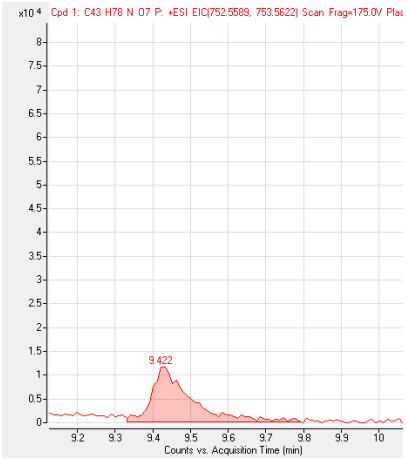
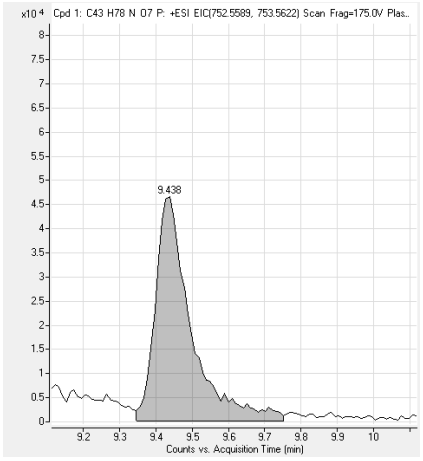
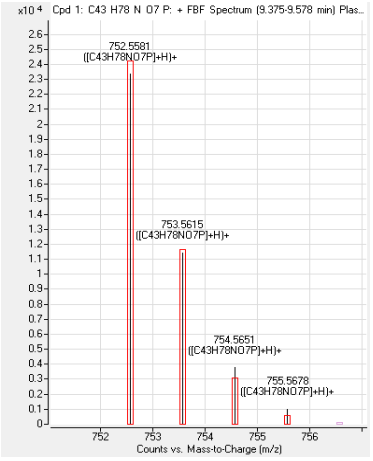
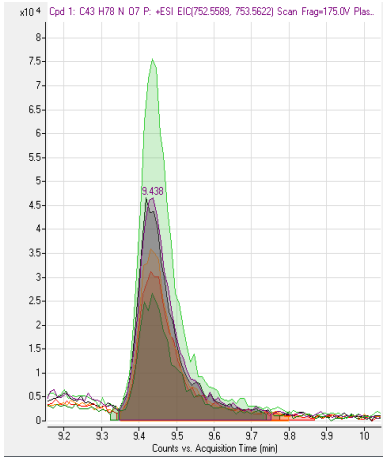
Supplementary Figure 3.1 Flowchart for identification of metabolites that make up the red meat metabolite score in the trial. GPE, glycerophosphoethanolamine; GPC, glycerophosphocholine; IARC, International Agency for Research on Cancer; TMAO, trimethylamine N-oxide.



Supplementary Figure 3.2 The correlations between meat scores and 7-day diet diary (7dDD) measured meat intake. A, the means and 95% confidence intervals (CIs) of red meat consumption measured by 7dDD in quintiles of the derived red meat metabolite score (139 metabolites) in the exploratory set in the EPIC-Norfolk study (n=11,432); B, the means and 95% CIs of processed meat consumption measured by 7dDD in quintiles of the processed meat metabolite score (82 metabolites) in the exploratory set in the EPIC-Norfolk study (n=11,432); C, the means and 95% CIs of poultry consumption measured by 7dDD in quintiles of the poultry metabolite score (139 metabolites) in the exploratory set in the EPIC-Norfolk study (n=11,432); D. the correlations matrix for consumption of types of meat (red meat, processed meat and poultry) measured by 7dDD and measured by derived metabolite scores in the validation set in the EPIC-Norfolk study (n=853).

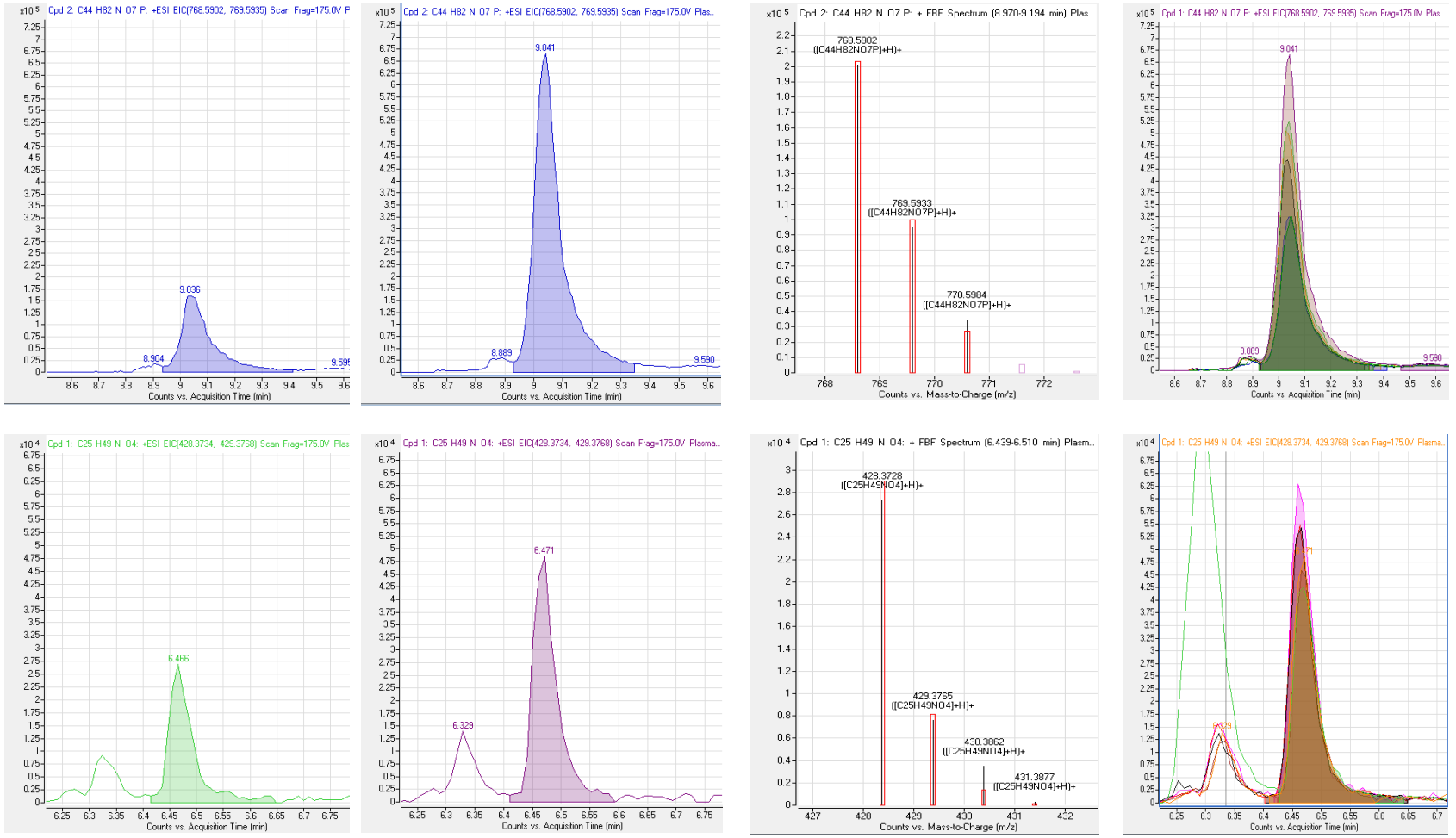


Supplementary Figure 3.3 Plasma levels of selected metabolites after consumption of pork and tofu in the randomised cross-over trial. Metabolites that were positively associated with red meat consumption in both the EPIC-Norfolk and the randomised cross-over trial are shown. Fold-change and *P* values are reported in Table 3.2. This figure is provided by Dr Roland Wedekind.

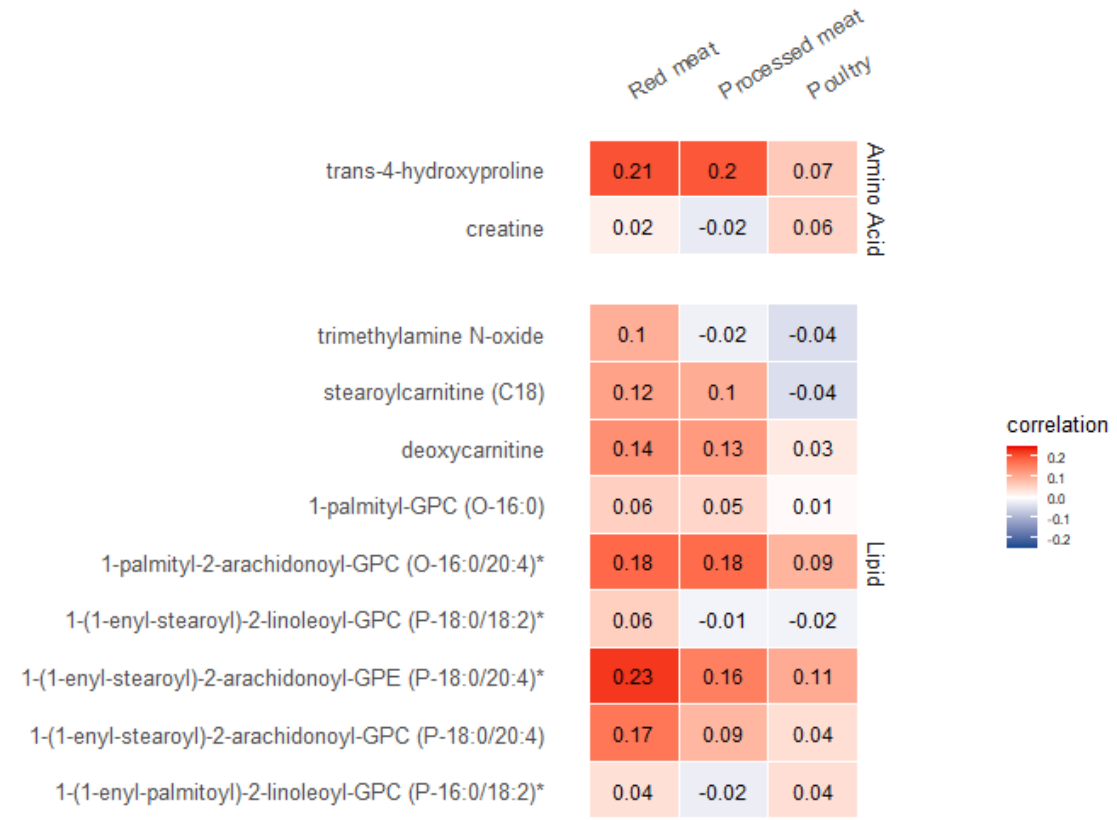
	Chromatogram in plasma after tofu intake	Chromatogram in plasma after pork intake	Isotope pattern	Chromatograms in plasma samples after pork intake
Deoxycarnitine				
1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4)				

1-(1-enyl-stearoyl)-2-linoleoyl-GPC (P-18:0/18:2)

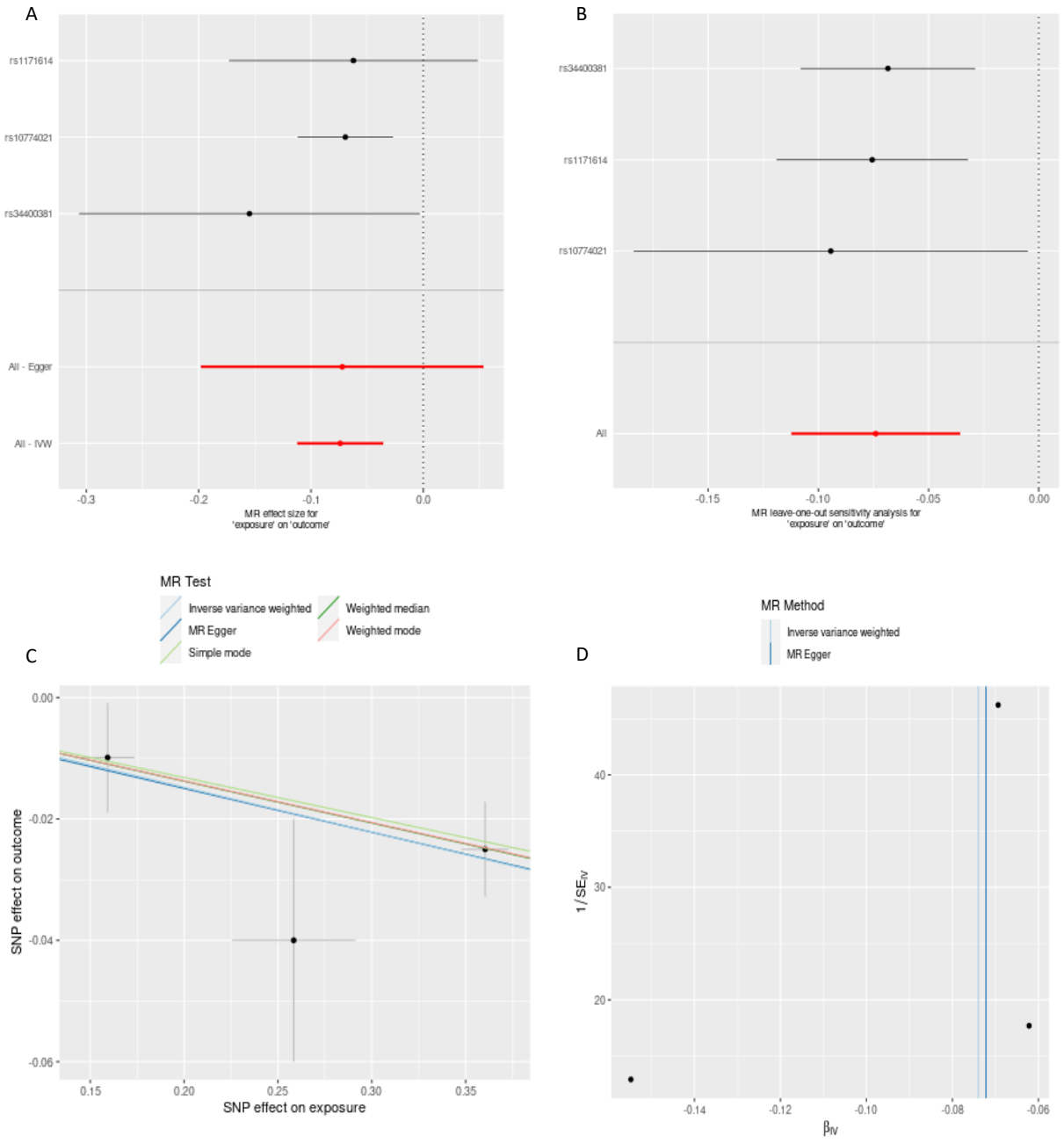
Stearoylcarnitine



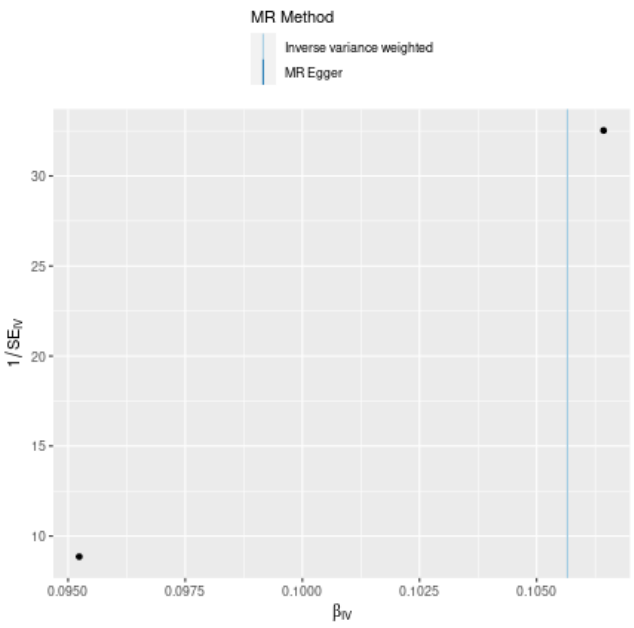
Supplementary Figure 3.4 Chromatographic tracing of selected metabolites after consumption of pork vs. tofu in the intervention study. Column 2 and 3 show the chromatogram of a compound after tofu and pork intake separately in the same participant. Isotope pattern was used as one indicator of the peak quality. The vertical lines represent the detected intensities of compounds. The boxes show the expected peaks. The plots indicate that high intensity compounds usually match very well with the expected isotope pattern. Column 5 shows the chromatogram of a compound in several samples of plasma after pork intake. It shows the variability of peak shapes and intensities (the variation of intensity of metabolites is reported in the boxplots in Supplementary Figure 3.3).



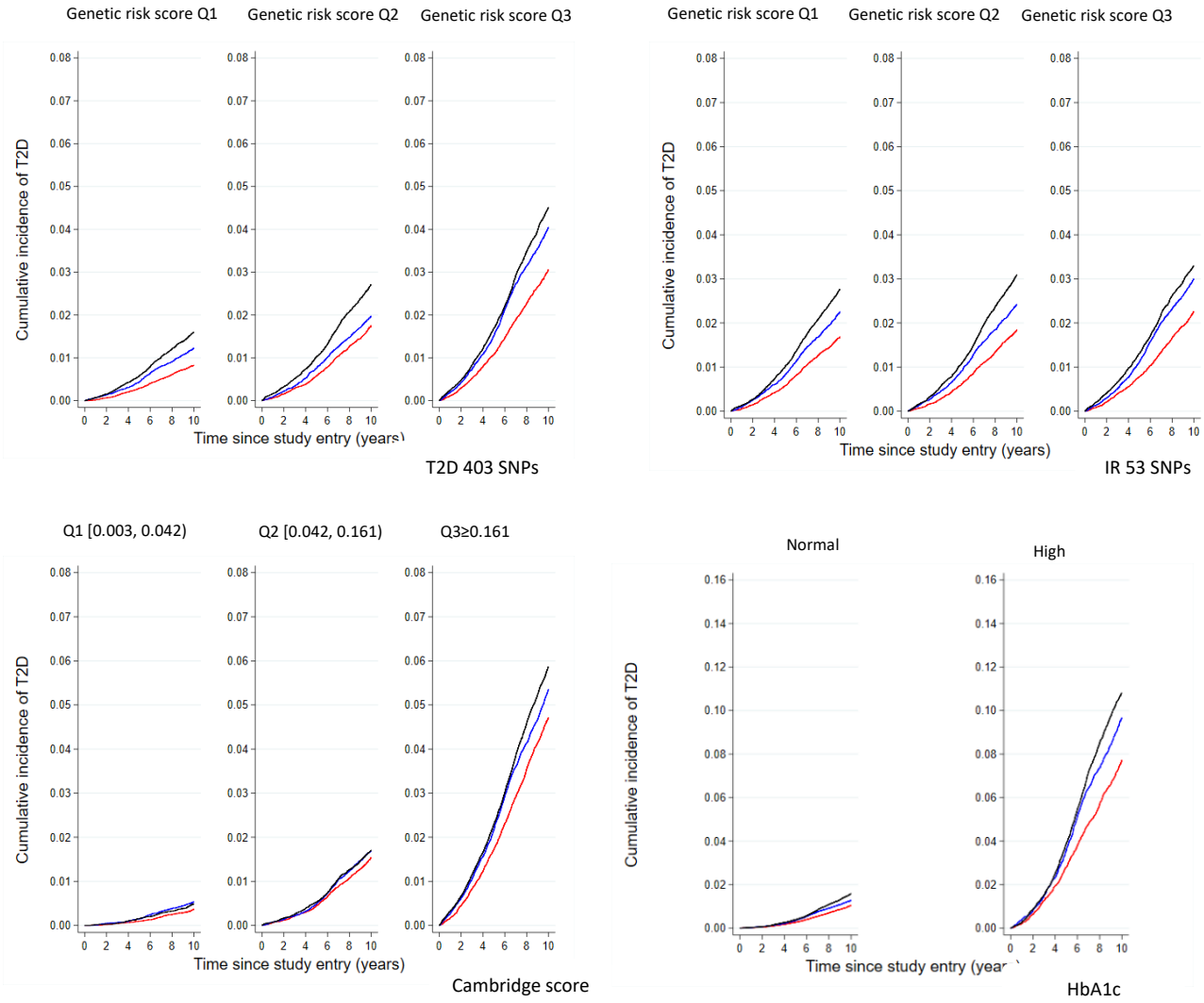
Supplementary Figure 3.5 Heatmap of correlations between types of meat consumption and top-ranked metabolites (n=11) in the red meat metabolite score that validated in the intervention study: EPIC-Norfolk study (n=11,432). The single asterisk in metabolite name represents the metabolite was annotated based on in-silico predictions which indicates the compound has not been confirmed based on a standard but its identity is confident.



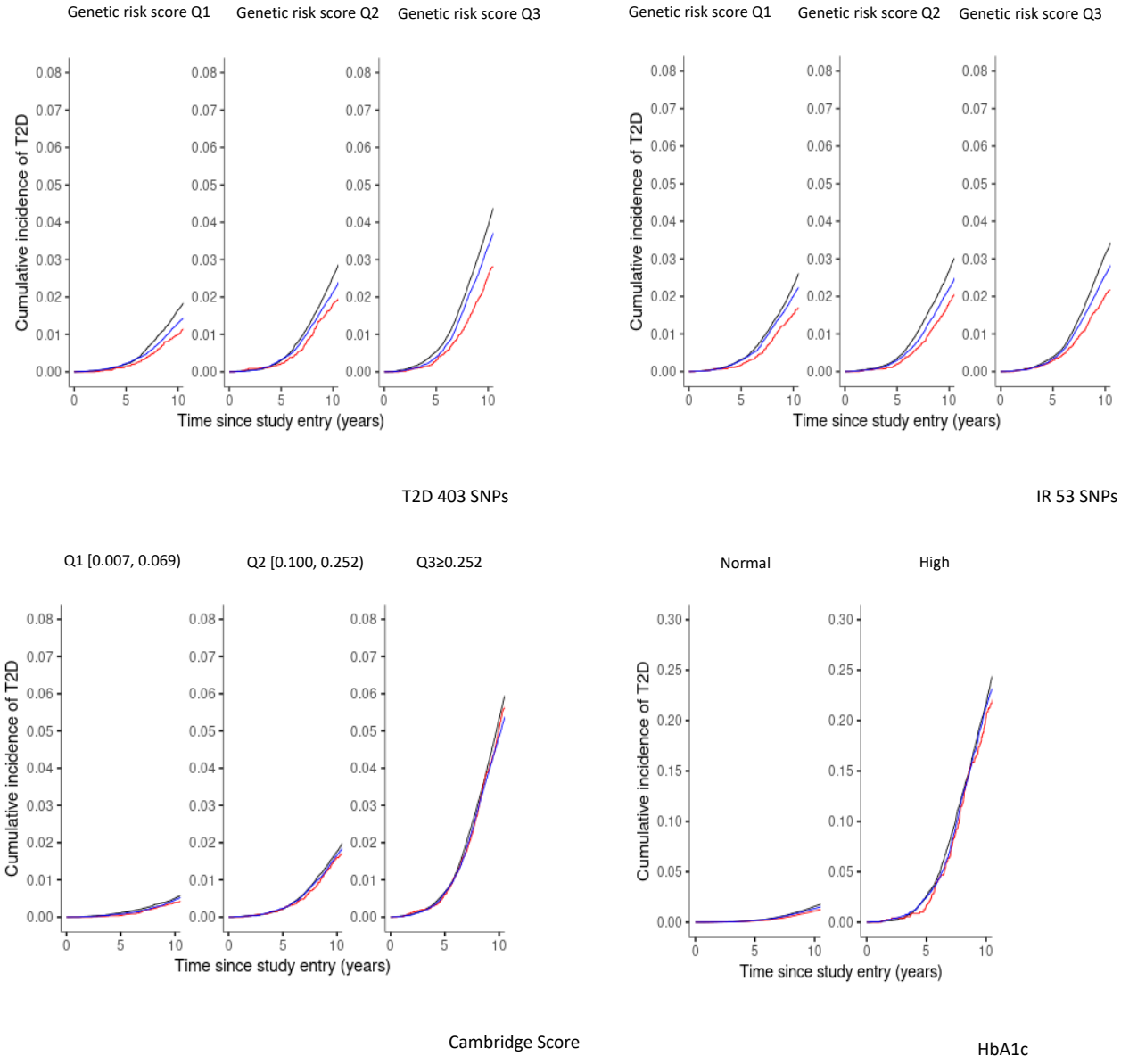
Supplementary Figure 4.1 Plots of Mendelian randomisation (MR) sensitivity analyses of deoxycarnitine on T2D. A, forest plot of individual single nucleotide polymorphism (SNP) ratio estimates ($\beta_{SNP\text{-}outcome}/\beta_{SNP\text{-}exposure}$); b, leave-one-out plot of inverse variance weighted estimates after removing each SNP; C, scatter plot presenting results using different MR multi-locus methods (see legend); D, funnel plot presenting MR causal estimates against their precision: each point represent a genetic variant; x-axis, ratio estimate in log scale; y-axis, 1/standard error of ratio estimate. Asymmetry in this plot may be indicative of directional pleiotropy.



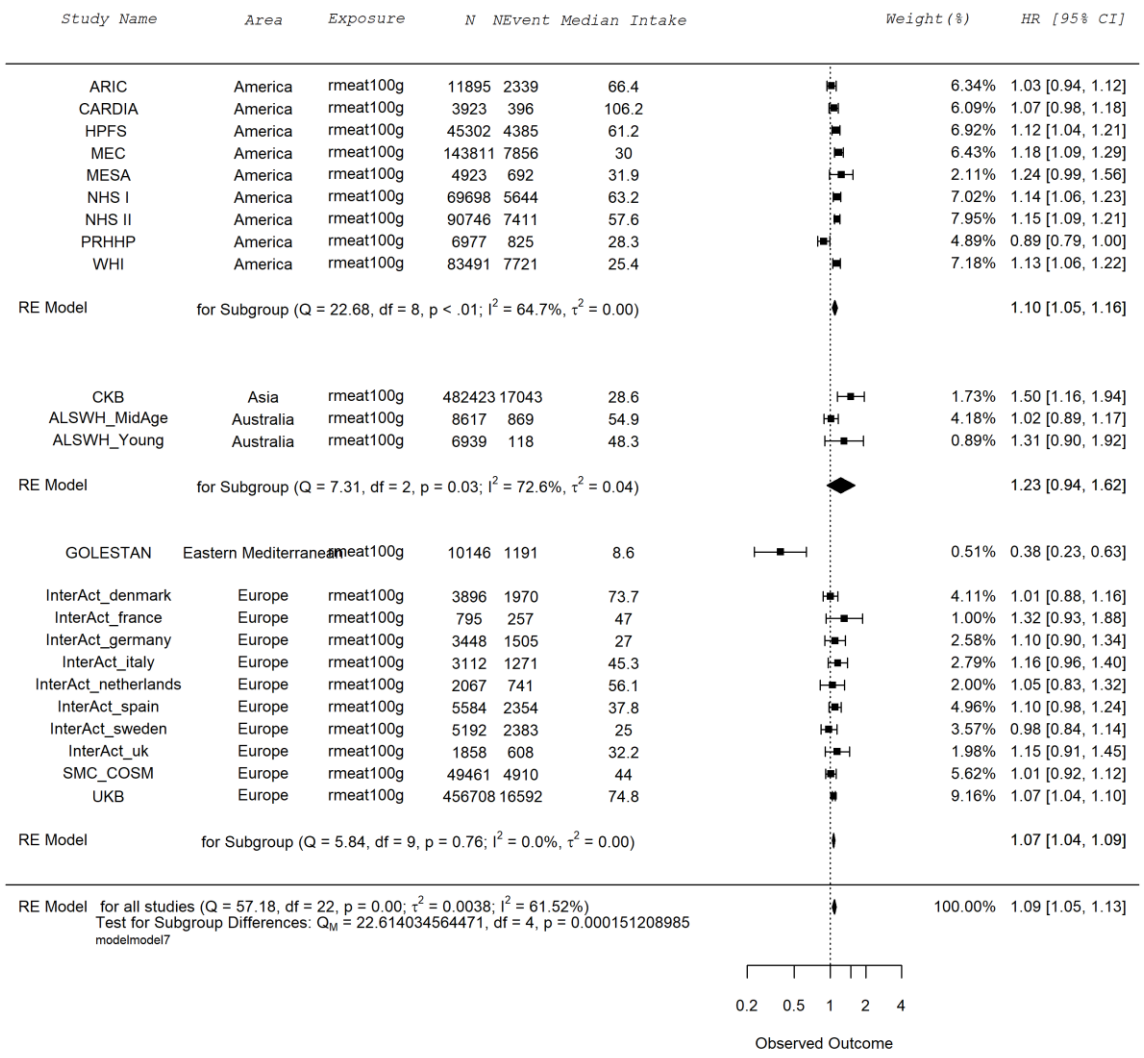
Supplementary Figure 4.2 Funnel plot of Mendelian randomisation (MR) sensitivity analysis of 1-(1-enyl-stearoyl)-2-arachidonoyl-GPE (P-18:0/20:4) on T2D. This plot presenting MR causal estimates against their precision: each point represent a genetic variant; x-axis, ratio estimate in log scale; y-axis, 1/standard error of ratio estimate. Asymmetry in this plot may be indicative of directional pleiotropy.



Supplementary Figure 5.2 Unadjusted cumulative incidence of T2D over 10 years in subgroups by meat intake and genetic or clinical risk factors in the EPIC-InterAct (n=20,628) study. Cambridge score, Cambridge diabetes risk score. The HbA1c groups were classified according to thresholds for diagnosing diabetes (<42 mmol/mol as the normal group, ≥42 and <48 mmol/mol as the high group, ≥48 mmol/mol were not included for analysis).

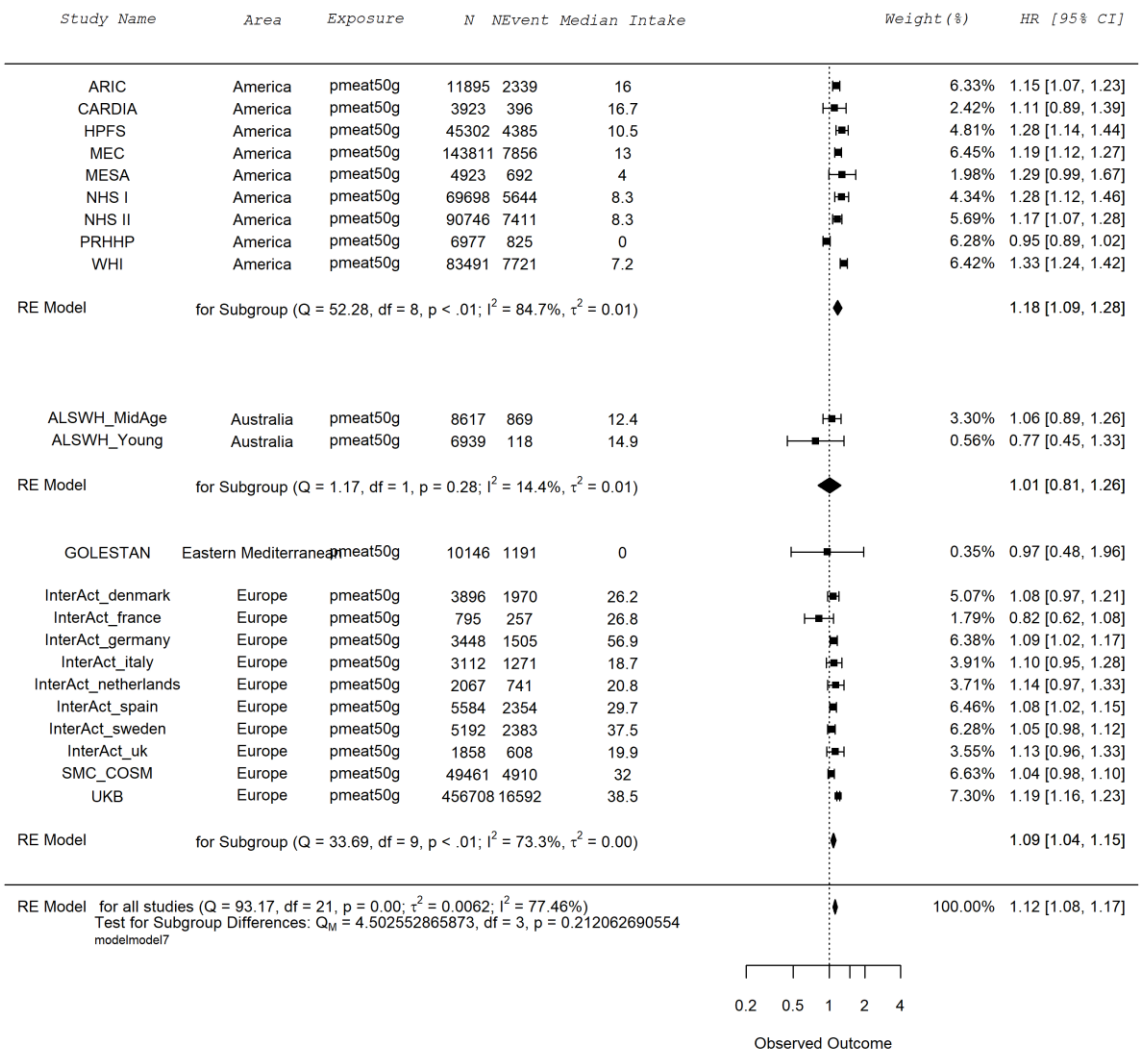


Supplementary Figure 5.3 Unadjusted cumulative incidence of T2D over 10 years in subgroups by meat intake and genetic or clinical risk factors in the UK Biobank (n=316,222) study. Cambridge score, Cambridge diabetes risk score. The HbA1c groups were classified according to thresholds for diagnosing diabetes (<42 mmol/mol as the normal group, ≥42 and <48 mmol/mol as the high group, ≥48 mmol/mol were not included for analysis).



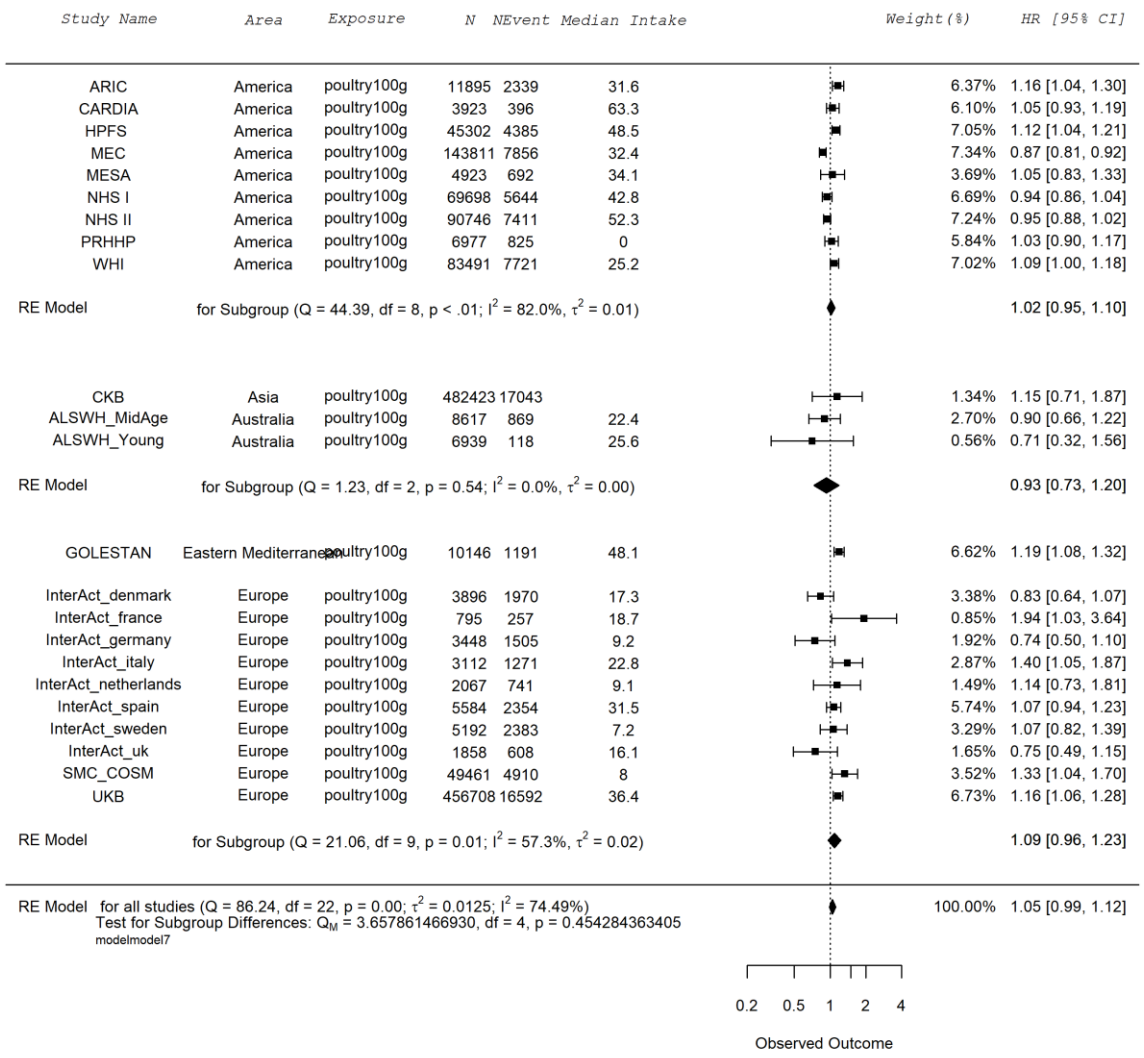
Supplementary Figure 6.1 Hazard ratios and 95% confidence intervals (CIs) for association between red meat consumption (per 100 g/d) and incident type 2 diabetes (secondary outcome) in the InterConnect project. Combined n=1,501,177; total incident type 2 diabetes cases=72,489. Associations are adjusted for age, sex, education, smoking, physical activity, alcohol intake, total energy intake, BMI, and other food intakes.

ALSWH, Australian Longitudinal Study on Women's Health; ARIC, Atherosclerosis Young Adults Study; CARDIA, Coronary Artery Risk Development in Young Adults; CKB, the China Kadoorie Biobank; COSM, the Cohort of Swedish Men; EPIC-InterAct, the European Prospective Investigation into Cancer-InterAct; HPFS, Health Professionals Follow-up Study; MEC, Multiethnic Cohort Study; MESA, the Multi-Ethnic Study of Atherosclerosis; NHS, Nurses' Health Study; PRHHP, the Puerto Rico Heart Health Program; SMC, the Swedish Mammography Cohort; WHI, Women's Health Initiative Study; UKB, UK Biobank.



Supplementary Figure 6.2 Hazard ratios and 95% confidence intervals (CIs) for association between processed meat consumption (per 50 g/d) and incident type 2 diabetes (secondary outcome) in the InterConnect project. Combined n=1,018,754; total incident type 2 diabetes cases=55,446. Associations are adjusted for age, sex, education, smoking, physical activity, alcohol intake, total energy intake, BMI, and other food intakes.

ALSWH, Australian Longitudinal Study on Women's Health; ARIC, Atherosclerosis Young Adults Study; CARDIA, Coronary Artery Risk Development in Young Adults; CKB, the China Kadoorie Biobank; COSM, the Cohort of Swedish Men; EPIC-InterAct, the European Prospective Investigation into Cancer-InterAct; HPFS, Health Professionals Follow-up Study; MEC, Multiethnic Cohort Study; MESA, the Multi-Ethnic Study of Atherosclerosis; NHS, Nurses' Health Study; PRHHP, the Puerto Rico Heart Health Program; SMC, the Swedish Mammography Cohort; WHI, Women's Health Initiative Study; UKB, UK Biobank.



Supplementary Figure 6.3 Hazard ratios and 95% confidence intervals (CIs) for association between poultry consumption (per 100 g/d) and incident type 2 diabetes (secondary outcome) in the InterConnect project. Combined n=1,501,177; total incident type 2 diabetes cases=72,489. Associations are adjusted for age, sex, education, smoking, physical activity, alcohol intake, total energy intake, BMI, and other food intakes.

ALSWH, Australian Longitudinal Study on Women's Health; ARIC, Atherosclerosis Young Adults Study; CARDIA, Coronary Artery Risk Development in Young Adults; CKB, the China Kadoorie Biobank; COSM, the Cohort of Swedish Men; EPIC-InterAct, the European Prospective Investigation into Cancer-InterAct; HPFS, Health Professionals Follow-up Study; MEC, Multiethnic Cohort Study; MESA, the Multi-Ethnic Study of Atherosclerosis; NHS, Nurses' Health Study; PRHHP, the Puerto Rico Heart Health Program; SMC, the Swedish Mammography Cohort; WHI, Women's Health Initiative Study; UKB, UK Biobank.

Supplementary Tables

Supplementry table 3.1 The definition of non-communicable diseases outcomes in the exploratory analyses for the association between red meat metabolite score and health outcomes

Disease		Definition	notes	prevalent cases for exclusion
Incident cardiovascular disease	Incident coronary heart disease	ICD-9 codes: 410-414; ICD-10 codes: I20-I25	Incident cases were defined either by hospital admissions data or death certificate.	Prevalent coronary heart disease was defined by a self-reported history of either angina or myocardial infarction.
	Incident cerebral stroke	ICD-9 codes: 433-435; ICD-10 codes: I63, I65, I66	Incident cases were defined either by hospital admissions data or death certificate.	Prevalent stroke was defined based on a self-reported history of stroke (any kind) by a doctor.
	Incident haemorrhagic stroke	ICD-9 codes: 430-432; ICD-10 codes: I60-I62	Incident cases were defined either by hospital admissions data or death certificate.	Prevalent stroke was defined based on a self-reported history of stroke (any kind) by a doctor.
	Incident atrial fibrillation	ICD-9 codes: 427.3; ICD-10 codes: I48	Incident cases were defined either by hospital admissions data or death certificate.	Prevalent atrial fibrillation (AF) was defined by self-reported intake of drugs that were used for treatment of AF in clinical practice at the time of the baseline survey (digitalis or vitamin Kantagonists; PMID 25059930).

Following the last page

Disease		Definition	notes	prevalent cases for exclusion
Incident cardiovascular disease	Incident heart failure	ICD-9 codes: 428; ICD-10 codes: I50	Incident cases were defined either by hospital admissions data or death certificate.	We defined prevalent heart failure by self-reported intake of drugs that were recommended for treatment of heart failure, namely loop diuretics in combination with digitalis or angiotensin-converting enzyme inhibitors (PMID 21835284).
Incident liver disease	Incident liver disease	ICD-10 codes: B15-19, C22, E83, E88, I85, K70, K72-76, R18, Z94	Incident cases were defined either by hospital admissions data or death certificate.	Prevalent liver disease was defined based on self-reported diagnosis of any liver disease by a doctor.
Incident renal disease	Incident renal disease	ICD-9 codes: 580-589, 593; ICD-10 codes: N00-N19, N25-N29	Incident cases were defined either by hospital admissions data or death certificate.	Prevalent kidney disease was defined as an eGFR<50 ml/min/1.73m^2.
Incident gastrointestinal cancer	Incident colon cancer	ICD-9 codes: 153.0-153.9; ICD-10 codes: C18	Incident cases were defined either by hospital admissions data or death certificate.	Prevalent cases were defined based on a self-reported history of any cancer.
	Incident rectal cancer	ICD-9 codes: 154.0-154.1, 159.0; ICD-10 codes: C19-C20	Incident cases were defined either by hospital admissions data or death certificate.	Prevalent cases were defined based on a self-reported history of any cancer.
	Incident stomach cancer	ICD-9 codes: 151; ICD-10 codes: C16	Incident cases were defined either by hospital admissions data or death certificate.	Prevalent cases were defined based on a self-reported history of any cancer.

Following the last page

Disease		**Definition**	**notes**	**prevalent cases for exclusion**
Incident fractures	Incident fractures	ICD-10 codes: S02, S12, S22, S32, S42, S52, S62, S72, S82, S92, S120-S122, S127-S129, S220-S225, S228, S229, S320-SS325, S327, S328, S520-S529, S620-S627, S720-S729, S820-S829, S920-S929, T02, T08, T10	Incident cases were defined based on hospital admission data.	Prevalent cases were reported based on any reported fracture at baseline examinations.
All-cause mortality	All-cause mortality		Mortality from all causes was defined from death certificates.	

Supplementary table 5.1 SNPs that made up for GRSs and components for Cambridge diabetes risk score

Trait	Chr	Position	SNP	Effect Allele	Other Allele	Weight
T2D	1	40035928	rs3768321	T	G	0.086
T2D	1	51256091	rs58432198	C	T	0.068
T2D	1	62579891	rs12140153	G	T	0.068
T2D	1	115144899	rs184660829	C	T	2.086
T2D	1	117532790	rs1127215	C	T	0.049
T2D	1	120526982	rs1493694	T	C	0.086
T2D	1	150786038	rs10305745	A	G	0.247
T2D	1	151017991	rs145904381	T	C	0.174
T2D	1	177889025	rs539515	C	A	0.049
T2D	1	205114873	rs12048743	G	C	0.039
T2D	1	206593900	rs9430095	C	G	0.039
T2D	1	214150821	rs79687284	C	G	0.148
T2D	1	214159256	rs340874	C	T	0.068
T2D	1	214175531	rs114526150	G	T	0.113
T2D	1	219584164	rs553014999	C	T	0.642
T2D	1	219748818	rs2820446	C	G	0.058
T2D	1	229672955	rs348330	G	A	0.049
T2D	1	235690800	rs291367	G	A	0.039
T2D	10	12307894	rs11257655	T	C	0.086
T2D	10	71321279	rs177045	G	A	0.068
T2D	10	71321658	rs61850200	C	G	0.039
T2D	10	71332301	rs41277236	T	C	0.086
T2D	10	71347311	rs549498088	T	C	0.445
T2D	10	71466578	rs2642588	G	T	0.049
T2D	10	80952826	rs703972	G	C	0.068
T2D	10	81096589	rs1317617	G	A	0.039
T2D	10	89769340	rs11202627	T	C	0.058
T2D	10	93924663	rs7078559	T	C	0.03
T2D	10	94462427	rs10882101	T	C	0.058
T2D	10	94479107	rs1112718	A	G	0.058
T2D	10	114699835	rs536643418	G	C	0.405
T2D	10	114702962	rs140242150	A	G	0.307
T2D	10	114703136	rs7918400	T	C	0.058
T2D	10	114740337	rs184509201	C	G	0.191
T2D	10	114751173	rs180988137	G	A	0.157
T2D	10	114757956	rs78025551	C	G	0.049
T2D	10	114758349	rs7903146	T	C	0.315
T2D	10	114871594	rs34855922	A	G	0.049
T2D	10	122915345	rs72631105	A	G	0.058
T2D	10	124193181	rs2280141	T	G	0.049
T2D	11	1704596	rs12802972	A	G	0.03

T2D	11	2118860	rs11042596	G	T	0.039
T2D	11	2151761	rs555759341	C	G	0.322
T2D	11	2182519	rs571342427	C	T	0.519
T2D	11	2197286	rs4929965	A	G	0.068
T2D	11	2372356	rs4930091	C	T	0.039
T2D	11	2579163	rs2283164	A	G	0.077
T2D	11	2634177	rs80102379	G	T	0.14
T2D	11	2672821	rs231349	T	C	0.068
T2D	11	2691500	rs231361	A	G	0.077
T2D	11	2755548	rs2283220	A	G	0.049
T2D	11	2850828	rs234853	G	A	0.077
T2D	11	2857194	rs2237895	C	A	0.113
T2D	11	2858546	rs2237897	C	T	0.207
T2D	11	2908754	rs445084	G	A	0.03
T2D	11	14763828	rs141521721	A	C	0.122
T2D	11	17408404	rs5213	C	T	0.068
T2D	11	17470143	rs67254669	G	A	0.637
T2D	11	28534898	rs4923543	A	G	0.039
T2D	11	32460873	rs7943101	T	C	0.039
T2D	11	32927778	rs145678014	G	T	0.104
T2D	11	33091735	rs528122639	A	G	0.737
T2D	11	34642668	rs286925	A	G	0.039
T2D	11	34982148	rs2767036	C	A	0.039
T2D	11	43877934	rs1061810	A	C	0.049
T2D	11	45912013	rs7115753	A	G	0.039
T2D	11	47529947	rs7124681	A	C	0.039
T2D	11	65294799	rs1783541	T	C	0.058
T2D	11	68997225	rs61881115	G	A	0.049
T2D	11	69448758	rs11820019	T	C	0.148
T2D	11	72460398	rs77464186	A	C	0.104
T2D	11	92708710	rs10830963	G	C	0.095
T2D	11	93013531	rs57235767	C	T	0.039
T2D	11	128042575	rs10893829	T	C	0.058
T2D	11	128234144	rs10750397	A	G	0.049
T2D	11	128398938	rs67232546	T	C	0.058
T2D	11	128583975	rs112595469	T	C	0.095
T2D	12	4031104	rs10848958	C	T	0.039
T2D	12	4300172	rs11063028	C	T	0.058
T2D	12	4376089	rs4238013	C	T	0.058
T2D	12	4384696	rs3217792	C	T	0.113
T2D	12	4384844	rs76895963	T	G	0.482
T2D	12	4399050	rs3217860	G	A	0.049
T2D	12	12871099	rs2066827	G	T	0.049
T2D	12	26453283	rs718314	G	A	0.049
T2D	12	27965150	rs10842994	C	T	0.077
T2D	12	66221060	rs2258238	T	A	0.095

T2D	12	66358347	rs1042725	T	C	0.049
T2D	12	71522953	rs1796330	G	C	0.049
T2D	12	95928560	rs2197973	T	C	0.039
T2D	12	97562756	rs759111467	A	G	1.122
T2D	12	97779248	rs557027608	A	G	0.85
T2D	12	97848775	rs77864822	A	G	0.077
T2D	12	108629780	rs1426371	G	A	0.049
T2D	12	118412373	rs34965774	A	G	0.058
T2D	12	118489636	rs12578639	A	T	0.039
T2D	12	121297815	rs11065299	A	G	0.058
T2D	12	121380541	rs73226260	G	A	0.122
T2D	12	121416864	rs1800574	T	C	0.131
T2D	12	121432117	rs56348580	G	C	0.049
T2D	12	121501461	rs28638142	A	C	0.077
T2D	12	121882395	rs73224262	T	C	0.215
T2D	12	123450765	rs4148856	C	G	0.049
T2D	12	124468572	rs7978610	G	C	0.239
T2D	12	124509177	rs825452	A	G	0.039
T2D	12	133069698	rs12811407	A	G	0.049
T2D	13	26776999	rs34584161	A	G	0.049
T2D	13	31042452	rs11842871	G	T	0.039
T2D	13	33554302	rs576674	G	A	0.049
T2D	13	51096095	rs963740	A	T	0.039
T2D	13	58366634	rs9537803	C	T	0.039
T2D	13	58965435	rs9569864	C	T	0.049
T2D	13	59077406	rs9563615	A	T	0.049
T2D	13	59184234	rs76251711	G	A	0.148
T2D	13	80717156	rs1359790	G	A	0.086
T2D	13	109947213	rs7987740	T	C	0.039
T2D	13	110431626	rs4771648	G	A	0.039
T2D	14	23288935	rs17122772	G	C	0.039
T2D	14	33302882	rs17522122	T	G	0.039
T2D	14	38848419	rs8017808	G	T	0.039
T2D	14	79932041	rs17836088	C	G	0.058
T2D	14	91963722	rs8010382	G	A	0.039
T2D	14	103894071	rs62007683	G	T	0.039
T2D	15	38834033	rs8032939	C	T	0.058
T2D	15	38873115	rs34715063	C	T	0.095
T2D	15	41809205	rs11070332	A	G	0.049
T2D	15	42201410	rs543786825	T	C	1.147
T2D	15	53091553	rs2456530	T	C	0.058
T2D	15	53747228	rs528350911	G	C	0.239
T2D	15	57456802	rs117483894	G	A	0.095
T2D	15	62394264	rs8037894	G	C	0.049
T2D	15	63871292	rs7178762	C	T	0.039
T2D	15	68080886	rs4776970	A	T	0.039

T2D	15	75932129	rs13737	G	T	0.049
T2D	15	77818128	rs1005752	A	C	0.077
T2D	15	90423293	rs4932265	T	C	0.068
T2D	15	91511260	rs12910825	G	A	0.049
T2D	16	295795	rs6600191	T	C	0.058
T2D	16	3583173	rs3751837	T	C	0.039
T2D	16	28915217	rs8046545	G	A	0.039
T2D	16	30045789	rs11642430	G	C	0.039
T2D	16	30419384	rs199795270	C	G	0.223
T2D	16	53501946	rs4281707	G	A	0.039
T2D	16	53758720	rs78020297	A	G	0.086
T2D	16	53800954	rs1421085	C	T	0.122
T2D	16	69651866	rs862320	C	T	0.039
T2D	16	75234872	rs72802342	C	A	0.157
T2D	16	75516534	rs3115960	G	C	0.03
T2D	16	81534790	rs2925979	T	C	0.049
T2D	16	89564055	rs12920022	A	T	0.049
T2D	17	3828086	rs1043246	G	C	0.049
T2D	17	3860356	rs3826482	A	T	0.03
T2D	17	4045440	rs1377807	C	G	0.049
T2D	17	7549681	rs1641523	C	T	0.049
T2D	17	7740170	rs62059712	T	C	0.068
T2D	17	9785187	rs7222481	C	G	0.039
T2D	17	17661802	rs4925109	A	G	0.049
T2D	17	29413019	rs71372253	C	T	0.077
T2D	17	36046451	rs10962	C	G	0.049
T2D	17	36063685	rs2189301	G	A	0.049
T2D	17	36099952	rs10908278	T	A	0.077
T2D	17	40731411	rs34855406	C	G	0.049
T2D	17	47060322	rs35895680	C	A	0.058
T2D	17	52140805	rs569511541	G	A	2.032
T2D	17	61965043	rs2727301	T	C	0.039
T2D	17	62203304	rs60276348	T	C	0.049
T2D	17	65648427	rs11657492	G	T	0.058
T2D	17	65820153	rs558308082	C	G	0.713
T2D	17	65892507	rs61676547	C	G	0.058
T2D	18	7070642	rs7240767	C	T	0.039
T2D	18	36278709	rs62080313	C	T	0.058
T2D	18	52604955	rs76197067	G	A	0.986
T2D	18	53050646	rs72926932	C	A	0.086
T2D	18	53452144	rs28719468	C	T	0.039
T2D	18	54675384	rs17684074	G	C	0.039
T2D	18	56876228	rs9957145	G	A	0.049
T2D	18	57848369	rs523288	T	A	0.049
T2D	18	58056566	rs74452128	C	A	0.14
T2D	18	60668270	rs10469140	G	A	0.03

T2D	18	60845884	rs12454712	T	C	0.049
T2D	19	4948862	rs7249758	A	G	0.049
T2D	19	5224998	rs116953931	A	G	0.077
T2D	19	7240848	rs75253922	C	T	0.049
T2D	19	7970635	rs4804833	A	G	0.049
T2D	19	12938471	rs755734872	T	C	0.863
T2D	19	13038415	rs3111316	A	G	0.049
T2D	19	19388500	rs8107974	T	A	0.095
T2D	19	19396616	rs188247550	T	C	0.14
T2D	19	33890838	rs10406327	C	G	0.039
T2D	19	44938870	rs745903616	A	G	0.476
T2D	19	45411941	rs429358	T	C	0.077
T2D	19	46157019	rs10406431	A	G	0.049
T2D	19	46178661	rs2238689	C	T	0.039
T2D	19	46351837	rs533172266	T	C	0.846
T2D	19	47569003	rs3810291	A	G	0.049
T2D	2	422144	rs62107261	T	C	0.113
T2D	2	653575	rs35913461	C	T	0.058
T2D	2	16574669	rs11680058	A	G	0.058
T2D	2	25643221	rs17802463	G	T	0.039
T2D	2	27730940	rs1260326	C	T	0.068
T2D	2	43207872	rs28525376	G	T	0.03
T2D	2	43430440	rs6708643	A	G	0.039
T2D	2	43698028	rs80147536	A	T	0.122
T2D	2	58981064	rs10193538	T	G	0.039
T2D	2	59307725	rs6545714	G	A	0.039
T2D	2	60583665	rs243024	A	G	0.058
T2D	2	65287896	rs2249105	A	G	0.095
T2D	2	65355270	rs2052261	G	A	0.068
T2D	2	65655012	rs2028150	C	G	0.049
T2D	2	96913918	rs79046683	T	G	0.85
T2D	2	118071061	rs562386202	G	A	1.163
T2D	2	121318166	rs11688931	C	G	0.039
T2D	2	121347612	rs11688682	G	C	0.049
T2D	2	121378852	rs66477705	T	C	0.086
T2D	2	147861633	rs35999103	T	C	0.049
T2D	2	158339550	rs13426680	A	G	0.086
T2D	2	161135544	rs3772071	T	C	0.049
T2D	2	165513091	rs10195252	T	C	0.068
T2D	2	165573194	rs13024606	T	C	0.086
T2D	2	219859171	rs113414093	A	G	0.113
T2D	2	227101411	rs2972144	G	A	0.095
T2D	20	21466795	rs13041756	C	T	0.058
T2D	20	32596704	rs2268078	A	G	0.039
T2D	20	42905415	rs76811102	T	C	0.086
T2D	20	43001721	rs4810426	T	C	0.086

T2D	20	43023355	rs191830490	G	A	0.215
T2D	20	43042364	rs1800961	T	C	0.166
T2D	20	43233649	rs11696357	A	G	0.058
T2D	20	45317678	rs560716466	A	G	0.307
T2D	20	45598564	rs6063048	G	A	0.049
T2D	20	48832135	rs11699802	C	T	0.039
T2D	20	51223594	rs34454109	A	T	0.039
T2D	20	57394628	rs6070625	G	C	0.049
T2D	20	57551099	rs862016	G	A	0.068
T2D	20	62450664	rs6011155	T	C	0.039
T2D	20	62693175	rs59944054	A	G	0.058
T2D	22	30609554	rs6518681	G	A	0.086
T2D	22	32348841	rs117001013	C	T	0.068
T2D	22	41489920	rs5758223	A	G	0.039
T2D	22	44324730	rs738408	T	C	0.049
T2D	22	50356850	rs1801645	C	T	0.039
T2D	22	50604696	rs112915006	G	A	0.077
T2D	3	12336507	rs11709077	G	A	0.131
T2D	3	12489342	rs17819328	G	T	0.058
T2D	3	23455582	rs35352848	T	C	0.068
T2D	3	23510044	rs17013314	G	A	0.104
T2D	3	46925539	rs11926707	C	T	0.239
T2D	3	47242923	rs75423501	G	A	0.049
T2D	3	49980596	rs4688760	T	C	0.039
T2D	3	53127677	rs2581787	T	G	0.039
T2D	3	54828827	rs76263492	T	G	0.086
T2D	3	63962339	rs3774723	G	A	0.068
T2D	3	64460694	rs74368513	G	A	0.27
T2D	3	64701146	rs9860730	A	G	0.058
T2D	3	72865183	rs13085136	C	T	0.077
T2D	3	77671721	rs2272163	C	A	0.039
T2D	3	123065778	rs11708067	A	G	0.086
T2D	3	124926637	rs649961	T	C	0.039
T2D	3	129333182	rs9828772	C	G	0.058
T2D	3	129470067	rs559138871	T	C	0.399
T2D	3	150066540	rs62271373	A	T	0.086
T2D	3	152086533	rs13065698	A	G	0.049
T2D	3	152417881	rs74653713	C	A	0.095
T2D	3	152433628	rs35497231	C	T	0.039
T2D	3	168218841	rs7629630	A	T	0.049
T2D	3	170733076	rs9873618	G	A	0.068
T2D	3	183738460	rs2872246	A	C	0.039
T2D	3	185503456	rs6780171	A	T	0.131
T2D	3	185514421	rs150111048	G	A	0.113
T2D	3	185541213	rs11717959	G	T	0.039
T2D	3	185829891	rs1516728	A	T	0.03

T2D	3	186665645	rs3887925	T	C	0.068
T2D	3	186675277	rs7645517	A	G	0.077
T2D	3	187740899	rs4686471	C	T	0.058
T2D	4	616608	rs111827885	C	T	0.166
T2D	4	744972	rs1531583	T	G	0.122
T2D	4	1010077	rs35654957	C	T	0.03
T2D	4	1784403	rs56337234	C	T	0.058
T2D	4	3241845	rs362307	T	C	0.077
T2D	4	6302519	rs1801212	A	G	0.049
T2D	4	6306763	rs10937721	C	G	0.058
T2D	4	17792869	rs12640250	C	A	0.039
T2D	4	45186139	rs10938398	A	G	0.049
T2D	4	52818664	rs2102278	G	A	0.039
T2D	4	53207093	rs114447556	T	C	0.058
T2D	4	83578271	rs12642790	A	G	0.039
T2D	4	89740894	rs1903002	G	C	0.039
T2D	4	89857291	rs576406049	T	C	0.501
T2D	4	95091911	rs6821438	A	G	0.039
T2D	4	104140848	rs1580278	C	A	0.039
T2D	4	137083193	rs1296328	A	C	0.039
T2D	4	153513369	rs7669833	T	A	0.058
T2D	4	157652753	rs28819812	C	A	0.039
T2D	4	185717759	rs58730668	T	C	0.068
T2D	5	14610134	rs3845281	G	A	0.077
T2D	5	14751305	rs146886108	C	T	0.344
T2D	5	14753745	rs17250977	G	A	0.113
T2D	5	14768092	rs6885132	C	G	0.068
T2D	5	14768766	rs76549217	T	C	0.131
T2D	5	44534364	rs62368490	T	C	0.095
T2D	5	44682589	rs6884702	G	A	0.039
T2D	5	51791225	rs17261179	T	C	0.039
T2D	5	52100489	rs3811978	G	A	0.058
T2D	5	52315682	rs62357230	A	G	0.086
T2D	5	52774510	rs62370480	A	G	0.039
T2D	5	53271420	rs702634	A	G	0.049
T2D	5	53412620	rs279744	C	A	0.039
T2D	5	55808475	rs465002	T	C	0.104
T2D	5	55848669	rs2431115	A	G	0.039
T2D	5	55861595	rs9687832	A	G	0.077
T2D	5	56196604	rs96844	G	A	0.039
T2D	5	67714246	rs4976033	G	A	0.049
T2D	5	75003678	rs2307111	T	C	0.049
T2D	5	76424949	rs4457053	G	A	0.058
T2D	5	78430607	rs1316776	C	A	0.049
T2D	5	86577352	rs7719891	G	A	0.039
T2D	5	101232944	rs138337556	G	A	0.445

T2D	5	102338739	rs78408340	G	C	0.385
T2D	5	102422968	rs115505614	T	C	0.174
T2D	5	133414622	rs244665	A	G	0.03
T2D	5	133864599	rs329122	A	G	0.039
T2D	5	157928196	rs3934712	C	T	0.049
T2D	6	7035734	rs112498319	C	A	0.03
T2D	6	7231843	rs9379084	G	A	0.104
T2D	6	7255650	rs9505097	C	T	0.049
T2D	6	20679709	rs7756992	G	A	0.14
T2D	6	32573415	rs601945	G	A	0.058
T2D	6	34247047	rs77136196	T	C	0.104
T2D	6	34524698	rs2233632	T	C	0.039
T2D	6	40409243	rs34298980	T	C	0.039
T2D	6	43760327	rs11967262	G	C	0.039
T2D	6	43814190	rs6458354	C	T	0.049
T2D	6	50788778	rs3798519	C	A	0.058
T2D	6	51180765	rs2465043	G	A	0.03
T2D	6	67387490	rs555402748	T	C	1.3
T2D	6	107431688	rs4946812	G	A	0.039
T2D	6	126792095	rs11759026	G	A	0.068
T2D	6	127416930	rs2800733	A	G	0.049
T2D	6	137300960	rs9494624	A	G	0.039
T2D	6	139835329	rs2982521	A	T	0.049
T2D	6	140249466	rs616279	A	G	0.039
T2D	6	160770312	rs474513	A	G	0.039
T2D	6	164133001	rs4709746	C	T	0.058
T2D	7	14898282	rs17168486	T	C	0.068
T2D	7	15063569	rs10228066	T	C	0.068
T2D	7	15206239	rs2908334	T	C	0.03
T2D	7	23434606	rs78840640	G	C	0.104
T2D	7	23512896	rs4279506	G	C	0.058
T2D	7	28198677	rs1708302	C	T	0.095
T2D	7	30728452	rs917195	C	T	0.049
T2D	7	44255643	rs878521	A	G	0.058
T2D	7	44365549	rs116913033	C	T	0.039
T2D	7	102038318	rs56376556	T	C	0.077
T2D	7	102486254	rs11496066	T	C	0.077
T2D	7	102987583	rs62482405	G	T	0.049
T2D	7	103444978	rs39328	T	C	0.039
T2D	7	117495667	rs6976111	A	C	0.039
T2D	7	130027037	rs2268382	C	A	0.03
T2D	7	130457914	rs1562396	G	A	0.058
T2D	7	150537635	rs62492368	A	G	0.049
T2D	7	156930550	rs6459733	G	C	0.058
T2D	8	9974824	rs17689007	G	A	0.039
T2D	8	10808687	rs57327348	A	T	0.039

T2D	8	19830921	rs10096633	C	T	0.068
T2D	8	30863938	rs10954772	T	C	0.039
T2D	8	41508577	rs13262861	C	A	0.068
T2D	8	41509915	rs4736819	T	C	0.039
T2D	8	41552046	rs148766658	C	T	0.086
T2D	8	95685147	rs11786992	A	C	0.03
T2D	8	95961626	rs10097617	T	C	0.039
T2D	8	96092422	rs187936726	G	A	0.104
T2D	8	97737741	rs149364428	A	G	0.239
T2D	8	110123183	rs12680028	C	G	0.039
T2D	8	118185025	rs3802177	G	A	0.104
T2D	8	118404672	rs80244329	G	A	0.104
T2D	8	128711742	rs17772814	G	A	0.077
T2D	8	129568078	rs1561927	C	T	0.039
T2D	8	145507304	rs4977213	C	T	0.049
T2D	8	145879883	rs12719778	T	C	0.039
T2D	9	3965689	rs510807	A	C	0.03
T2D	9	4243045	rs79103584	T	A	0.131
T2D	9	4291928	rs10974438	C	A	0.049
T2D	9	19067833	rs7022807	G	A	0.039
T2D	9	20241069	rs7867635	C	T	0.039
T2D	9	20662703	rs7847880	C	T	0.039
T2D	9	22043612	rs1412830	C	T	0.039
T2D	9	22133773	rs76011118	A	G	0.104
T2D	9	22134068	rs10811660	G	A	0.239
T2D	9	22134172	rs10757283	T	C	0.104
T2D	9	22157908	rs1333052	A	C	0.03
T2D	9	22301092	rs1575972	T	A	0.095
T2D	9	28410683	rs1412234	C	T	0.039
T2D	9	34074476	rs12001437	C	T	0.039
T2D	9	81359113	rs11137820	C	G	0.039
T2D	9	81905590	rs17791513	A	G	0.095
T2D	9	84308948	rs2796441	G	A	0.068
T2D	9	97001682	rs55653563	A	C	0.039
T2D	9	97497494	rs12236906	T	C	0.14
T2D	9	136149229	rs505922	C	T	0.049
T2D	9	139235606	rs78403475	G	C	0.058
T2D	9	139241030	rs28505901	G	A	0.086
T2D	9	139507212	rs11793035	C	T	0.039
T2D	9	139737088	9:139737088_G_A	A	G	1.008
Insulin resistance	1	39895460	rs683135	A	G	0.008
Insulin resistance	1	50815783	rs17386142	C	T	0.014
Insulin resistance	1	110500175	rs11577194	T	C	0.006
Insulin resistance	1	172312769	rs9425291	A	G	0.009
Insulin resistance	1	219722104	rs4846565	G	A	0.013
Insulin resistance	2	65287896	rs2249105	A	G	0.009

Insulin resistance	2	165513091	rs10195252	T	C	0.017
Insulin resistance	2	219349752	rs492400	T	C	0.006
Insulin resistance	2	227099180	rs2943645	T	C	0.019
Insulin resistance	3	12116620	rs308971	G	A	0.021
Insulin resistance	3	15185634	rs3864041	T	C	0.006
Insulin resistance	3	47375955	rs295449	A	G	0.007
Insulin resistance	3	52896855	rs11130329	A	C	0.012
Insulin resistance	3	123082416	rs9881942	A	G	0.008
Insulin resistance	3	135926622	rs645040	T	G	0.008
Insulin resistance	4	3480136	rs2699429	C	T	0.006
Insulin resistance	4	89741269	rs3822072	A	G	0.012
Insulin resistance	4	157734675	rs6822892	A	G	0.014
Insulin resistance	5	53272664	rs4865796	A	G	0.015
Insulin resistance	5	55806751	rs459193	G	A	0.015
Insulin resistance	5	67714246	rs4976033	G	A	0.009
Insulin resistance	5	112711486	rs6887914	C	T	0.008
Insulin resistance	5	118729286	rs1045241	C	T	0.007
Insulin resistance	5	158022041	rs2434612	G	A	0.009
Insulin resistance	5	173350405	rs966544	G	A	0.007
Insulin resistance	6	35004819	rs12525532	T	C	0.011
Insulin resistance	6	43815364	rs6937438	A	G	0.007
Insulin resistance	6	127452935	rs2745353	T	C	0.011
Insulin resistance	6	130398731	rs9492443	C	T	0.008
Insulin resistance	6	139828916	rs3861397	G	A	0.008
Insulin resistance	7	15883727	rs17169104	G	C	0.012
Insulin resistance	7	130466854	rs972283	G	A	0.011
Insulin resistance	8	9185146	rs2126259	T	C	0.024
Insulin resistance	8	19830769	rs1011685	C	T	0.011
Insulin resistance	8	72469742	rs4738141	G	A	0.008
Insulin resistance	8	126528955	rs7005992	C	G	0.010
Insulin resistance	9	78034169	rs498313	A	G	0.007
Insulin resistance	10	64869239	rs10995441	G	T	0.008
Insulin resistance	11	63862612	rs11231693	A	G	0.021
Insulin resistance	12	14571671	rs17402950	G	A	0.014
Insulin resistance	12	26453283	rs718314	G	A	0.010
Insulin resistance	12	124449223	rs7973683	C	A	0.011
Insulin resistance	13	111628195	rs7323406	A	G	0.007
Insulin resistance	15	39464167	rs7176058	A	G	0.008
Insulin resistance	15	73081067	rs8032586	C	T	0.009
Insulin resistance	17	4657034	rs754814	T	C	0.007
Insulin resistance	18	47174679	rs7227237	C	T	0.009
Insulin resistance	19	7293119	rs8101064	T	C	0.026
Insulin resistance	19	7970635	rs4804833	A	G	0.010
Insulin resistance	19	8615589	rs4804311	A	G	0.011
Insulin resistance	19	33899065	rs731839	G	A	0.015
Insulin resistance	20	45602638	rs6066149	G	A	0.007

Insulin resistance	22	38563471	rs132985	C	T	0.009
BMI	1	47684677	rs977747	T	G	0.017
BMI	1	49589847	rs657452	A	G	0.023
BMI	1	50559820	rs11583200	C	T	0.018
BMI	1	72751185	rs3101336	C	T	0.033
BMI	1	75002193	rs12566985	G	A	0.024
BMI	1	78446761	rs12401738	A	G	0.021
BMI	1	96924097	rs11165643	T	C	0.022
BMI	1	110154688	rs17024393	C	T	0.066
BMI	1	177889480	rs543874	G	A	0.048
BMI	1	201784287	rs2820292	C	A	0.020
BMI	10	87410904	rs7899106	G	A	0.040
BMI	10	102395440	rs17094222	C	T	0.025
BMI	10	104869038	rs11191560	C	T	0.031
BMI	10	114758349	rs7903146	C	T	0.023
BMI	11	8673939	rs4256980	G	C	0.021
BMI	11	27684517	rs11030104	A	G	0.041
BMI	11	43864278	rs2176598	T	C	0.020
BMI	11	47650993	rs3817334	T	C	0.026
BMI	11	115022404	rs12286929	G	A	0.022
BMI	12	50247468	rs7138803	A	G	0.032
BMI	12	122781897	rs11057405	G	A	0.031
BMI	13	28017782	rs9581854	T	C	0.030
BMI	13	54102206	rs12429545	A	G	0.033
BMI	13	66205704	rs9540493	A	G	0.017
BMI	13	79580919	rs1441264	A	G	0.018
BMI	14	25928179	rs10132280	C	A	0.023
BMI	14	29736838	rs12885454	C	A	0.021
BMI	14	30515112	rs11847697	T	C	0.049
BMI	14	79899454	rs7141420	T	C	0.024
BMI	15	51748610	rs3736485	A	G	0.018
BMI	15	68077168	rs16951275	T	C	0.031
BMI	15	73093991	rs7164727	T	C	0.018
BMI	16	3627358	rs758747	T	C	0.023
BMI	16	19935389	rs12446632	G	A	0.040
BMI	16	28333411	rs2650492	A	G	0.021
BMI	16	28889486	rs3888190	A	C	0.031
BMI	16	30015337	rs4787491	G	A	0.016
BMI	16	31129895	rs9925964	A	G	0.019
BMI	16	49062590	rs2080454	C	A	0.017
BMI	16	53803574	rs1558902	A	T	0.082
BMI	17	2005136	rs9914578	G	C	0.020
BMI	17	5283252	rs1000940	G	A	0.019
BMI	17	78615571	rs12940622	G	A	0.018
BMI	18	21104888	rs1808579	C	T	0.017
BMI	18	40147671	rs7239883	G	A	0.016

BMI	18	56883319	rs7243357	T	G	0.022
BMI	18	57829135	rs6567160	C	T	0.056
BMI	19	18454825	rs17724992	A	G	0.019
BMI	19	34309532	rs29941	G	A	0.018
BMI	19	45395619	rs2075650	A	G	0.026
BMI	19	46202172	rs2287019	C	T	0.036
BMI	19	47569003	rs3810291	A	G	0.028
BMI	2	632348	rs13021737	G	A	0.060
BMI	2	25150296	rs10182181	G	A	0.031
BMI	2	26928811	rs11126666	A	G	0.021
BMI	2	59305625	rs1016287	T	C	0.023
BMI	2	63053048	rs11688816	G	A	0.017
BMI	2	143043285	rs2121279	T	C	0.025
BMI	2	164567689	rs1460676	C	T	0.020
BMI	2	181550962	rs1528435	T	C	0.018
BMI	2	208255518	rs17203016	G	A	0.021
BMI	2	213413231	rs7599312	G	A	0.022
BMI	2	219349752	rs492400	C	T	0.016
BMI	2	227092802	rs2176040	A	G	0.014
BMI	20	51087862	rs6091540	C	T	0.019
BMI	21	40291740	rs2836754	C	T	0.016
BMI	3	25106437	rs6804842	G	A	0.019
BMI	3	61236462	rs2365389	C	T	0.020
BMI	3	81792112	rs3849570	A	C	0.019
BMI	3	85807590	rs13078960	G	T	0.030
BMI	3	141275436	rs16851483	T	G	0.048
BMI	3	185824004	rs1516725	C	T	0.045
BMI	4	45182527	rs10938397	G	A	0.040
BMI	4	77129568	rs17001654	G	C	0.031
BMI	4	103188709	rs13107325	T	C	0.048
BMI	4	145659064	rs11727676	T	C	0.036
BMI	5	75015242	rs2112347	T	G	0.026
BMI	5	153537893	rs7715256	G	T	0.016
BMI	6	34563164	rs205262	G	A	0.022
BMI	6	40348653	rs2033529	G	A	0.019
BMI	6	50845490	rs2207139	G	A	0.045
BMI	6	108977663	rs9400239	C	T	0.019
BMI	6	120185665	rs9374842	T	C	0.019
BMI	6	137675541	rs13201877	G	A	0.023
BMI	6	163033350	rs13191362	A	G	0.028
BMI	7	75163169	rs1167827	G	A	0.020
BMI	7	76608143	rs2245368	C	T	0.032
BMI	7	93197732	rs9641123	C	G	0.019
BMI	7	95169514	rs6465468	T	G	0.017
BMI	8	76806584	rs17405819	T	C	0.022
BMI	8	81375457	rs16907751	C	T	0.035

BMI	8	85079709	rs2033732	C	T	0.019
BMI	9	15634326	rs4740619	T	C	0.018
BMI	9	28414339	rs10968576	G	A	0.025
BMI	9	111932342	rs6477694	C	T	0.017
BMI	9	120378483	rs1928295	T	C	0.019
BMI	9	129460914	rs10733682	A	G	0.017

Supplementary table 5.2 Associations of meat intake, genetic and clinical risk indexes with T2D in the EPIC-InterAct (n=20,628) and the UK Biobank (n=316,222) studies

Group		N	N cases	HR (95% CI)	*P* value
EPIC-InterAct					
Red meat intake	Low	6,373	2,500	1 [reference]	
	Intermediate	6,863	3,014	1.16 (1.04, 1.29)	0.006
	High	7,392	3,572	1.23 (1.08, 1.40)	0.002
GRS of T2D	Low	5,458	1,520	1 [reference]	
	Intermediate	6,585	2,710	1.79 (1.63, 1.98)	<0.001
	High	8,585	4,856	3.37 (3.08, 3.70)	<0.001
GRS of insulin resistance	Low	6,547	2,654	1 [reference]	
	Intermediate	6,828	2,987	1.18 (1.08, 1.29	<0.001
	High	7,717	3,958	1.33 (1.21, 1.47)	<0.001
	Low	6,587	2,721	1 [reference]	
GRS of BMI	Intermediate	6,905	3,080	1.03 (0.94, 1.13)	0.54
	High	7,136	3,285	0.97 (0.88, 1.07)	0.54
Cambridge diabetes risk score	Low	4,581	583	1 [reference]	
	Intermediate	5,979	2,061	3.05 (2.74, 3.40)	<0.001
	High	10,068	6,442	9.64 (8.68, 10.71)	<0.001
HbA1c	Normal	15,316	4,612	1 [reference]	
	High	3,491	2,723	5.22 (4.67, 5.85)	<0.001
UK Biobank					
Red meat intake	Low	31,065	729	1 [reference]	
	Intermediate	125,808	3,884	0.97 (0.88, 1.06)	0.47
	High	159,349	5,905	1.02 (0.93, 1.12)	0.63
GRS of T2D	Low	105,405	2,095	1 [reference]	
	Intermediate	105,399	3,348	1.63 (1.52, 1.74)	<0.001
	High	105,418	5,075	2.58 (2.43, 2.74)	<0.001
GRS of insulin resistance	Low	106,352	3,042	1 [reference]	
	Intermediate	105,568	3,517	1.19 (1.13, 1.26)	<0.001
	High	104,302	3,959	1.36 (1.29, 1.44)	<0.001
GRS of BMI	Low	105,405	3,151	1 [reference]	
	Intermediate	105,409	3,519	0.99 (0.93, 1.04)	0.61
	High	105,408	3,848	0.94 (0.89, 0.99)	0.02
Cambridge diabetes risk score	Low	104,031	721	1 [reference]	
	Intermediate	104,587	2,536	3.05 (2.74, 3.40)	<0.001
	High	107,604	7,261	9.64 (8.68, 10.71)	<0.001
HbA1c	Normal	301,051	6,349	1 [reference]	
	High	10,101	2,840	8.50 (8.01, 9.03)	<0.001
Meta-analysis					
Red meat intake	Low	37,438	3,229	1 [reference]	
	Intermediate	132,671	6,898	1.05 (0.98, 1.12)	0.02
	High	166,741	9,477	1.09 (1.01, 1.17)	0.03

Following the last page

Group		N	N cases	HR (95% CI)	*P* value
GRS of T2D	Low	110,863	3,615	1 [reference]	
	Intermediate	111,984	6,058	1.68 (1.59, 1.77)	<0.001
	High	114,003	9,931	2.79 (2.65, 2.94)	<0.001
GRS of insulin resistance	Low	112,899	5,696	1 [reference]	
	Intermediate	112,396	6,504	1.19 (1.13, 1.25)	<0.001
	High	112,019	7,917	1.35 (1.29, 1.42)	<0.001
GRS of BMI	Low	111,992	5,872	1 [reference]	
	Intermediate	112,314	6,599	1.00 (0.95, 1.05)	0.89
	High	112,544	7,133	0.95 (0.90, 0.99)	0.02
Cambridge diabetes risk score	Low	108,612	1,304	1 [reference]	
	Intermediate	110,566	4,597	3.05 (2.83, 3.29)	<0.001
	High	117,672	13,703	9.64 (8.95, 10.38)	<0.001
HbA1c	Normal	316,367	10,961	1 [reference]	
	High	13,592	5,563	7.64 (7.24, 8.05)	<0.001

HR, hazard ratio; CI, confidence interval; GRS, genetic risk score.

Supplementary table 5.3 Association between meat consumption and incident T2D in subpopulations in meta-analysis of EPIC-InterAct and UKBB

Group	Meat	N	N cases	HR	95% CI	*P* value
GRS of T2D						
Low	Low	12,141	1,858	1 [reference]		
Low	Intermediate	43,608	2,559	1.03	(0.89, 1.19)	0.74
Low	High	55,114	3,136	1.07	(0.92, 1.25)	0.37
Intermediate	Low	12,245	2,270	1 [reference]		
Intermediate	Intermediate	44,063	3,394	0.90	(0.80, 1.02)	0.09
Intermediate	High	55,676	4,269	1.00	(0.89, 1.13)	0.9
High	Low	13,052	2,974	1 [reference]		
High	Intermediate	45,000	4,794	1.17	(1.06, 1.29)	0.002
High	High	55,951	5,892	1.18	(1.06, 1.32)	0.002
GRS of insulin resistance						
Low	Low	12,543	2,196	1 [reference]		
Low	Intermediate	44,203	3,321	1.07	(0.95, 1.21)	0.26
Low	High	55,208	4,072	1.12	(0.99, 1.27)	0.08
Intermediate	Low	12,525	2,422	1 [reference]		
Intermediate	Intermediate	44,202	3,528	1.05	(0.94, 1.18)	0.39
Intermediate	High	55,508	4,395	1.14	(1.01, 1.28)	0.04
High	Low	12,370	2,484	1 [reference]		
High	Intermediate	44,266	3,898	1.05	(0.94, 1.18)	0.37
High	High	56,025	4,830	1.07	(0.95, 1.21)	0.25
Cambridge diabetes risk score						
Low	Low	16,649	1,965	1 [reference]		
Low	Intermediate	44,868	1,847	1.27	(1.07, 1.51)	0.007
Low	High	48,165	1,490	1.29	(1.07, 1.56)	0.009
Intermediate	Low	11,236	2,028	1 [reference]		
Intermediate	Intermediate	44,462	2,984	1.14	(1.02, 1.27)	0.02
Intermediate	High	55,970	3,503	1.21	(1.08, 1.37)	0.001
High	Low	9,553	3,109	1 [reference]		
High	Intermediate	43,341	5,916	1.02	(0.95, 1.11)	0.54
High	High	62,606	8,304	1.17	(1.08, 1.27)	<0.001
HbA1c						
Normal	Low	35,113	5,357	1 [reference]		
Normal	Intermediate	126,459	7,475	1.15	(1.07, 1.25)	<0.001
Normal	High	157,795	8,833	1.26	(1.16, 1.37)	<0.001
High	Low	1,734	1,208	1 [reference]		
High	Intermediate	4,840	2,170	1.13	(0.99, 1.28)	0.08
High	High	7,018	2,953	1.16	(1.02, 1.33)	0.28

HR, hazard ratio; CI, confidence interval; GRS, genetic risk score.

Supplementary table 5.4 Cumulative incidence rates by subgroups defined by meat intake and other risk factors in the EPIC-InterAct (n=20,628) and the UK Biobank (n=316,222) studies

		EPIC-InterAct		UK Biobank	
Group	**Meat**	**Cumulative incidence, %**	**ARI_adj %, (95% CI)**	**Cumulative incidence, %**	**ARI_adj %, (95% CI)**
GRS of T2D					
Low	Low	0.83	Reference	1.01	Reference
Low	Intermediate	1.22	0.13 (-0.06, 0.4)	0.92	-0.09 (-0.24, 0.11)
Low	High	1.59	0.2 (-0.02, 0.6)	0.97	-0.03 (-0.2, 0.18)
Intermediate	Low	1.75	Reference	1.81	Reference
Intermediate	Intermediate	1.97	-0.21 (-0.39, 0.07)	1.68	-0.13 (-0.36, 0.15)
Intermediate	High	2.71	0.11 (-0.22, 0.54)	1.74	-0.08 (-0.32, 0.22)
High	Low	3.05	Reference	2.56	Reference
High	Intermediate	4.04	1.08 (0.71, 2.35)	2.62	0.07 (-0.27, 0.45)
High	High	4.51	0.91 (0.32, 2.22)	2.88	0.33 (-0.03, 0.82)
GRS of insulin resistance					
Low	Low	1.68	Reference	1.53	Reference
Low	Intermediate	2.10	0.32 (0.01, 0.82)	1.47	-0.06 (-0.29, 0.21)
Low	High	2.62	0.34 (-0.02, 0.92)	1.63	0.09 (-0.17, 0.42)
Intermediate	Low	1.78	Reference	1.84	Reference
Intermediate	Intermediate	2.34	0.18 (-0.12, 0.58)	1.85	0.01 (-0.27, 0.33)
Intermediate	High	2.97	0.5 (0.14, 1.24)	1.93	0.09 (-0.2, 0.43)
High	Low	2.21	Reference	2.01	Reference
High	Intermediate	2.78	0.45 (0.04, 1.15)	1.91	-0.1 (-0.35, 0.2)
High	High	3.22	0.49 (-0.01, 1.35)	2.02	0.01 (-0.27, 0.33)
Cambridge diabetes risk score					
Low	Low	0.35	Reference	0.39	Reference
Low	Intermediate	0.50	0.08 (0, 0.24)	0.52	0.13 (0, 0.41)
Low	High	0.48	0.04 (-0.05, 0.18)	0.59	0.2 (0.09, 0.61)
Intermediate	Low	1.52	Reference	1.60	Reference
Intermediate	Intermediate	1.58	0.14 (-0.09, 0.42)	1.94	0.35 (0.03, 0.89)
Intermediate	High	1.68	0.22 (-0.04, 0.59)	2.09	0.49 (0.2, 1.16)
High	Low	4.65	Reference	5.07	Reference
High	Intermediate	5.16	0.45 (-0.08, 1.12)	4.81	-0.26 (-0.74, 0.31)
High	High	5.63	1.29 (0.79, 2.61)	5.48	0.42 (-0.17, 1.15)
HbA1c					
Normal	Low	1.04	Reference	1.12	Reference
Normal	Intermediate	1.27	0.16 (0.02, 0.36)	1.29	0.18 (0.06, 0.36)
Normal	High	1.57	0.21 (0.05, 0.48)	1.44	0.32 (0.24, 0.6)
High	Low	7.72	Reference	20.25	Reference
High	Intermediate	9.66	2.27 (0.22, 6.39)	21.38	1.13 (-2.14, 5.09)
High	High	10.80	2.6 (0.34, 7.52)	22.33	2.07 (-1.25, 6.41)

ARI_adj, absolute risk increase, adjusted for age, sex, genetic ancestry, smoking, physical activity, BMI, and family history of diabetes in both cohorts and further included country and total energy intake in EPIC-InterAct; CI, confidence interval; GRS, genetic risk score.

Supplementary table 5.5 Joint effects of meat intake, genetics and clinical risks with T2D incidence in the EPIC-InterAct (n=20,628) and the UK Biobank (n=316,222) studies after meta-analysis

Group	Meat	N	N cases	HR	95% CI		*P* value
GRS of T2D							
Low	Low	12,141	1,858	1 [reference]			
Low	Intermediate	43,608	2,559	1.07	(0.93,	1.24)	0.35
Low	High	55,114	3,136	1.21	(1.05,	1.39)	0.007
Intermediate	Low	12,245	2,270	1.99	(1.73,	2.28)	<0.0001
Intermediate	Intermediate	44,063	3,394	1.83	(1.60,	2.07)	<0.0001
Intermediate	High	55,676	4,269	1.93	(1.68,	2.22)	<0.0001
High	Low	13,052	2,974	3.00	(2.61,	3.42)	<0.0001
High	Intermediate	45,000	4,794	3.47	(3.06,	3.89)	<0.0001
High	High	55,951	5,892	3.41	(2.99,	3.86)	<0.0001
GRS of insulin resistance							
Low	Low	12,543	2,196	1 [reference]			
Low	Intermediate	44,203	3,321	1.07	(0.95,	1.20)	0.28
Low	High	55,208	4,072	1.14	(1.01,	1.28)	0.04
Intermediate	Low	12,525	2,422	1.19	(1.05,	1.35)	0.007
Intermediate	Intermediate	44,202	3,528	1.25	(1.11,	1.39)	0.0002
Intermediate	High	55,508	4,395	1.37	(1.21,	1.53)	<0.0001
High	Low	12,370	2,484	1.45	(1.27,	1.65)	<0.0001
High	Intermediate	44,266	3,898	1.47	(1.31,	1.64)	<0.0001
High	High	56,025	4,830	1.49	(1.33,	1.68)	<0.0001
Cambridge diabetes risk score							
Low	Low	16,649	1,965	1 [reference]			
Low	Intermediate	44,868	1,847	1.29	(1.10,	1.53)	0.02
Low	High	48,165	1,490	1.33	(1.12,	1.59)	0.02
Intermediate	Low	11,236	2,028	3.38	(2.90,	3.94)	<0.0001
Intermediate	Intermediate	44,462	2,984	3.72	(3.21,	4.30)	<0.0001
Intermediate	High	55,970	3,503	3.97	(3.42,	4.62)	<0.0001
High	Low	95,53	3,109	10.21	(8.83,	11.84)	<0.0001
High	Intermediate	43,341	5,916	10.87	(9.44,	12.49)	<0.0001
High	High	62,606	8,304	12.49	(10.83,	14.41)	<0.0001
HbA1c							
Normal	Low	35,113	5,357	1 [reference]			
Normal	Intermediate	126,459	7,475	1.12	(1.00,	1.25)	0.05
Normal	High	157,795	8,833	1.17	(1.02,	1.33)	0.02
High	Low	1,734	1,208	4.83	(3.97,	5.92)	<0.0001
High	Intermediate	4,840	2,170	6.19	(5.09,	7.49)	<0.0001
High	High	7,018	2,953	6.18	(5.02,	7.65)	<0.0001

HR, hazard ratio; CI, confidence interval; GRS, genetic risk score.

Supplementary Table 6.1 Characteristics of 23 cohorts to study the association between meat consumption and incident type 2 diabetes in InterConnect

Study	Location	Participants	Recruitment time-frame	Baseline sample size
America				
ARIC	USA	Ethnically-representative men and women in 4 communities	1987-1989	15,792
CARDIA	USA	Black and white men and women aged 18-30 years representative of 4 cities	1985-1986	5,115
PRHHP	Puerto Rico	Rural and urban men aged 45-64 years	1965-1968	9,824
WHI	USA	Postmenopausal women aged 50-79 years at 40 centres.	1994-1998	9,676
MESA	USA	Ethically-stratified asymptomatic men and women aged 45-84 years	2000-2002	6,814
MEC	USA	Men and women in Hawaii and California aged 45-75 years, for 5 ethnic groups (Japanese Americans, African Americans, European Americans, Latinos and Native Hawaiians)	1993-1996	215,000
NHS I	USA	Married registered nurses, women aged 30-55 years, who lived in the 11 most populous states (California, Connecticut, Florida, Maryland, Massachusetts, Michigan, New Jersey, New York, Ohio, Pennsylvania, and Texas) in the US	1976	121,700
HPFS	USA	Men in health professions aged 40-75 years	1986	51,529
NHS II	USA	Women between 25-42 years old	1989	116,430
Eastern Mediterranean				
Golestan	Iran	Healthy men and women aged 40-75 years from urban (20%) and rural areas (80%), including Turkmen (74 %) and Non-Turkmen (26%) ethnicity	2004-2008	50,045
Europe				
COSM	Sweden	Men born aged 45-79 years living in Västmanland and Örebro counties	1997	45,906
SMC	Sweden	Women aged 39-76 living in Uppsala Västmanland counties	1987-1990	66,651
EPIC-InterAct	European	A case-cohort of T2D occurring in the EPIC cohorts between 1991 and 2007 from 8 of the 10 EPIC countries and a subcohort of individuals randomly selected from those with available stored blood and buffy coat, stratified by centre	1991-1998	12403 cases, 16154 subcohort.
UK Biobank	UK			
Western Pacific				
CKB	China	Men and women aged 30-79 years from five urban and five rural regions	2004-2008	512,891
ALSWH-Young	Australia	Women aged 25-30 years	2003	9,081
ALSWH-MidAge	Australia	Women aged 50-55 years	2001	11,226

ALSWH, Australian Longitudinal Study on Women's Health; ARIC, Atherosclerosis Young Adults Study; CARDIA, Coronary Artery Risk Development in Young Adults; CKB, China Kadoorie Biobank; COSM, Cohort of Swedish Men; EPIC, European Prospective Investigation into Cancer (in Denmark France, Germany, Italy, Netherlands, Spain, Sweden, and UK); HPFS, Health Professionals Follow-up Study; MEC, Multi-ethnic Cohort Study; MESA, Multi-Ethnic Study of Atherosclerosis; NHS, Nurses' Health Study; PRHHP, Puerto Rico Heart Health Program; SMC, Swedish Mammography Cohort; WHI, Women's Health Initiative Study. The COSM and SMC used the same protocol and were made available as one combined dataset.

Supplementary Table 6.2 Details of exposure variables used to study the association between meat consumption and incident type 2 diabetes in InterConnect

Study	Dietary assessment method	Dietary assessment validation	Units provided
America			
ARIC	Interviewer-administered FFQ	√	Frequencies (standerdised serving size)
CARDIA	Interviewer-administered dietary history		Serving per day
PRHHP	24-hour diet recall		Quarter cup per day
WHI	FFQ		Serving per day
MESA	FFQ		Serving per day
MEC	FFQ	√	g/d
NHS I	FFQ	√	g/d
HPFS	FFQ	√	g/d
NHS II	FFQ	√	g/d
Eastern Mediterranean			
Golestan	FFQ		g/d
Europe			
COSM/SMC	FFQ	√	g/d
EPIC-InterAct Denmark	FFQ		
EPIC-InterAct France	Quantitative questionnaire	√	g/d
EPIC-InterAct Germany	Quantitative questionnaire	√	g/d
EPIC-InterAct Italy	FFQ or Quantitative questionnaire	√	g/d
EPIC-InterAct Netherlands	Quantitative questionnaire	√	g/d
EPIC-InterAct Spain	Quantitative questionnaire	√	g/d
EPIC-InterAct Sweden	FFQ	√	g/d
EPIC-InterAct UK	FFQ	√	g/d
UK Biobank	FFQ, multiple 24-hour diet recalls	X	g/d
Western Pacific			
CKB	FFQ, multiple 24-hour diet recalls	√	g/d
ALSWH-Young	FFQ	√	g/d
ALSWH-MidAge	FFQ	√	g/d

ALSWH, Australian Longitudinal Study on Women's Health; ARIC, Atherosclerosis Young Adults Study; CARDIA, Coronary Artery Risk Development in Young Adults; CKB, China Kadoorie Biobank; COSM, Cohort of Swedish Men; EPIC, European Prospective Investigation into Cancer; HPFS, Health Professionals Follow-up Study; MEC, Multi-ethnic Cohort Study; MESA, Multi-Ethnic Study of Atherosclerosis; NHS, Nurses' Health Study; PRHHP, Puerto Rico Heart Health Program; SMC, Swedish Mammography Cohort; WHI, Women's Health Initiative Study. The COSM and SMC used the same protocol and were made available as one combined dataset.

Supplementary Table 6.3 Portion sizes used to study the association between meat consumption and incident type 2 diabetes in InterConnect

Study	Variable	Original reporting quantity	Assigned portion, g
ARIC	Red meat as sandwich or mixed dish	Standardise serving size	85
	Red meat as main dish	Standardise serving size	113
	Hamburger	Standardise serving size	113
	Hot dog	Standardise serving size	57
	Processed meat	Standardise serving size	57
	Bacon	Standardise serving size	57
	Chicken without skin	Standardise serving size	113
	Chicken with skin	Standardise serving size	113
CARDIA	Beef	Serving sizes were not specified	113
	Veal	Serving sizes were not specified	113
	Lamb	Serving sizes were not specified	113
	Fresh pork	Serving sizes were not specified	113
	Cured pork	Serving sizes were not specified	57
	Cold cuts and sausage	Serving sizes were not specified	57
	Game	Serving sizes were not specified	113
	Poultry	Serving sizes were not specified	113
	Fried chicken	Serving sizes were not specified	113
MESA	Red meat	Serving sizes were not specified	113
	Processed meat	Serving sizes were not specified	57
	Poultry	Serving sizes were not specified	113
PRHHP	Beef or veal	Ounces per day	28.3
	Pork	Ounces per day	28.3
	Ham	Ounces per day	28.3
	Lunch meat or sausage	Ounces per day	28.3
	Pigs feet	100g per day	100
	Chicken	Ounces per day	28.3
WHI	Ground meat (hamburger/meatloaf/picadillo)	Med serving/d	85
	Beef/pork/lamb m dish (steak/roast/ham)	Med serving/d	113
	Beef/pork/lamb as sandwich (steak/barbeque)	Med serving/d	85
	Stew/potpie/casserole w/meat or chicken	Med serving/d	42.5
	Chilli with meat and beans	Med serving/d	42.5
	Fried chicken	Med serving/d	113
	Chicken and turkey (roast/stew/broil)	Med serving/d	113
	Lunch meat (ham/turkey/lean meat)	Med serving/d	57
	Other lunch meat (bologna/salami/spam)	Med serving/d	57
	Hot dogs/chorizo/sausage/bratwurst	Med serving/d	85
	Bacon/breakfast sausage/scrapple	Med serving/d	28

ARIC, Atherosclerosis Young Adults Study; CARDIA, Coronary Artery Risk Development in Young Adults; MESA, Multi-Ethnic Study of Atherosclerosis; PRHHP, Puerto Rico Heart Health Program; WHI, Women's Health Initiative Study.

Supplementary Table 6.4 Covariate variables used to study the association between meat consumption and incident type 2 diabetes in InterConnect

Variables	ARIC	WHI	PRHHP	MESA	CARDIA	MEC	NHS I	NHS II	HPFS	ALSWH MidAge	ALSWH Young	CKB	Golestan
Age	√	√	√	√	√	√	√	√	√	√	√	√	√
Sex	√	√	√	√	√	√				√	√	√	√
Education	√	√	√	√	√	√	√	√	√	√	√	√	√
Smoking	√	√	√	√	√	√	√	√	√	√	√	√	√
Physical activity	√	√	√	√	√	√	√	√	√	√	√	√	√
Alcohol	√	√	√	√	√	√	√	√	√	√	√	√	
TEI	√	√	√	√	√	√	√	√	√	√	√		√
BMI	√	√	√	√	√	√	√	√	√	√	√	√	√
Waist	√	√		√	√	√	√	√	√	√	√	√	√
Comorbidity	√	√	√	√	√	√	√	√	√	√	√	√	√
Family history of T2D	√	√	√	√	√	√	√	√	√	√	√	√	√
Vegetable	√	√	√	√	√	√	√	√	√	√	√	√	√
Fruit	√	√	√	√	√	√	√	√	√	√	√	√	√
Fish	√	√	√	√	√	√	√	√	√	√	√	√	√
Legume	√	√	√	√	√	√	√	√	√	√	√	√	√
Soy		√		√		√	√	√	√	√	√	√	√
Nuts	√	√			√	√	√	√	√	√	√		√
Potatoes	√	√	√	√		√	√	√	√	√	√		√
Dairy	√	√	√	√	√	√	√	√	√	√	√	√	√
Egg	√	√	√	√	√	√	√	√	√	√	√		√
Cereals	√	√	√	√	√	√	√	√	√	√	√		√
Whole grain	√	√		√	√	√	√	√	√	√	√		
Pasta	√	√		√	√	√	√	√	√	√	√		√
Rice	√	√	√	√		√	√	√	√	√	√		√
SSB	√	√	√	√	√	√	√	√	√				√
Coffee	√	√	√	√	√	√	√	√	√				
Tea	√			√	√		√	√	√				
Fibre		√		√	√		√	√	√	√	√		
Cooking fat	√	√	√	√	√	√	√	√	√	√	√		√

Following the last page

Variables	EPIC-InterAct France	EPIC-InterAct Italy	EPIC-InterAct Spain	EPIC-InterAct UK	EPIC-InterAct Netherlands	EPIC-InterAct Germany	EPIC-InterAct Sweden	EPIC-InterAct Denmark	SMC/ COSM	UK Biobank
Age	√	√	√	√	√	√	√	√	√	√
Sex	√	√	√	√	√	√	√	√	√	√
Education	√	√	√	√	√	√	√	√	√	√
Smoking	√	√	√	√	√	√	√	√	√	√
Physical activity	√	√	√	√	√	√	√	√	√	√
Alcohol	√	√	√	√	√	√	√	√	√	√
TEI	√	√	√	√	√	√	√	√	√	
BMI	√	√	√	√	√	√	√	√	√	√
Waist	√	√	√	√	√	√	√	√	√	√
Comorbidity	√	√	√	√	√	√	√	√	√	√
Family history of T2D	√	√	√	√	√	√	√	√	√	√
Vegetable	√	√	√	√	√	√	√	√	√	√
Fruit	√	√	√	√	√	√	√	√	√	√
Fish	√	√	√	√	√	√	√	√	√	√
Legume	√	√	√	√	√	√	√	√	√	
Soy									√	
Nuts	√	√	√	√	√	√	√	√	√	
Potatoes	√	√	√	√	√	√	√	√	√	
Dairy	√	√	√	√	√	√	√	√	√	√
Egg	√	√	√	√	√	√	√	√	√	√
Cereals	√	√	√	√	√	√	√	√	√	√
Whole grain	√								√	
Pasta									√	
Rice	√	√	√	√	√	√	√	√	√	
SSB	√	√	√	√	√	√	√	√	√	√
Coffee	√	√	√	√	√	√	√	√	√	√
Tea	√	√	√	√	√	√	√	√	√	√
Fibre	√	√	√	√	√	√	√	√	√	
Cooking fat	√	√	√	√	√	√		√	√	

TEI, total energy intake; SSB, sugar-sweetened beverages.

Following the last page

ALSWH, Australian Longitudinal Study on Women's Health; ARIC, Atherosclerosis Young Adults Study; CARDIA, Coronary Artery Risk Development in Young Adults; CKB, China Kadoorie Biobank; COSM, Cohort of Swedish Men; EPIC, European Prospective Investigation into Cancer; HPFS, Health Professionals Follow-up Study; MEC, Multi-ethnic Cohort Study; MESA, Multi-Ethnic Study of Atherosclerosis; NHS, Nurses' Health Study; PRHHP, Puerto Rico Heart Health Program; SMC, Swedish Mammography Cohort; WHI, Women's Health Initiative Study. The COSM and SMC used the same protocol and were made available as one combined dataset.