Disentangling the Underlying Mechanisms Linking Epigenetic, Metabolic and Environmental Determinants of Type 2 Diabetes

CAROLINA PATRICIA OCHOA ROSALES

Disentangling the Underlying Mechanisms Linking Epigenetic, Metabolic and Environmental Determinants of Type 2 Diabetes

Carolina Patricia Ochoa Rosales

ISBN 978-94-6361-460-3

Layout and printed by: Optima Grafische Communicatie (www.ogc.nl)

Disentangling the Underlying Mechanisms Linking Epigenetic, Metabolic and Environmental Determinants of Type 2 Diabetes

Ontrafeling van de onderliggende mechanismen die epigenetische, metabole en omgevingsdeterminanten van type 2 diabetes met elkaar verbinden

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof. dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on
Wednesday, 14th of October 2020 at 13:30 hrs

by
Carolina Patricia Ochoa Rosales

born in Valparaíso, Chile

Erasmus University Rotterdam

DOCTORAL COMMITTEE

Promotor:	Prof. Dr. M. Arfan Ikram
Other members:	Prof. Dr. Jill Pell Prof. Dr. Bruno H. C. Stricker Dr. Joyce B.J. van Meurs
Copromotor:	Dr. ir. Trudy Voortman

PARANYMPHS

Silvana C. E. Maas
Banafsheh Arshi

To Alejandro, Mario and Patricia

MANUSCRIPTS THAT FORM THE BASIS OF THIS THESIS

Chapter 2

Carolina Ochoa-Rosales*, Eralda Asllanaj*, Glisic Marija, Jana Nano, Taulant Muka, Oscar Franco. (2019). Chapter 11 - Chromatin landscape and epigenetic biomarkers for clinical diagnosis and prognosis of type 2 diabetes mellitus. In: Sharma S, ed. Prognostic Epigenetics: Academic Press; 2019:289-324. Doi:10.1016/B978-0-12-814259-2.00012-1.

Eralda Asllanaj, **Carolina Ochoa-Rosales***, Xiaofang Zhang*, Jana Nano, Wichor Bramer, Eliana Portilla, Kim Braun, Valentina González-Jaramillo, Wolfgang Ahrens, Arfan Ikram, Mohsen Ghanbari, Trudy Voortman, Oscar Franco, Taulant Muka, Marija Glisic. (2020). Sexually Dimorphic DNA-methylation in Cardiometabolic Health: A Systematic Review. Maturitas 2020;135:6-26. 10.1016/j.maturitas.2020.02.005.

Chapter 3

Fariba Ahmadizar, **Carolina Ochoa-Rosales**, Marja Glisic, Oscar Franco, Taulant Muka, Bruno Stricker. (2019). Associations of statin use with glycaemic traits and incident type 2 diabetes. Br J Clin Pharmacol 2019; 85:993-1002. Doi:85. 10.1111/bcp.13898.

Carolina Ochoa-Rosales, Eliana Portilla, Jana Nano, Rory Wilson, Benjamin Lehne, Pashupati Mishra, Xu Gao, Mohsen Ghanbari, Oscar Rueda-Ochoa, Diana Juvinao-Quintero, Marie Loh, Weihua Zhang, Jaspal Kooner, Hans Grabe, Stephan Felix, Ben Schöttker, Yan Zhang, Christian Gieger, Martina Müller-Nurasyid, Margit Heier, Annette Peters, Terho Lehtimäki, Alexander Teumer, Hermann Brenner, Melanie Waldenberger, M. Arfan Ikram, Joyce B.J. van Meurs, Oscar H. Franco, Trudy Voortman, John Chambers, Bruno H. Stricker, Taulant Muka. (2020). Epigenetic Link Between Statin Therapy and Type 2 Diabetes. Diabetes Care 2020; 43:875-84. dc191828. 10.2337/dc19-1828.

Chapter 4

Carolina Ochoa-Rosales, Niels van der Schaft, Kim Braun, Frederick K. Ho, Fanny Petermann-Rocha, Jill P. Pell, M. Arfan Ikram, Carlos A. Celis-Morales*, Trudy Voortman*. C-reactive protein partially mediates the inverse association between coffee consumption and risk of type 2 diabetes (Submitted for publication).

Chapter 5

Carolina Ochoa-Rosales, Niels van der Schaft, Frederick K. Ho, Jill P Pell, M. Arfan Ikram, Carlos A. Celis-Morales*, Trudy Voortman* Type 2 diabetes and sex differences in the association between serum uric acid and risk of fatal and nonfatal cardiovascular related outcomes (Manuscript).

* Denotes equal contribution

TABLE OF CONTENTS

Chapter 1.

General introduction

INTRODUCTION

The global problem of type 2 diabetes

Type 2 diabetes mellitus is a metabolic disease characterized by chronic insulin resistance and declining pancreatic beta-cell function, with consequent impaired fasting or postprandial glycemia.[1] Type 2 diabetes gives rise to several health complications[2] such as damage to blood vessels in e.g. the heart, eyes, kidneys, nerves leading to other diseases/disability and premature death. For example, type 2 diabetes is a major risk factor for cardiovascular events, conferring about two-fold higher risk of vascular disease independent of other conventional risk factors.[3] The World Health Organization (WHO) has estimated that in the past decades diabetes prevalence around the world has doubled, rising from 4.7% in 1980 to 8.5% in 2014, the vast majority (90-95%) of cases being type 2 diabetes.[4] With this rise in prevalence and its health complications, diabetes has become the seventh leading cause of death in 2016 and the fourth main noncommunicable disease, accounting for 4% of the deaths worldwide.[5] Its prevalence is expected to keep increasing.

Type 2 diabetes is not only a burden for individual patients because of the adverse effects on their health and quality of life, but also a major public health concern given the economic burden that diabetes means to nations worldwide. The global cost of type 2 diabetes, including both direct treatment costs and indirect costs due to type 2 diabetes-associated morbidity and premature death, was estimated to be 1.8% of global gross domestic product (GDP) for 2015.[6] Therefore, it is imperative to invest in improving strategies to prevent or delay the onset of type 2 diabetes, as well as for implementing early detection methods.

Sex differences in type 2 diabetes and cardiometabolic health

Although cardiometabolic disease events remain a leading cause of death worldwide for both sexes,[7] several epidemiological studies have reported differences in intermediate cardiovascular risk factors and risk of type 2 diabetes between men and women. Moreover, the global trend in the last decades shows that the age-standardized diabetes prevalence has risen in both sexes, but that this increment is stronger in men than in women. Between 1980 and 2014 the prevalence of diabetes increased from 4.3% to 9.0% in men while from 5.0% to 7.9% in women.[8] Recent reports summarizing the available evidence concluded that, overall, adult men are at a higher risk of developing type 2 diabetes and cardiovascular disease as compared to adult women.[9,10] This was the case across the majority,[11] but not all,[12] of the ethnicities investigated. However, evidence of disparities in type 2 diabetes between younger men and women remain controversial.[13] Although cardiometabolic risk factors

and diseases are unequally distributed between sexes, the potential mechanisms underlying these differences are not completely understood.[9] Sex hormones may play a role, because the advantage that women seem to have over men progressively disappears with ageing, especially after menopause.[14] Nevertheless, to shed light on potential biological mechanisms explaining the observed sex differences sill further research is needed.

Epigenetics: a potential novel mechanism involved in type 2 diabetes onset

Type 2 diabetes is a multifactorial product of the complex interplay of environmental and genetic factors. History of parental diabetes may confer up to 6-fold higher risk for diabetes as compared with those with no family history.[15] It was hypothesized that type 2 diabetes-associated genetic variants identified in large population studies[16,17] may account for the heritability of the disease. However, studies evaluating the predicting value of a genetic component added to the classic risk factors to predict type 2 diabetes risk showed a slight improvement in the prediction as compared to models including the classic factors alone.[18-21] For example, a study building a prediction model to assess type 2 diabetes risk in adults reported C statistics (a measure indicating to what extent the model can discriminate the risk of type 2 diabetes) up to 0.900 when using only classic risk factors, such as age, family history of diabetes, body mass index, smoking, blood pressure and plasma high-density lipoprotein cholesterol, triglycerides and glucose; while a C statistics of 0.901 was observed when adding a genotype score to the model.[21] Furthermore, this suggests that there is a proportion of type 2 diabetes risk and its heritability that is not fully explained by these classic and genetic markers. In search for novel markers that may explain the remaining variation, epigenetics has risen as potential mechanism.

Epigenetics is the study of the heritable changes in the function of genes that do not comprise variations in the DNA sequence of nucleotides, but refer to covalent modifications of histones and DNA,[22] altering chromatin structure and consequently DNA transcription. Some epigenetic mechanisms are changes in DNA methylation and modifications of histone tails (acetylation, methylation, phosphorylation and ubiquitination).[23] This type of gene expression regulation is needed during developmental stages where it controls tissue differentiation and cellular responsiveness. This is how cells in the body, containing the same DNA sequence give origin to specialized organs and tissues with different functions. Some of these epigenetic marks are heritable from cell to cell and transgenerationally from parent to offspring to grand-offspring,[24] but epigenetic patterns are also most sensitive to suffer modifications due to the early exposure to environmental influences, which can thereby

determine future phenotype and affect health.[25] These influences can be dietary, physical, chemical, stochasticity and random chance. Moreover, the epigenome can still be affected by unfavorable lifetime environmental exposures. Some risk factors previously associated with type 2 diabetes and cardiometabolic health such as smoking, sedentarism, drugs, obesity, ageing, unhealthy diet and several other lifestyle and biological factors have also been associated with epigenetic changes which in turn may cause disease,[26,27] including diabetes.[28]

Among the epigenetic changes, DNA methylation is one of the most studied and has been more frequently associated with gene silencing. DNA methylation refers to a reversible modification of the DNA, where a methyl (-CH_3) group is attached to a cytosine at the 5′-position, that is located 5′ to a guanosine. This dinucleotide is commonly named CpG island, wherein p refers to the phosphodiester linkage between the cytosine and guanosine nucleotides.[29] The transfer of the methyl group is performed by a family of enzymes called DNA methyltransferases (DNMTs).[30]

Modifications in DNA methylation patterns may be in the biological pathway of the disease pathophysiology through dysregulation of gene expression that may in turn contribute to e.g. insulin resistance and type 2 diabetes onset. Indeed, several studies have compared DNA methylation patterns between type 2 diabetes patients and healthy subjects, as well as risk of future type 2 diabetes, and have observed numerous differentially methylated CpG sites in diabetes.[27,31-39] However the evidence is still controversial, with an existing lack of replication across studies.[27] If methylation patterns indeed differ in early stages of disease development, DNA methylation marks can be promising candidates to be implemented as biomarkers in clinical practice[22] for early diagnose of diseases. For example, insulin resistance starts appearing years before the clinical manifestation and diagnose of type 2 diabetes, while consequent organ damage is already occurring. Evidence suggests that early insulin therapy can help correct the underlying pathogenetic abnormalities in type 2 diabetes and improve long-term glycemic control.[40] Hence, efforts are being made in order to use DNA methylation signatures as biomarkers to identify individuals at risk of type 2 diabetes and to monitor the progression of this disease.[41-44]

Taking into consideration the abundant evidence identifying DNA methylation patterns of type 2 diabetes, it is of great interest to explore to what extent known and novel environmental factors exert their diabetogenic effect by promoting DNA methylation changes, which in turn may lead to the onset of this disease. Furthermore, the missing heritability and the proportion of the disease risk that is not fully explained by the currently used measures in medical practice are leading scientists

to study lesser-understood environmental exposures affecting cardiometabolic health and their potential effect on the epigenome. However, this is an incipient field and further efforts are needed.

Environmental risk factors in type 2 diabetes pathophysiology

Well-known classic environmental risk factors for type 2 diabetes are older age, obesity, family history of diabetes, smoking, increased levels of plasma triglycerides and glucose, low levels of plasma high-density lipoprotein cholesterol and high blood pressure. In the search for novel factors and markers, several studies have recently explored the role of lesser-known factors in type 2 diabetes onset. Oxidative stress and inflammatory states,[45] as well as some medications such as anti-hypertensives, antidepressants and statins,[46-48] have been hypothesized to promote the development of type 2 diabetes.

Several studies suggest that oxidative stress plays a role in the pathogenesis of chronic inflammation and related diseases like type 2 diabetes and its complications.[45,49,50] Oxidative stress occurs during the state of redox equilibrium disruption, an imbalance between the production of reactive oxygen species (ROS) and antioxidant capacity, and the body's response to eliminate them leads to chronic inflammation.[51] Experimental studies have suggested that oxidative stress decreases insulin secretion in beta cells and impairs glucose uptake in adipose and muscular tissues.[49,52,53] Obesity, a state of excessive accumulation of adipose tissue, may induce systemic oxidative stress leading to impaired production of adipokines in the adipocytes, such as leptin and adiponectin.[54] Adiponectin, downregulated in obese states, has shown to exert an anti-inflammatory effect;[55] as well as improvement of insulin sensitivity;[56] while leptin, increased in obese subjects, is involved in the regulation of the energy-balance by controlling food intake and energy expenditure,[56] and has been positively correlated with markers of inflammation such as C-reactive protein.[57]

Several studies have reported pro-inflammatory cytokines in association with type 2 diabetes,[58-62] which is called a state of chronic inflammation.[50] Furthermore, a recent study investigating the relationship between classic and novel inflammatory markers with type 2 diabetes found that C-reactive protein (CRP), Extracellular Newly identified Receptor for Advanced Glycation End-products binding protein (EN-RAGE), interleukin 13 (IL13), interleukin 17 (IL17), interleukin 18 (IL18), interleukin 1 Receptor Antagonist (IL1ra), Complement Factor (CFH), Complement 3 and Tumor Necrosis Factor Receptor 2 (TNFRII) were associated with incident type 2 diabetes in Caucasian population.[63] The available evidence has given rise to the study of the potential use of anti-inflammatory medications to improve glycemic

and metabolic profiles,[64,65] albeit the underlying pathophysiological mechanisms are not fully understood.[50,66]

Additional to the classic risk factors, the potential role of oxidative stress and the novel inflammatory biomarkers identified, recent evidence has shown that some medications may contribute to the early onset of type 2 diabetes, being among them statins,[46,67] antidepressants[47] and beta-blockers.[48] Statins class of drugs are have a lipid lowering effect, and are well known to effectively reduce the risk of cardiovascular events and CVD-related mortality.[68] This drug targets the enzyme HMG-CoA reductase (3-hydroxy-3-methyl-glutaryl-coenzyme A reductase), inhibiting the *de novo* synthesis of cholesterol. Nevertheless, recent meta-analyses[46,67,69] of epidemiological and also experimental[70,71] studies have provided evidence of a diabetogenic effect of statins, presumably through insulin secretion, sensitivity, and beta-cell dysfunction. However, evidence on the mechanistic pathways underlying these associations is lacking. Interestingly, statins have been studied in relation to epigenetic changes,[72] nevertheless a possible link with DNA methylation and type 2 diabetes has not been explored.

Protective strategies against type 2 diabetes development: diet

Most of the underlying risk factors responsible for the enormous rise in diabetes prevalence are modifiable risk factors such as lifestyle. Among factors that can be improved to help prevent type 2 diabetes, healthy dietary patterns have shown to play a protective role against type 2 diabetes onset. Diets rich in fruits, vegetables, wholegrains, legumes, nuts and seeds; while lower in red or processed meat, refined grains and sugared beverages; and moderate in alcohol intake have been related to decreased risk of type 2 diabetes and improvement of glycemic control.[73] Part of this effect of a healthy diet may be explained by its high content in polyphenols and other antioxidant compounds, who can inhibit ROS formation or interact with ROS to prevent tissue damaged, thus modulating the inflammatory response involved in chronic inflammatory diseases such as type 2 diabetes.[51]. Furthermore, the total antioxidant capacity of a diet,[74] as well as dietary polyphenols alone,[75,76] have demonstrated to have anti-diabetic effects. Foods that are particularly high in antioxidants and polyphenols are mainly plant-based foods, and include vegetables, fruits, cocoa, tea, extra-virgin olive oil, red wine and coffee. When it comes to beverages, a study assessing the total antioxidant capacity of a Mediterranean diet showed that coffee, followed by red wine, was the beverage with the highest antioxidant capacity.[77] The rich antioxidant capacity of coffee may be due to the high content of several bioactive compounds and micronutrients such as chlorogenic acids, caffeine, cafestol, kahweol, melanoidins, polyphenols and trigonelline, although the exact

composition depends mainly on the genus and roasting process.[78] While findings from a meta-analysis on clinical trials provided inconsistent results on the association between coffee or caffeine intake and serum levels of inflammation markers (C-reactive protein, adiponectin, interleukins), a recent study reported positive associations between coffee consumption of ≥ 4 cups/day and favorable profiles of blood markers related to metabolic and inflammatory pathways such as C-peptide, insulin-like growth factor binding protein 3 (IGFBP-3), estrone, total estradiol, free estradiol, leptin, C-reactive protein (CRP), interleukin-6 (IL6), tumor necrosis factor receptor 2 (TNFR-2), sex hormone-binding globulin (SHBG), total testosterone, total adiponectin, and high-molecular-weight (HMW) adiponectin.[79] Interestingly, existing evidence from a meta-analysis and umbrella review concluded that coffee is associated with decreased risk of developing T2D.[80,81] Whether the apparent protective effect that coffee consumption exerts on type 2 diabetes risk is mediated by a potential modulation of the inflammatory response needs further investigation.

STUDY DESIGN AND POPULATIONS

The work of this thesis includes reviews and original studies. The exponential increase in the number of published studies demands the use of a more practical approach to gather the current knowledge. In a systematic review, a well-defined search procedure and strict statistical protocols are followed to extract data from selected existing literature in order to summarize the current evidence.[82] Therefore, systematic reviews are evidence-based approaches to provide an unbiased overview on the topic of interest. In chapter 2 of this thesis, systematic and narrative reviews were performed to summarize the current evidence on DNA methylation, type 2 diabetes and related cardiometabolic traits.

The original studies presented in this thesis were embedded in prospective cohort studies. In the prospective cohort design, a representative sample of a population is followed in time until the occurrence of endpoints, such as type 2 diabetes. This design allows to characterize the exposure, i.e., risk or protective factor, before the occurrence of the event of interest or disease onset, as well as the identification of biomarkers with potential predictive value, that appear before the diagnosis of the disease and that may change over time.[83] Hence, a prospective cohort design is useful in the study of human complex diseases, wherein environmental factors or genes-environment interactions play a role.[83] Further advantages of the prospective cohort designs are the avoidance of recall bias and participants' selection bias. Among the disadvantages, the prospective cohort design needs a long duration of follow-up, a

large sample size and regular follow-up checks in order to produce sufficient incident cases. Such characteristics make this kind of study costly.[83] The prospective cohort studies in which the original studies included in this thesis were embedded in, were mainly the Rotterdam Study (RS)[84] and the UK Biobank (UKBB).[85,86] Other prospective cohorts contributing to this work were the Kooperative Gesundheitsforschung in der Region Augsburg-F4 (KORA-F4),[87] The London Life Sciences Prospective Population Study (LOLIPOP),[88] the Epidemiologische Studie zu Chancen der Verhütung, Früherkennung und optimierten Therapie chronischer Erkrankungen in der älteren Bevölkerung (ESTHER),[89] and the Study of Health Pomerania-Trend (SHIP-Trend).[90]

Chapters 3 and 4 of this thesis used data from the Rotterdam Study.[84] RS is a population-based prospective cohort study of middle-aged and elder participants living in the well-defined Ommoord district in the city of Rotterdam, the Netherlands. The aim of this study is to investigate the risk factors and occurrence of chronic diseases such as cardiovascular, endocrine, oncological, respiratory, hepatic, neurological, ophthalmic, psychiatric and dermatological diseases. The Rotterdam Study comprises three sub-cohorts. The first one (RS-I) started in 1990 and recruited 7,983 subjects aged 55 years or above. The second sub-cohort (RS-II) was initiated in 2000 and enrolled 3,011 individuals who had turned 45 years old since 1989. The third sub-cohort (RS-III) commenced in 2006 and comprised 3,932 participants aged 45 years and above. Currently, the Rotterdam Study is composed of a total of 14,926 subjects. Follow up visits were performed every four or five years. The visits to the research center included physical and functional measures, completion of questionnaires and blood samples were taken to assess concentrations of lipids, glycemic traits, and other biomarkers as well as for genetics and epigenetic measurements. DNA methylation levels in blood were determined in a random sample of 1,454 subjects from the third visit of the second cohort (RS-II-3) and in a non-overlapping sample of the first and second visit of the third cohort (RS-III-1 and RS-III-2). Data on dietary intake was measured at baseline visits of all three cohorts using validated semi-quantitative FFQs and in home interviews by a trained interviewer. Data on medication were obtained from digital pharmacy and physician's records.

Specifically, **chapter 3.2** used additional data from KORA-F4,[87] a follow-up study to KORA-S4 cohort, established in 1996 in Augsburg, Germany; LOLIPOP,[88] a prospective cohort of South-Asians residing in London, United Kingdom, aged 35 to 75 years; ESTHER, [89] a population-based study recruited in Saarland, Germany, with participants aged 50 to 75 years; and SHIP-Trend is a population-based study of participants from northeast Germany.[90]

Chapters 4 and 5 used data from the UK Biobank[85,86] The UKBB is a prospective population-based cohort study in the United Kingdom that recruited 502,549 individuals aged 37 to 73 years. This study was established to investigate genetic and non-genetic determinants of many diseases and health-related outcomes in the middle-aged and elderly. Enrollment and data assessment was carried out between April 2006 and December 2010, at 22 research centers across England, Scotland and Wales. This design allowed socioeconomic and ethnic heterogeneity, as well as an urban–rural mix among the participants. During the visits to the research centers participants had physical and functional measures, had to complete questionnaires, and their biological samples were collected to determine biomarkers and other laboratory assays, as well as for genetic assessment.

AIM OF THIS THESIS AND OUTLINE

In this thesis, I sought to investigate potential underlying mechanisms explaining associations of protective and adverse determinants of type 2 diabetes.

Chapter 2 focuses on epigenetics and type 2 diabetes and other cardiometabolic determinants. Specifically, in **chapter 2.1** we reviewed the clinical use of epigenetic marks as biomarkers for diagnosis and prognosis of type 2 diabetes; and in **chapter 2.2** we reviewed epigenetic sex differences in cardiometabolic traits.

Chapter 3 aimed to dissect the association between a novel risk factor for type 2 diabetes, statin therapy, and the incidence of the disease. Particularly, in **chapter 3.1** we explored the association between statin treatment and type 2 diabetes, including the study of the different types of statins. In **chapter 3.2** we sought to identify DNA methylation signatures associated with statin therapy that may mediate the diabetogenic effect of statins use.

Chapter 4 devoted to explore the potential role of in inflammatory markers as mediators in the observed beneficial effect of coffee consumption on type 2 diabetes onset.

In chapter 5 I studied the effect modification role of type 2 diabetes and sex differences in the adverse association of blood uric acid levels with all-cause and specific cause mortality.

REFERENCES

1. 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.
2. Guthrie RA, Guthrie DW. Pathophysiology of diabetes mellitus. Crit Care Nurs Q 2004;27:113-25.
3. Emerging Risk Factors C, Sarwar N, Gao P, et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet 2010;375:2215-22.
4. Roglic G. WHO Global report on diabetes: A summary. International Journal of Noncommunicable Diseases 2016;1:3-8.
5. Organization GWH. World health statistics 2018: monitoring health for the SDGs, sustainable development goals. 2018 ed2018.
6. Bommer C, Heesemann E, Sagalova V, et al. The global economic burden of diabetes in adults aged 20-79 years: a cost-of-illness study. Lancet Diabetes Endocrinol 2017;5:423-30.
7. Collaborators GBDCoD. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1736-88.
8. NCD- NRFC. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants (vol 387, pg 1513, 2016). Lancet 2017;389:E2-E.
9. Huebschmann AG, Huxley RR, Kohrt WM, Zeitler P, Regensteiner JG, Reusch JEB. Sex differences in the burden of type 2 diabetes and cardiovascular risk across the life course. Diabetologia 2019;62:1761-72.
10. Sattar N. Gender aspects in type 2 diabetes mellitus and cardiometabolic risk. Best Pract Res Clin Endocrinol Metab 2013;27:501-7.
11. Ferguson LD, Ntuk UE, Celis-Morales C, et al. Men across a range of ethnicities have a higher prevalence of diabetes: findings from a cross-sectional study of 500 000 UK Biobank participants. Diabet Med 2018;35:270-6.
12. Hilawe EH, Yatsuya H, Kawaguchi L, Aoyama A. Differences by sex in the prevalence of diabetes mellitus, impaired fasting glycaemia and impaired glucose tolerance in sub-Saharan Africa: a systematic review and meta-analysis. Bull World Health Organ 2013;91:671-82D.
13. Lipscombe LL, Hux JE. Trends in diabetes prevalence, incidence, and mortality in Ontario, Canada 1995-2005: a population-based study. Lancet 2007;369:750-6.
14. Mehta LS, Beckie TM, DeVon HA, et al. Acute Myocardial Infarction in Women A Scientific Statement From the American Heart Association. Circulation 2016;133:916-47.
15. Meigs JB, Cupples LA, Wilson PW. Parental transmission of type 2 diabetes: the Framingham Offspring Study. Diabetes 2000;49:2201-7.
16. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 2010;42:579-89.
17. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 2012;44:981-90.

18. 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.
19. Wilson PWF, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB. Prediction of incident diabetes mellitus in middle-aged adults - The Framingham Offspring Study. Archives of Internal Medicine 2007;167:1068-74.
20. Lyssenko V, Laakso M. Genetic Screening for the Risk of Type 2 Diabetes Worthless or valuable? Diabetes Care 2013;36:S120-S6.
21. Meigs JB, Shrader P, Sullivan LM, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med 2008;359:2208-19.
22. Berdasco M, Esteller M. Clinical epigenetics: seizing opportunities for translation. Nat Rev Genet 2019;20:109-27.
23. Murrell A, Rakyan VK, Beck S. From genome to epigenome. Hum Mol Genet 2005;14 Spec No 1:R3-R10.
24. Youngson NA, Whitelaw E. Transgenerational epigenetic effects. Annual Review of Genomics & Human Genetics 2008;9:233-57.
25. Faulk C, Dolinoy DC. Timing is everything The when and how of environmentally induced changes in the epigenome of animals. Epigenetics 2011;6:791-7.
26. Biswas S, Rao CM. Epigenetics in cancer: Fundamentals and Beyond. Pharmacology & Therapeutics 2017;173:118-34.
27. Muka T, Nano J, Voortman T, et al. The role of global and regional DNA methylation and histone modifications in glycemic traits and type 2 diabetes: A systematic review. Nutr Metab Cardiovasc Dis 2016;26:553-66.
28. Muka T, Nano J, Voortman T, et al. The role of global and regional DNA methylation and histone modifications in glycemic traits and type 2 diabetes: A systematic review. Nutr Metab Cardiovasc Dis 2016;26:553-66.
29. Weber M, Schubeler D. Genomic patterns of DNA methylation: targets and function of an epigenetic mark. Curr Opin Cell Biol 2007;19:273-80.
30. Turek-Plewa J, Jagodzinski PP. The role of mammalian DNA methyltransferases in the regulation of gene expression. Cell Mol Biol Lett 2005;10:631-47.
31. Dayeh T, Tuomi T, Almgren P, et al. DNA methylation of loci within ABCG1 and PHOSPHO1 in blood DNA is associated with future type 2 diabetes risk. Epigenetics 2016;11:482-8.
32. Dayeh T, Volkov P, Salo S, et al. Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. PLoS Genet 2014;10:e1004160.
33. Gillberg L, Jacobsen SC, Ribel-Madsen R, et al. Does DNA methylation of PPARGC1A influence insulin action in first degree relatives of patients with type 2 diabetes? PLoS One 2013;8:e58384.
34. Kulkarni H, Kos MZ, Neary J, et al. Novel epigenetic determinants of type 2 diabetes in Mexican-American families. Hum Mol Genet 2015;24:5330-44.
35. Ling C, Del Guerra S, Lupi R, et al. Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion. Diabetologia 2008;51:615-22.
36. Ling C, Groop L. Epigenetics: a molecular link between environmental factors and type 2 diabetes. Diabetes 2009;58:2718-25.

37. Liu ZH, Chen LL, Deng XL, et al. Methylation status of CpG sites in the MCP-1 promoter is correlated to serum MCP-1 in Type 2 diabetes. J Endocrinol Invest 2012;35:585-9.
38. Zou L, Yan S, Guan X, Pan Y, Qu X. Hypermethylation of the PRKCZ Gene in Type 2 Diabetes Mellitus. J Diabetes Res 2013;2013:721493.
39. Hidalgo B, Irvin MR, Sha J, et al. Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network study. Diabetes 2014;63:801-7.
40. Campbell RK, White JR, Jr. Insulin therapy in type 2 diabetes. J Am Pharm Assoc (Wash) 2002;42:602-11.
41. Abbasi A, Peelen LM, Corpeleijn E, et al. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. BMJ 2012;345:e5900.
42. Buijsse B, Simmons RK, Griffin SJ, Schulze MB. Risk assessment tools for identifying individuals at risk of developing type 2 diabetes. Epidemiol Rev 2011;33:46-62.
43. Gillberg L, Ling C. The potential use of DNA methylation biomarkers to identify risk and progression of type 2 diabetes. Front Endocrinol (Lausanne) 2015;6:43.
44. Herder C, Kowall B, Tabak AG, Rathmann W. The potential of novel biomarkers to improve risk prediction of type 2 diabetes. Diabetologia 2014;57:16-29.
45. Asmat U, Abad K, Ismail K. Diabetes mellitus and oxidative stress-A concise review. Saudi Pharm J 2016;24:547-53.
46. Casula M, Mozzanica F, Scotti L, et al. Statin use and risk of new-onset diabetes: A meta-analysis of observational studies. Nutr Metab Cardiovasc Dis 2017;27:396-406.
47. Barnard K, Peveler RC, Holt RI. Antidepressant medication as a risk factor for type 2 diabetes and impaired glucose regulation: systematic review. Diabetes Care 2013;36:3337-45.
48. Gress TW, Nieto FJ, Shahar E, Wofford MR, Brancati FL. Hypertension and antihypertensive therapy as risk factors for type 2 diabetes mellitus. Atherosclerosis Risk in Communities Study. N Engl J Med 2000;342:905-12.
49. Penckofer S, Schwertz D, Florczak K. Oxidative stress and cardiovascular disease in type 2 diabetes: the role of antioxidants and pro-oxidants. J Cardiovasc Nurs 2002;16:68-85.
50. Donath MY, Shoelson SE. Type 2 diabetes as an inflammatory disease. Nat Rev Immunol 2011;11:98-107.
51. Hussain T, Tan B, Yin Y, Blachier F, Tossou MC, Rahu N. Oxidative Stress and Inflammation: What Polyphenols Can Do for Us? Oxid Med Cell Longev 2016;2016:7432797.
52. Matsuoka T, Kajimoto Y, Watada H, et al. Glycation-dependent, reactive oxygen species-mediated suppression of the insulin gene promoter activity in HIT cells. J Clin Invest 1997;99:144-50.
53. Rudich A, Tirosh A, Potashnik R, Hemi R, Kanety H, Bashan N. Prolonged oxidative stress impairs insulin-induced GLUT4 translocation in 3T3-L1 adipocytes. Diabetes 1998;47:1562-9.
54. Furukawa S, Fujita T, Shimabukuro M, et al. Increased oxidative stress in obesity and its impact on metabolic syndrome. J Clin Invest 2004;114:1752-61.
55. Ohashi K, Shibata R, Murohara T, Ouchi N. Role of anti-inflammatory adipokines in obesity-related diseases. Trends Endocrinol Metab 2014;25:348-55.

56. Leal Vde O, Mafra D. Adipokines in obesity. Clin Chim Acta 2013;419:87-94.
57. van Dielen FM, van't Veer C, Schols AM, Soeters PB, Buurman WA, Greve JW. Increased leptin concentrations correlate with increased concentrations of inflammatory markers in morbidly obese individuals. Int J Obes Relat Metab Disord 2001;25:1759-66.
58. Kanmani S, Kwon M, Shin MK, Kim MK. Association of C-Reactive Protein with Risk of Developing Type 2 Diabetes Mellitus, and Role of Obesity and Hypertension: A Large Population-Based Korean Cohort Study. Sci Rep 2019;9:4573.
59. Lainampetch J, Panprathip P, Phosat C, et al. Association of Tumor Necrosis Factor Alpha, Interleukin 6, and C-Reactive Protein with the Risk of Developing Type 2 Diabetes: A Retrospective Cohort Study of Rural Thais. J Diabetes Res 2019;2019:9051929.
60. Liu C, Feng X, Li Q, Wang Y, Li Q, Hua M. Adiponectin, TNF-alpha and inflammatory cytokines and risk of type 2 diabetes: A systematic review and meta-analysis. Cytokine 2016;86:100-9.
61. Odegaard AO, Jacobs DR, Jr., Sanchez OA, Goff DC, Jr., Reiner AP, Gross MD. Oxidative stress, inflammation, endothelial dysfunction and incidence of type 2 diabetes. Cardiovasc Diabetol 2016;15:51.
62. Sung KC, Ryu S, Sung JW, et al. Inflammation in the Prediction of Type 2 Diabetes and Hypertension in Healthy Adults. Arch Med Res 2017;48:535-45.
63. Brahimaj A, Ligthart S, Ghanbari M, et al. Novel inflammatory markers for incident pre-diabetes and type 2 diabetes: the Rotterdam Study. Eur J Epidemiol 2017;32:217-26.
64. Pollack RM, Donath MY, LeRoith D, Leibowitz G. Anti-inflammatory Agents in the Treatment of Diabetes and Its Vascular Complications. Diabetes Care 2016;39 Suppl 2:S244-52.
65. Tsalamandris S, Antonopoulos AS, Oikonomou E, et al. The Role of Inflammation in Diabetes: Current Concepts and Future Perspectives. Eur Cardiol 2019;14:50-9.
66. Donath MY, Dalmas E, Sauter NS, Boni-Schnetzler M. Inflammation in obesity and diabetes: islet dysfunction and therapeutic opportunity. Cell Metab 2013;17:860-72.
67. Thakker D, Nair S, Pagada A, Jamdade V, Malik A. Statin use and the risk of developing diabetes: a network meta-analysis. Pharmacoepidemiol Drug Saf 2016;25:1131-49.
68. Cholesterol Treatment Trialists C, Baigent C, Blackwell L, et al. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. Lancet 2010;376:1670-81.
69. Cederberg H, Stancakova A, Yaluri N, Modi S, Kuusisto J, Laakso M. Increased risk of diabetes with statin treatment is associated with impaired insulin sensitivity and insulin secretion: a 6 year follow-up study of the METSIM cohort. Diabetologia 2015;58:1109-17.
70. Chen YH, Chen YC, Liu CS, Hsieh MC. The Different Effects of Atorvastatin and Pravastatin on Cell Death and PARP Activity in Pancreatic NIT-1 Cells. J Diabetes Res 2016;2016:1828071.
71. Lorza-Gil E, Salerno AG, Wanschel AC, et al. Chronic use of pravastatin reduces insulin exocytosis and increases beta-cell death in hypercholesterolemic mice. Toxicology 2016;344-346:42-52.
72. Allen SC, Mamotte CDS. Pleiotropic and Adverse Effects of Statins-Do Epigenetics Play a Role? J Pharmacol Exp Ther 2017;362:319-26.

73. Ley SH, Hamdy O, Mohan V, Hu FB. Prevention and management of type 2 diabetes: dietary components and nutritional strategies. Lancet 2014;383:1999-2007.
74. Mancini FR, Affret A, Dow C, et al. Dietary antioxidant capacity and risk of type 2 diabetes in the large prospective E3N-EPIC cohort. Diabetologia 2018;61:308-16.
75. Cao H, Ou J, Chen L, et al. Dietary polyphenols and type 2 diabetes: Human Study and Clinical Trial. Crit Rev Food Sci Nutr 2019;59:3371-9.
76. Guasch-Ferre M, Merino J, Sun Q, Fito M, Salas-Salvado J. Dietary Polyphenols, Mediterranean Diet, Prediabetes, and Type 2 Diabetes: A Narrative Review of the Evidence. Oxid Med Cell Longev 2017;2017:6723931.
77. Saura-Calixto F, Goni I. Antioxidant capacity of the Spanish Mediterranean diet. Food Chem 2006;94:442-7.
78. Ludwig IA, Clifford MN, Lean ME, Ashihara H, Crozier A. Coffee: biochemistry and potential impact on health. Food Funct 2014;5:1695-717.
79. Hang D, Kvaerner AS, Ma W, et al. Coffee consumption and plasma biomarkers of metabolic and inflammatory pathways in US health professionals. Am J Clin Nutr 2019;109:635-47.
80. Carlstrom M, Larsson SC. Coffee consumption and reduced risk of developing type 2 diabetes: a systematic review with meta-analysis. Nutr Rev 2018;76:395-417.
81. Poole R, Kennedy OJ, Roderick P, Fallowfield JA, Hayes PC, Parkes J. Coffee consumption and health: umbrella review of meta-analyses of multiple health outcomes. BMJ 2017;359:j5024.
82. Impellizzeri FM, Bizzini M. Systematic review and meta-analysis: a primer. Int J Sports Phys Ther 2012;7:493-503.
83. Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. Nat Rev Genet 2006;7:812-20.
84. Ikram MA, Brusselle G, Ghanbari M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. Eur J Epidemiol 2020;35:483-517.
85. Collins R. What makes UK Biobank special? Lancet 2012;379:1173-4.
86. Palmer LJ. UK Biobank: bank on it. Lancet 2007;369:1980-2.
87. Rathmann W, Haastert B, Icks A, et al. High prevalence of undiagnosed diabetes mellitus in Southern Germany: target populations for efficient screening. The KORA survey 2000. Diabetologia 2003;46:182-9.
88. Chahal NS, Lim TK, Jain P, Chambers JC, Kooner JS, Senior R. Does subclinical atherosclerosis burden identify the increased risk of cardiovascular disease mortality among United Kingdom Indian Asians? A population study. Am Heart J 2011;162:460-6.
89. Raum E, Rothenbacher D, Low M, Stegmaier C, Ziegler H, Brenner H. Changes of cardiovascular risk factors and their implications in subsequent birth cohorts of older adults in Germany: a life course approach. Eur J Cardiovasc Prev Rehabil 2007;14:809-14.
90. Volzke H, Alte D, Schmidt CO, et al. Cohort profile: the study of health in Pomerania. Int J Epidemiol 2011;40:294-307.

Chapter 2.

Epigenetics of type 2 diabetes

Chapter 2.1.

Chromatin landscape and epigenetic biomarkers for clinical diagnosis and prognosis of type 2 diabetes mellitus

Carolina Ochoa-Rosales*, Eralda Asllanaj*, Glisic Marija, Jana Nano, Taulant Muka, Oscar Franco.

Chapter 11 - Chromatin landscape and epigenetic biomarkers for clinical diagnosis and prognosis of type 2 diabetes mellitus. *In: Sharma S, ed. Prognostic Epigenetics: Academic Press; 2019:289-324.*

ABSTRACT

Type 2 diabetes and its accompanying complications constitute a major health burden worldwide, which can be partly attributed to the interplay between genetics and environments. Extensive research over the last decades has shown that our genome is not the only determinant of disease risk. Epigenetic marks induced by lifestyle and environmental factors are associated with altered gene expression patterns in important tissues, leading to altered susceptibility to disease later in life. Hence, the identification of epigenetic biomarkers unfolds the possibility for a novel personalized disease prevention strategy and at the same time holds the potential to be a promising prognostic tool for diabetes. So far, evidence on the predictive value of epigenetics in diabetes management is very limited. Unlike in cancer pathology, where examples of important epigenetic tools are now widely used in clinical practice as predictive/diagnostic biomarkers, for complex pathophysiological diseases such as diabetes, this still remains a challenge. These topics are discussed extensively in this chapter.

1. INTRODUCTION

Diabetes has become a major public health problem with type 2 diabetes (T2D) being the predominant condition that accounts for at least 90% of the cases [1]. According to the World Health Organization (WHO) reports in 2012, the estimated number of people living with T2D by 2030 would have been 366 million [2]. However, these numbers are expected to be higher since until 2017, 425 million people were reported having diabetes [3] overcoming all predictions [4].

T2D is characterized by insulin deficiency and insulin resistance [5], and its major complications comprise macrovascular, microvascular and neurologic changes which can lead to organ damage including heart, kidneys, eyes, feet and nerves [5]. According to the American Diabetes Association (ADA), diabetes diagnosis is defined as: fasting plasma glucose ≥ 126 mg/dL (≥ 7.0 mmol/L) where fasting is defined as no caloric intake for at least 8 hours or 2-h plasma glucose ≥ 200 mg/dL (≥ 11.1 mmol/L) during a 75-g oral glucose tolerance test (OGTT, the test should be performed as described by the WHO, using a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water) or A1C ≥ 6.5% (≥ 48 mmol/mol) or in a patient with classic symptoms of hyperglycaemia or hyperglycaemic crisis, a random plasma glucose ≥ 200 mg/dL (≥ 11.1 mmol/L) [6]. These definitions go in line with the current WHO diagnostic criteria, except for the glycosylated haemoglobin (HbA1c) test, which remains controversial[7].

Studies investigating the aetiology of T2D have been primarily focused into the genetic determinants of the disease. Recent evidence shows epigenetics could be a major player in the pathophysiology of the disease, through which environmental and lifestyle factors could affect T2D pathogenesis (Figure 1) [8]. Lifestyle and other environmental factors could lead to changes in DNA methylation and histone modifications, which on the other hand, might affect the development of pancreatic β cells and the function of insulin secretion, contributing to the decline of insulin sensitivity resulting in the occurrence of T2D [8]. Animal and human studies investigating the genome-wide maps of epigenetic markers using islet tissue have provided a reliable resource for understanding the importance of the epigenetic mechanisms in T2D susceptibility [9].

In clinical practice, biomarkers are used routinely to identify individuals at risk and are of great importance in disease diagnosis. For T2D, fasting blood glucose, HbA1c and 2-hours oral glucose are commonly used, but they come with some drawbacks. Blood glucose levels do not reflect the impaired b-cell function or insulin resistance [10];

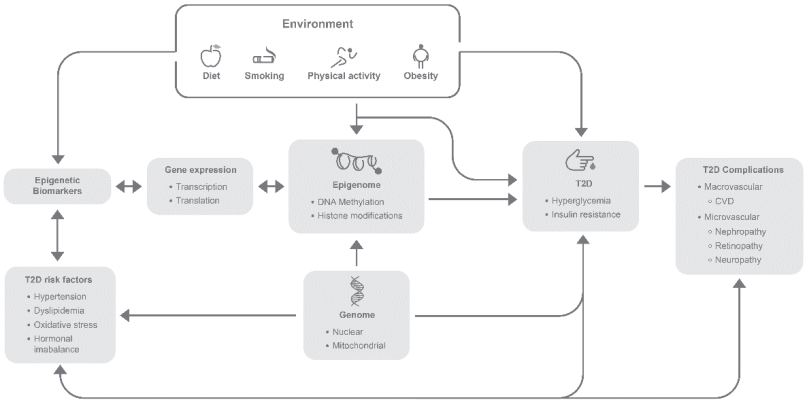


Figure 1. A conceptual model linking epigenomics to T2D and T2D complications.
Epigenomics represents a critical link between genomic coding and phenotype expression that is influenced by both underlying genetic and environmental factors. Epigenetic biomarkers which can be influenced by T2D risk factors and remodeled epigenetic patterns may contribute to the development of T2D and its complications.

the optimal value for HbA1c for diagnosis of prediabetes state still remain controversial [11 12], whereas the 2-hours oral glucose tolerance test is a time consuming procedure. It is not known whether the deterioration in glucose tolerance and beta-cell function is linear or whether there is an accelerated loss of function at some point prior to the onset of diabetes [7]. Therefore, the early identification of high-risk individuals demands novel biomarkers that adequately account for the inter-individual variance in the different pathological mechanisms underlying impaired fasting glucose (IFG) and impaired glucose tolerance (IGT), each of which have distinct progression patterns towards diabetes [13]. Moreover, the different kind of complications from diabetes might leave different methylation signatures which could result in type-specific and predictive signatures with potential use as future prognostic biomarkers for T2D. Further, considering the rapidly increase in incidence and prevalence of T2D, it has become relevant to extend current knowledge and discover new biomarkers that could be identified and/or monitored during the diagnosis and progression of the disease.

In this chapter, the topics of epigenetic alterations, particularly DNA methylation and histone modifications and the importance of epigenetic biomarkers for risk prediction, diagnosis and prognosis of T2D will be discussed.

2. EPIGENETIC ALTERATIONS INVOLVED IN GLUCOSE HOMEOSTASIS AND INSULIN METABOLISM

The association between glucose homeostasis related traits and DNA methylation has been assessed through different approaches, such as global DNA methylation assessment, DNA methylation in candidate genes, and Epigenome-Wide Association Studies (EWAS).

Global DNA methylation refers to the overall level of 5-methylcitosine in the genome, expressed as percentage of total cytosine. Repetitive and transposable elements, such as LINE-1 and Alu, represent a large portion of the human genome and contain much of the CpG methylation found in normal human postnatal somatic tissues [14]. Given the existing correlation of methylation at such elements with the total genomic methylation content, they are considered surrogate markers for global genome methylation [14].

In a candidate gene methylation approach, the association is evaluated only for specific genes of interest that have been selected based on their possible role in the phenotype of interest. Therefore, the methylation level is assessed only in specific regions of the DNA.

EWAS, scan genome-wide epigenetic variants, such as DNA methylation, which might be associated with the phenotype of interest. EWAS are mainly performed using microarrays, which profile the methylation level of thousands of CpG islands in the genome, surveying multiple samples.

Information about the function of the genes mentioned in this chapter that have been studied in relation with T2D and glycaemic traits can be found in Table 1.

2.1. Glucose homeostasis

Epigenetic alterations can have great influence on islet cells and glucose homeostasis that can alter their pathophysiological processes and consequently result in T2D [15].

2.1.1. DNA methylation

Different studies have investigated the association between global DNA methylation and glucose levels, reporting inconsistent results [16-18]. Increased levels of plasma glucose were associated with higher methylation levels in LINE-1 when assessed in adipose tissue, blood or skeletal muscle [16 17 19]. However, one study showed no association or an inverse association between LINE -1 methylation or other markers

of global DNA methylation and glucose levels assessed in B and NK lymphocytes human cells [18]. This stresses the relevance of using cell type-specific assays when investigating epigenetic signatures in clinical tissue samples especially those characterized by a high heterogeneity in cell types frequency and phenotype, such as blood.

Candidate gene studies have revealed lower methylation levels of *GIPR* gene and *PPARGC1A* gene in blood and skeletal muscle [20 21]. Both genes are believed to contribute to improve insulin sensitivity, mitochondrial biogenesis and browning of white adipose tissue. In another study, the authors reported that high glucose levels affect human pancreatic islet gene expression and several of these genes also exhibit epigenetic changes [22]. This might contribute to the impaired insulin secretion seen in T2D. Moreover, one study reported that increased levels of plasma glucose might be associated with higher methylation levels of LY86 gene in blood, which has been suggested to play a role in inflammation, obesity and insulin resistance [23]. Volkmar et al. have investigated DNA methylation in human pancreatic islets by exposing the pancreatic cells from nondiabetic donors to high glucose levels [9]. The study reported a non-significant association between DNA methylation and the 16 CpG sites tested, concluding that the methylated changes in the islets from T2D patients would not likely be a cause of hyperglycaemia [9].

Furthermore, studies conducted using placenta tissue and cord blood have yielded interesting results between methylation levels and fasting glucose. Lower DNA methylation levels of *ADIPOQ, LPL, IGF1R* and *IGFBP3* on the fetal side of the placenta were associated with higher maternal 2-h post oral glucose tolerance test levels during pregnancy, although the association did not remain significant with 2 h post-oral glucose tolerance test levels [24]. Also, the maternal gestational glucose levels were positively associated with placental DNA methylation, and negatively associated with cord blood DNA methylation of the *PPARGC1A* promoter in a CpG site-specific manner. The researchers concluded that epigenetic alteration of the *PPAGRC1A* promoter may be one of the potential mechanisms underlying the metabolic programming in offspring exposed to intrauterine hyperglycaemia [25]. Another study investigating whether epigenetic dysregulations of the insulin-like growth factor system in placenta were exposed to maternal impaired glucose tolerance, confirmed their hypothesis [26]. Also, in this study, maternal glucose 2 h post oral glucose tolerance test and fasting glucose at the second trimester of pregnancy were negatively correlated with *GF1R-I4* (7 CpGs) and *IGFBP3-L1* DNA methylation levels [26]. Both *GF1R-I4* and *IGFBP3-L1* are important genes in foetal metabolic programming and impaired glucose tolerance.

Limited evidence exists on EWAS and glucose metabolism. Also, the existing evidence so far is inconclusive and inconsistent with studies reporting no association [27] and another reporting a positive association between epigenome-wide DNA methylation levels and fasting glucose [28]. Using whole blood samples from a population-based prospective study, one study recently reported 6 CpG sites related to fasting glucose and 2-hour glucose, independent of age, sex, smoking, and estimated white blood cell proportions [26]. Moreover this study showed that effect strengths were reduced on average by around 30% after adjustment for BMI, suggesting an influence of BMI on the investigated phenotypes [26]. The findings provide evidence for the first time that DNA methylation may be associated with glucose metabolism, a relationship which can be measured in DNA isolated from whole blood.

2.1.2. Histone modifications

Histone modifications may also play a pivotal role in glucose metabolism, but this is an understudied research topic [29]. Studies have shown that *TXNIP* gene might be important in glucose metabolism, especially in diabetes-related phenotypes [30 31]. *TXNIP* is a key component of pancreatic β-cell biology, nutrient sensing, energy metabolism, and regulation of cellular redox [30 31]. Moreover, *TXNIP* expression is highly induced by glucose through activation of the carbohydrate response element-binding protein, which binds the *TXNIP* promoter, making it an attractive target for diabetes therapy. Previous studies have identified several critical transcription factors, enzymes important in histone activation and acetylation, like the ChREBP and p300, as the specific chromatin modification mediating this glucose-induced transcription of beta cell *TXNIP* [31]. Recently, another study published similar results confirming the findings [32]. They found that the glucose-induced *TXNIP* gene expression is greatly reduced by p300 silencing, and Ep300 cells are protected from high glucose-induced cell death and have elevated insulin secretion [32]. In the current study, elevated levels of EP300 and *TXNIP* gene expression in human diabetic islets were correlated with reduced glucose-stimulated *TXNIP* genes expression [32]. These data provide evidence that histone acetylation could be a key regulator of glucose-induced increase in *TXNIP* gene expression and thereby glucotoxicity-induced apoptosis.

2.2. Insulin metabolism

A common feature of T2D that affects the liver and the peripheral tissues is insulin resistance (IR). The most relevant tissues that develop insulin resistance are liver cells, skeletal muscle, and adipose tissue [33]. Impaired response to insulin fails to clear the blood stream from glucose, and additionally, stimulates the secretion of adipokines from the adipose tissue which may further negatively affect the whole body glucose homeostasis [33].

2.2.1. DNA Methylation

Several studies have investigated the association between global DNA methylation and insulin metabolism, focusing on fasting plasma insulin levels, insulin secretion [19], and insulin resistance as measured by homeostatic model assessment [34 35]. The studies on insulin secretion and insulin resistance did not report significant associations between insulin metabolism and global DNA methylation. However, one study reported an interaction of global DNA methylation with circulating folate concentrations in relation to insulin resistance [34]. The authors found that a lower degree of methylation and lower plasma folate concentrations were associated with higher insulin resistance [34]. Folate metabolism is linked to phenotypic changes through DNA methylation by the knowledge that folate, a coenzyme of one-carbon metabolism, is directly involved in methyl group transfer for DNA methylation, making them important epigenetic players. Another study assessed global DNA methylation as a percentage of 20-deoxycytidine plus 5-methyl-deoxy-cytidine (5mdC) in genomic DNA and reported a positive association between insulin levels and global DNA methylation assessed in lymphocyte B cells but no association in natural killer cells [36]. Zhao et al. assessed global DNA methylation in Alu elements in peripheral blood leukocytes, which was quantified by bisulphite pyrosequencing [35]. The study showed a positive association with insulin resistance and reported that a 10% increase in mean Alu methylation was associated with an increase of 4.55 units in homeostatic model assessment [35].

Many candidate gene studies have examined methylation sites in or near known candidate genes in relation to plasma insulin, insulin expression and insulin resistance [15]. Most of them reported a positive correlation between plasma insulin and methylation at *PPARGC1A* in the liver and at *HTR2A* and LY86 in blood cells. Lower levels of methylation at *PPARGC1A* were identified in skeletal muscle and lower levels of methylation were identified at the insulin promoter gene associated with increased levels of plasma insulin or mRNA insulin expression [15]. Moreover, inverse associations were found between insulin resistance and the degree of methylation of *TFAM* and *GIPR3* genes in blood cells and *PPARGC1A* gene in skeletal muscle [15]. Furthermore, studies in pregnant women, reported a negative association of methylation levels of the maternal side of placenta of *ADIPOQ* gene with insulin resistance [37]. While another study reported a positive correlation between the methylation of *IGFBP3* with fasting insulin levels and insulin resistance [26].

Further, a few EWAS have been performed in regard to insulin metabolism [38 39]. Hidalgo et al. reported a significant association between the methylation of a CpG site in *ABCG1* gene on chromosome 21 with insulin and homeostatic model assess-

ment-IR, suggesting that methylation of the CpG site within *ABCG1* merits further evaluation as a novel disease risk marker [39].

The majority of the above mentioned genes are reported to have important functions in metabolic traits and have been associated to insulin metabolism through different biological mechanisms. The *HTR2C* gene is involved in energy expenditure and polymorphisms in this gene coding for many receptors are thought to influence insulin homeostasis [40]. While, *PPARGC1A* upregulates transcription of genes involved in mitochondrial oxidative metabolism and biogenesis as well as skeletal muscle glucose transport. Because mitochondrial defects have been associated with peripheral insulin resistance in healthy subjects it has been suggested that reduced *PPARGC1A* expression in skeletal muscle may be a primary feature of insulin resistance [41]. Furthermore, *PPARGC1A* is involved in biological functions with implications in insulin action including protection against oxidative stress, formation of muscle fiber types as well as regulation of microvascular flow [41 42]. Moreover, there is evidence linking the LY86, TFAM and GIPR3 genes to insulin resistance, mainly their respective encoded proteins that play crucial roles in the pathophysiological regulation of inflammation and insulin resistance [42].

2.2.2. Histone modifications

The evidence pertaining the possible role of histone modifications in insulin metabolism is also very limited. One study investigated the effects of insulin on alterations in post-translational modifications of histone H3 in L6 myoblasts under a hyperglycaemic condition [43]. The authors demonstrated that insulin induced intracellular generated oxidative stress is involved in modulating multiple histone modifications under hyperglycaemic conditions [43]. Their results also revealed that phosphorylation of histone H3 at Ser 10 was independent of known histone kinases and suggest the role of serine/threonine phosphatase in modulating insulin signalling, suggesting a possible role of phosphatase and its inhibitor in diabetes [43].

3. EPIGENETIC ALTERATIONS IN DIABETES

T2D is a complex disease, product of the interaction of genetic and environmental factors [44] (Figure 1). Epigenetic mechanisms could underlie the connection between environmental exposures and pathology of T2D [45]. For this reason, in recent years, it has been of great interest to study DNA methylation and histone modifications in relation to T2D.

3.1. DNA methylation

When comparing diabetic versus non-diabetic individuals, no difference in global DNA methylation has been reported in overall peripheral blood [46 47], lymphocytes or monocytes populations assessed separately [18], pancreatic islets [9], omental visceral adipose tissue and subcutaneous adipose tissue [17]. Also studies that used skeletal muscle and subcutaneous adipose tissue from monozygotic twins discordant for T2D did not report any differences in global DNA methylation [16]. However, significant differences in the degree of global DNA methylation have also been reported. Luttmer et al. reported hypomethylation in blood samples from T2D patients [48], whereas Simar et al. reported an increased degree of global DNA methylation specifically in B-cells and natural killer cells from T2D donors [18].

In a candidate gene approach, DNA methylation at several selected genes has been investigated in different tissues, comparing diabetic and non-diabetic donors.

The genes that have been reported to have higher methylation levels in T2D patients are: *IGFBP7* [49], *IGFBP1* [50], *TLR2* [51], *SLC30A8* [52], *GCK* [53], *PRKCZ* [54], *CTGF* [46] and leptin gene in peripheral blood; *PPARGC1A* [55], *PDX-1* [56], insulin promoter gene [57] and *GLP1R* in pancreatic islets [56] and *APN* in adipose tissue [58].

On the other hand, in diabetic donors, lower levels of methylation have been found at genes: *GIPR* [59], *CAMK1D, CRY2, CALM2* [60], *MCP1* [61], *TLR4, FFAR3* [51], *PP2Ac* [62] and *CTGF* [46] in the in peripheral blood samples; *UBASH3A* in B-cells [18] and *PDK4* in skeletal muscle tissue [63]. Additionally, differential methylation between type 2 diabetic patients and matched controls has been found at *TCF7L2* [64].

Further, no clear difference was observed for genes IRS-1 in the peripheral blood [65]; *GADPH, TFAM* and *TRIM3* in B-cells [18]; *GLP1R* in pancreatic islets [66] and *PPARGC1A* in the skeletal muscle [67].

When comparing monozygotic twins discordant for T2D, the genes that were hypermethylated in the diabetic subjects are: *HNF4A, KLF11, DUSP9, HHEX* and *PPARGC1A* in muscle tissue and *CIDEC, HNF4A, ADCY5, CDKN2B, IDE, KCNQ1, MTNR1B* and *TSPAN8* in subcutaneous adipose tissue. Whereas the hypomethylated genes found in the diabetic twins are *CDKN2A, KCNQ1* and *SLC30A8* in muscle tissue and *CAV1, CDKN2A, DUSP9, IRS1* and *WFS1* in subcutaneous adipose tissue. However, after adjustment for multiple testing only two methylation sites, at *CDKN2A* and *HNF4A* genes, in subcutaneous adipose tissue, remained significant [16].

EWAS have raised and evolved as the technology for epigenetics research gradually develops. Progressively, scientists have had more access to methylation arrays with a higher number of methylation probes and new sequencing techniques that have allowed researches to perform genome-wide studies in shorter times. Thus, genome-wide DNA methylation profiling to find associations with T2D has allowed identifying new genes differentially methylated that might provide new insights in the pathogenesis of the disease and the search for biomarkers. To date, EWAS in association with T2D have been performed in peripheral blood, pancreatic islets, skeletal muscle and subcutaneous adipose tissue.

For the study of methylation in pancreatic islets, a Human Methylation 27 BeadChip array was used. It was observed that 276 CpGs affiliated to promotors of 254 genes were displaying differentially methylated sites in tissue from T2D donors compared with non-diabetic controls [9]. Further, a Human Methylation 450 K BeadChip array was utilized to carry out the same approach. As a result, 1649 CpGs, annotated to 853 genes, were reported to have sites of differential methylation in diabetic islets [68]. In skeletal muscle tissue, methylation levels were compared between monozygotic twins discordant for T2D using a Human Methylation 27 Bead Chip array. The test identified one CpG, annotated at IL8, to be differentially methylated [16]. A similar analysis was performed in subcutaneous adipose tissue, revealing that in the diabetic donor, CpGs annotated to *ZNF668*, *HSPA2, C8orf31, CD320, SFT2D3, TWIST1* and *MYO5A* genes had methylation levels significantly different from the non-diabetic twins [16]. DNA methylation studies in blood samples reports dissimilar results, depending on the methylation assessment and multiple testing correction methods. When Human Methylation 450 K Bead Chip array was used, one study reported 51 significant CpGs associated to T2D (correction for multiple tests of FDR < 5%) [69], while other authors observed 5 CpGs associated to incident T2D (correction for multiple testing of $P < 5 \times 10^{-7}$) [30]. A different study used Methylated DNA immunoprecipitation sequencing (MeDIP-seq). It identified as the strongest signal, a differential methylated region at the promoter of *MALT1* (FDR < 5%). A study using the affimetrix SNP6 microarray with posterior in-deep sequencing of the tip-ranking regions showed that 13 out of 93 CpG sites exhibited differences between T2D patients and controls [70]. The researchers found that, among those sites, a methylation site located in the first intron of *FTO* was significantly hypomethylated in blood samples from diabetic subjects [70].

3.2. Histone modifications

Histone changes in acetylation or methylation patterns might induce modifications in chromatin structure and, as a consequence, it may promote dysregulated gene transcription and disease progression [71 72].

A limited number of studies have explored the association of histone modifications with T2D. Histone methyltransferase Set7 is an enzyme involved in histones methylation. Using peripheral blood mononuclear cells from diabetic patients, it was found a Set7-dependent monomethylation of lysine 4 of histone 3 (H3K4m1) on *NK-k*β p65 promoter, along with an upregulation of the enzyme [73]. Another study observed an increased level of histone H3 lysine 9 dimethylation (H3K9me2) around *IL-1A* promoter and *PTEN* coding regions in circulating monocytes from T2D patients relative to non-T2D controls [74]. A different study investigating histone acetylation found that histone 3 (H3) acetylation at *TNF-*α promoter and *COX-2* promoter was increased in peripheral blood monocytes from type 2 diabetics, compared to controls [75].

4. EPIGENETIC MECHANISMS AS BIOMARKERS FOR RISK PREDICTION

When T2D is not well managed, it can lead to health complications in different organs of the body, increasing the risk of disability and premature death [76]. Direct and indirect costs of T2D impose a large burden that impacts the economy of the country's [77]. These adverse consequences make the development of prevention strategies highly relevant, like the early identification of individuals at risk of developing T2D.

Recently, next to the traditional prevention methods, a novel approach to assess individuals risk is developed, by using statistical models that are able to predict future onset of the disease, based on epigenetic markers. The performance of a prediction model is evaluated by means of the area under the receiver operating characteristic (ROC) curve (AUC), being an AUC value of 1.0 a perfect discrimination of the outcome to be predicted. So far, the models using conventional components, such as anthropometric measurements, family history of diabetes, lifestyle factors and biomarkers like glucose, insulin and lipid levels, are able to predict T2D with an AUC that ranges from 0.7 to 0.9 [78]. However, a large number of predictor factors and measurements of different nature are required, and models do not perform similarly in different populations. As Genome-Wide Association Studies (GWAS) have been developed, genetic markers associated with T2D, glucose, insulin and

insulin resistance, have been included as predictors in prediction models. Unfortunately, the associated genetic variants have not improved the performance of the overall models [79 80], which might not be surprising since studies until now, agree that genetic variants explain up to 10% of the heritable risk [81].

Based on findings from previous reports in the field of epigenetics, a limited number of studies have been conducted to explore the suitability of the reported epigenetic markers to predict the onset of T2D and, whether the findings in blood samples can be used as surrogate markers for epigenetic modifications in target tissues for this disease [82 83].

4.1. DNA methylation

Investigating global DNA methylation, a study examined the predictive value for *LINE-1* as a risk marker for T2D and other metabolic disorders. Worsening of metabolic status or T2D onset after 1 year of follow-up was assessed. A model based on classic risk factors (age, sex, body mass index and physical activity) showed an AUC of 0.646[82]. The addition of *LINE-1* methylation measures to the previous model, significantly improved the predictive performance to an AUC of 0.650 [82]. The study was performed among European Spanish women 40 to 65 years old. No other evidence has been reporting the added value of global DNA methylation as a predictive tool in diabetes.

Methylation sites in the DNA, previously reported to be associated with T2D and glucose homeostasis in candidate gene studies or EWAS, have been taken into consideration in the search for biomarkers. One longitudinal study found that the methylation marker cg06500161, annotated to gene *ABCG1*, was associated with a 9% increased risk for future T2D, whereas methylation at cg02650017 annotated to *PHOSPHO1*, was associated with a 15% decreased risk for future T2D [83]. Nevertheless, no further investigation of the predictive value of these CpGs was performed.

The role of DNA methylation in predicting onset of T2D is still in its infancy of investigations. An ongoing longitudinal effort is combining clinical investigation, omics profiling (metabolomics, lipidomics, transcriptomics and epigenomics) with exercise and dietary interventions to provide novel diagnostic and predictive biomarkers to effectively detect the progression towards diabetes in high risk individuals, and also to predict responsiveness to lifestyle interventions known to be effective in the prevention of diabetes [84]. The study included 1455 participants from the DEXLIFE consortium and 400 participants in the intervention group. In the future, this com-

prehensive approach may provide some more insights on the contribution of DNA methylation sites as predictive markers.

4.2. Histone Modifications

Although a few studies have been conducted to examine the association between histone modifications and T2D [73-75], none of these markers have been proposed as possible novel biomarkers to identify subjects at high risk for T2D. The lack of consistent evidence and replicated studies may explain the absence of a reliable candidate marker serving for clinical purposes.

5. EPIGENETIC CHANGES ASSOCIATED WITH DIABETIC COMPLICATIONS

Diabetes is associated with significantly accelerated rates of several debilitating microvascular complications such as nephropathy, retinopathy, and macrovascular complications such as cardiovascular events. While several studies have been investigating genetic factors related to diabetes and associated complications, little is known about epigenetic changes that occur without alterations in the DNA sequence.

5.1. Epigenetic modifications in cardiovascular disease

Prospective studies have shown that diabetic patients have a two- to four fold risk to develop coronary artery disease, establishing that T2D is an independent risk factor for cardiovascular disease (CVD) [85]. About 70% of T2D patients at an age ≥65 years die from CVD, while T2D cases with no history of coronary artery disease have an equal cardiovascular risk as patients with previous myocardial infarction [86]. CVD and T2D share several common pathophysiological features like the classical cardiovascular risk factors, such as dyslipidaemia, hypertension and obesity. However, all of the known pathways do not explain the complex pathophysiology behind cardiovascular complications of diabetes. Up to date, the underlying mechanisms are often addressed within a specific pathological context, whereas an integrated approach should be preferred in order to capture all potential interlinks between T2D and CVD. New research investigations have linked the participation of epigenetic mechanisms in the process of inflammation, oxidative stress, and endothelial dysfunction, all representing the hallmark of cardiovascular complications of diabetes.

5.1.1. DNA methylation

Using human aortic endothelial cells exposed to high glucose and aortas of diabetic mice, one study found that the mitochondrial adaptor *p66Shc* was epigenetically upregulated by promoter CpG demethylation and H3 acetylation [87]. Moreover, the overexpression continued even after returning to normoglycemia and could only be inhibited after pharmacologic intervention, providing molecular insights for the progression of diabetic vascular complications despite glycaemic control, which might help to define novel therapeutic targets [87].

Although emerging data has linked some aspects of hypertrophy, heart failure, and arrhythmias in cardiomyocytes to DNA methylation and *PTMs* of histones, less evidence has been reported in hearts of diabetic patients [88]. Monkemann et al. reported that oxidative stress damages cardiomyocytes via *p53*-dependent apoptosis in diabetic cardiomyopathy [89]. Interestingly, in animal studies, the methylation of the *p21(WAF1/CIP1)* gene that encodes several protein kinases at *p53* showed the later to be an early step in the development of hyperglycemia-induced cardiomyopathy in diabetic rats [89].

Other epigenetic marks are linked to intermediate risk factors common to CVD and T2D and that could contribute in discovering diagnostic and prognostic factors. One study, investigating the association between *IGF2* methylation and lipid profile, showed that higher triglyceride/HDL-cholesterol ratio was associated with hypermethylation of *IGF2* gene, indicating that this gene might be an important marker of metabolic risk [90]. The *IGF2* gene provides instructions for making a protein called insulin-like growth factor 2, which plays an essential role in cell growth and insulin mechanisms. Another study that combined genome-wide transcriptome and CpG methylation profiling by array, reported many differentially methylated predicted sites in adipose tissue from insulin-resistant patients compared to controls, which included genes involved in insulin signalling and in the interaction with integrins [91].

Current therapies for diabetes are aimed to optimize glycaemic control and reduce the associated cardiovascular risk. Some preliminary studies have shown that DNA methylation plays an important role in the reversibility and treatment of diabetic complications such as CVD, including vascular inflammation [92]. Resveratrol is a polyphenol with antioxidative and anti-inflammatory properties. Numerous studies have shown that resveratrol might have cardiovascular protective effects and also might contribute in improving insulin sensitivity, reducing plasma glucose levels and reducing inflammation [92]. Lou et al. aimed to investigate the effects of resveratrol (trans-3, 5, 40-trihydroxystilbene) on the expression of pro-inflammatory cytokines

such as IL-1β, IL-6, TNF-α and IFN-γ in diabetic rat aortas and the potential epigenetic mechanisms involved [92]. It showed that the expression levels of pro-inflammatory cytokines were significantly lower in the resveratrol-treated diabetic group. Moreover, the untreated group showed reduced levels of DNA methylation at the specific cytosine phosphate guanosine sites of IL-1β, IL-6, TNF-α and IFN-γ and these levels were reversed by resveratrol [92]. Furthermore, incretins such as glucagon-like peptide 1 receptor *(GLP1R)* agonist are shown to have cardiovascular protection beyond glycaemic control in diabetes subjects[93]. Recent data show that methylation of *GLP1R* is associated with glycaemic control but also cardiometabolic risk factors, such as obesity [94]. Thus, although the studies of epigenetics marks and CVD in diabetes are scarce, they provide some insight on epigenetic modifications as possible targets to develop novel therapeutic agents at preventing and treating vascular complications.

5.1.2. Histone modifications

Gaikwad et al. reported deacetylation and dephosphorylation of histone H3 in the heart and kidney of diabetic Sprague-Dawley rats leading to changes in gene expression in the extracellular matrix and therefore hypertrophy [95]. A recent study used peripheral blood mononuclear cells to measure histone deacetylases (*HDACs*) activity and expression in relation to glycaemia, inflammation and insulin resistance in patients with T2D. Low-grade chronic inflammation and insulin resistance induced *HDAC3* activity and expression, and correlated positively with circulating levels of TNF-α, IL-6, and other proinflammatory markers, and negatively with *Sirt1* expression [96]. Using aortic endothelial cells, another study showed that exposure to high glucose correlates with the inverse acetylation of the histone *H3K9/K14* and modified DNA methylation patterns [97]. Several histone lysine modifications have also been described following transient high glucose levels that may account for a persistent transcriptional induction of the *RELA* gene, encoding for the *p65* subunit of *NF-kB*, even after subsequent incubation of endothelial cells with normal glucose concentrations [98]. Miao et al. recently compared patients previously included in the conventional treatment arm of the Diabetes Control and Complications Trial who developed diabetic microangiopathy (cases) to patients who were allocated the intensive treatment and had no progression of microvascular complications (controls) [99]. They reported a significantly greater number of promoter regions with enrichment in *H3K9Ac* (hyperacetylation) in monocytes, but not in lymphocytes, in cases versus controls [99]. These findings further support the existence of an epigenetic component in the metabolic memory—the concept that early glycaemic control is a major determinant of diabetic complications later in life.

5.2. Epigenetic modifications in diabetic nephropathy

Diabetic nephropathy (DN) is a major chronic complication of diabetes and the most common cause of end-stage kidney disease [100]. Approximately 50% of patients who have end-stage renal disease needing painful and costly dialysis are diabetic [100]. The underlying molecular mechanisms leading to DN are not fully elucidated. High glucose levels adversely impact all renal cell types including mesangial cells, tubular cells, podocytes and endothelial cells, and augment monocyte and macrophage infiltration [101]. High glucose conditions also increase the formation of advanced glycation end-products and production of growth factors such as transforming growth factor β1 (TGF-β1) and angiotensin II in renal cells [101]. Although several classic mechanisms and pathways leading to DN have been described over the years, new molecular and epigenetic mechanisms are emerging [102].

5.2.1. DNA methylation

The role of DNA methylation in DN has elicited much interest [102] mainly because most genome-wide association studies of DN have yielded few susceptibility loci. Studies of DNA methylation profiles in genomic DNA of diabetic patients with or without DN revealed differential methylation levels in several genes, including *UNC13B*, which has been suggested to mediate apoptosis in glomerular cells as a result of hyperglycaemia, and hence the association could be relevant to the initiation and pathogenesis of DN [103]. In DN, prolonged exposure to hyperglycemia induces production of cytokines, chemokines, and growth factors including *TGFβ1* and connective tissue growth factors, which leads to abnormal glomerular pathology. Brennan et al. measured DNA methylation in 192 candidate genes previously identified to be differentially expressed in *in vitro* models of DN and in renal biopsies from individuals with DN. The study found that 301 CpGs in 38 out of 192 genes were differentially methylated [104]. The gene ontology analysis of the differentially methylated genes revealed that the predominant biological function of the affected genes was organism development [104]. Additional studies using various DNA methylation assays and DNA collected from peripheral blood samples or saliva identified specific DNA methylation profiles for diabetic patients with and without nephropathy [105 106]. These studies proposed using DNA methylation profiles as biomarkers to help predict disease status and progression; however, they did not report associated gene expression data.

5.2.2. Histone modifications

Growing evidence suggests that histone post-translational modifications can have key roles in the pathogenesis of diabetes. *TGF-β1*-induced expression of plasminogen activator inhibitor-1 (*PAI-1*) and *p21* in renal mesangial cells plays a major role in

glomerulosclerosis and hypertrophy, key events in the pathogenesis of diabetic nephropathy. However, the involvement of histone acetyl transferases (*HATs*) and *HDACs* that regulate epigenetic histone lysine acetylation, and their interaction with *TGF-β1*-responsive transcription factors, are not clear. *In vitro* studies in rat mesangial cells, have shown that *TGF-β1,* which is a central mediator of fibrogenesis and high glucose treatment, increased *H3K9/14ac* enrichment near Smad and SP1 binding sites (proteins related to the control of gene expression and cell growth) at the plasminogen activator inhibitor 1 and *p21* gene promoters, together with *HATs p300* and *CREB*-binding protein [107]. Histone *H3K9/14ac* was found to have a key role in transcription of these genes in response to *TGF-β1* [107]. High glucose treatment also elicited similar histone post-translational modifications at fibrotic and cell-cycle gene (*p21*) promoters, which were blocked by an anti-*TGF-β* antibody [107]. Therefore, it is conceivable that histone hyperacetylation and related chromatin events involved in *TGF-β1*-mediated *PAI-1* and *p21* expression play important roles in the pathogenesis of DN and could therefore serve as potential therapeutic targets for diabetes-induced renal dysfunction. *In vivo* animal models of DN have also demonstrated changes in histone post-translational modifications. *NF-κB*-dependent inflammatory gene expression has been extensively studied due to the involvement of these target genes in the pathology of several inflammatory diseases, including atherosclerosis, insulin resistance, diabetes and its complications. Studies demonstrated that *H3K4me HMT SET7* could be a NF-κB coactivator at a subset of inducible inflammatory genes in monocytes[108]. *SET7/9* may therefore be a novel therapeutic target for inflammatory diseases, including diabetes, DN and related metabolic disorders. Interestingly, treatment of mice with losartan, an Ang II type 1 receptor blocker, ameliorated key indices of diabetic nephropathy, and reversed key changes in epigenetic enzymes and *H3K9ac* enrichment at promoters of genes encoding *PAI-1* and *RAGE*, but did not reverse all the diabetes-induced epigenetic changes [109]. Thus, the relative inefficiency of drugs commonly used for diabetic nephropathy, such as Ang II type 1 receptor blockers, to prevent progression to renal failure, in many patients could be due to the incomplete reversal of diabetic nephropathy-associated epigenetic changes [109]. Also, several clinical trials (NCT01038089, NCT00937222) have been conducted testing resveratrol, (class III histone deacetylases group) which is thought to protect against development of diabetic nephropathy via changes in phosphorylation of histone H3 and Sir-2. The activity of resveratrol shows great potential in the prevention and therapy of diabetes and its complications especially to DN, nevertheless the compound is still in its experimental phases. These and other studies emphasize the role of epigenetic mechanisms in the regulation of possible biological and genetic pathways relevant to DN.

5.3. Epigenetic modifications in diabetic retinopathy

Retinopathy, a sight-threatening disease, remains one of the most feared complications of diabetes. Although altered levels of glucose are the main initiators, progression of diabetic retinopathy continues even after the hyperglycaemic insult is reversed by good glycaemic control, suggesting a 'metabolic memory' phenomenon [110 111].

5.3.1. DNA methylation

Recent studies have implicated epigenetic modification in the metabolic memory phenomenon associated with the continued progression of diabetic complications. T2D activates the enzymes responsible for maintaining DNA methylation status in the retina, increasing the activities of DNA methyltransferases (DNMTs) and ten-eleven translocation enzymes (Tets) which can lead to hypo- or hypo- DNA methylation of many genes responsible for mitochondrial homeostasis [112-114]. A dynamic DNA methylation process of the matrix metalloproteinase gene *MMP-9* is shown to maintain its transcriptional activation, though the transcription factor binding sites of the *MMP-9* promoter, which are hypermethylated in diabetes. Due to concomitant increased binding of Tet at the same site, *MMP-9* DNA remains hypomethylated resulting in its transcriptional activation [114] and this continues to fuel into the mitochondrial damage [110]. Abrogation of *MMP-9* gene protects against the development of retinopathy and it is considered to play an important role in the apoptosis of retinal capillary cells.

5.3.2. Histone modifications

Histone acetylation and methylation (*HDAs, HATs, SETs* and *LSD1*) play also a crucial role in epigenetic modifications that occur during diabetic retinopathy. However, experimental studies of diabetic retinopathy have provided contradictory results for histone acetylation, with some showing increased global acetylation of retinal histones with activation of *HAT* and inhibition of *HDACs* [115], and others reporting significant increase in retinal histone acetylation [116 117]. The reasons for such discrepancies are still unknown. Although histone modifications and DNA methylation generally regulate gene transcription independently, DNA methylation and histone modifications are also shown to work in concordance; e.g., cooperation of DNA methylation with histone methyltransferase *SETDB1/ESET* results in trimethylation of *H3K9* (known for its function in condensing the chromatin) [110]. DNMTs and histone modifying enzymes, *SUV39h1* and *EZH2* lysine histone methyltransferase, can work in coordination, and the agents that interact with histone methyl transferases, in addition to regulating histone methylation, also regulate DNA methylation [118].

Some of the studies have shown that even after the reverse of the hyperglycemic insult, *H3K4* and *H4K20* at *Sod2* methylation levels continue to be altered. Also, the enzymes responsible for histone methylation (*SUV420h2*) and demethylation (*LSD1*) remain dysfunctional [116]. Upregulation of *MMP-9*, which encodes another enzyme implicated in diabetic retinopathy, was associated with reduced promoter *H3K9me2* and increased *H3K9ac* levels, along with increased recruitment of NF-κB in retinal endothelial cells from diabetic rats [119]. Furthermore, mass spectrometry studies demonstrated that hyperglycaemia causes acetylation of retinal histones, which was associated with increases in proinflammatory proteins [120]. The *HAT p300* was implicated in endothelial fibronectin expression related to diabetic retinopathy, and in gene expression relevant to diabetic cardiac hypertrophy [120].

5.4. Epigenetic modifications in diabetic neuropathy

Diabetic neuropathy is a common but irreversible complication that develops in up to 50% of patients with diabetes and results in sensory loss, pain, and risk of amputation [121]. The molecular mechanisms involved in the development of diabetic neuropathy is a complex process that includes activation of the polyol pathway, exaggerated oxidative stress, over activity of protein kinase C and increased formation of advanced glycation end-products in the presence of hyperglycaemia. Diabetic neuropathy entails decreased nerve conduction velocity, which indicates a prominent role for Schwann cells because they ensheath peripheral nerves and provide support for nerve conduction and axon regeneration [122]. Moreover, high glucose induces oxidative damage in Schwann cells, considered a major factor in diabetic complications [122]. With respect to diabetic neuropathy, several biological mechanisms have been studied, although the role of epigenetics has only recently been suggested [122].

5.4.1. DNA methylation

Previous studies have suggested that diabetic neuropathy involves dysregulation of the transcription factor peroxisome proliferator-activated receptor (*PPRG* or *PPARγ*) [123]. Although drugs that activate *PPARγ* improve glycaemic control in T2D, its role in metabolic memory has not been extensively examined. Because high glucose produces changes in gene expression that persist after cell division [124], a plausible mechanism for these stable changes is DNA methylation. Using quantitative PCR arrays for glucose and fatty acid metabolism, one study found that chronic high glucose induced a persistent increase in genes that promote glycolysis, while inhibiting those that oppose glycolysis and alternate metabolic pathways such as fatty acid metabolism, the pentose phosphate pathway, and trichloroacetic acid cycle [122]. These sustained effects were associated with decreased *PPARγ* binding and persis-

tently increased reactive oxygen species, cellular NADH, and altered DNA methylation [122]. Their results suggest that Schwann cells might exhibit features of metabolic memory that may be regulated at the transcriptional level therefore, targeting *PPAR* may prevent metabolic memory and the development of diabetic complications such as diabetic neuropathy.

5.4.2. Histone modifications

With respect to diabetic neuropathy and histone modifications, little evidence is available. Nevertheless, factors like dyslipidaemia, oxidative stress and inflammation have been reported to be particularly important for the development of neuropathy [125]. Therefore, one could speculate that some of the reported mechanisms linking diabetes with dyslipidaemia, oxidative stress and inflammation through histone modifications, could play a role also in diabetic neuropathy. Nevertheless, future studies are needed to shed light about the role that epigenetic mechanisms such as histone modifications play in diabetic neuropathy.

6. EPIGENETIC BASED BIOMARKERS FOR DIAGNOSIS AND PROGNOSIS OF T2D

Extensive research over the last decades, has shown that our genome is not the only determinant of disease risk, and that epigenetic marks induced by lifestyle and environmental factors are associated with altered gene expression patterns in important tissues, leading to altered susceptibility to disease in later life [8]. Thus, it should be possible to detect these altered epigenetic marks and use them as predictors of future metabolic capacity, early detection of disease and better prognosis (Figure 2). A premature diagnosis of disease is crucial as it might greatly improve clinical outcomes for patients. For example, population screening for cancers and the surveillance of high-risk patients allows an early diagnosis of cancer and therefore reduces morbidity by using less invasive treatment, resulting in fewer complications and side effects. However, constant screening for a broad range of diseases across whole populations is not routine and realistic and it poses a high economic burden to governments [77]. Hence, the detection of epigenetic alterations has become a promising tool in many health areas for the diagnosis and prognosis of disease and for the prediction of drug response.

6.1 Epigenetic based biomarkers for diagnosis of T2D

Although great progress has been made in the description of epigenetic modifications in normal and diseased tissues, the studies so far have been mainly focused

on cancer research. In prostate cancer, DNA hypermethylation at the glutathione S-transferase pi 1 (*GSTP1*) enzyme gene has been suggested as the most relevant candidate biomarker. It has been detected in 69% of proliferative inflammatory atrophy lesions, that are considered precursor lesions for the development of prostate cancer and/or high-grade prostatic intraepithelial neoplasia, underlining its importance for diagnosis [126]. This epigenetic mark was consistently validated in many studies, showing a sensitivity of 82% and a specificity of 95% [127], compared to serum prostate-specific antigen (PSA, 20%), which is the only biomarker currently used for the detection and monitoring of this cancer. However, *GSTP1* can also be hypermethylated in some other types of cancers. Therefore, combinations with additional biomarker genes and PSA testing have been recommended in order to increase specificity. Nevertheless, later studies that combined DNA hypermethylation of *GSTP1, APC, RASSF1, PTGS2, MDR1* and *TIG1* resulted in both sensitivity and specificity up to 100% [128 129]. Furthermore, for the detection of glioblastoma, the hypermethylation at the promoter of the gene O-6-Methylguanine-DNA Methyltransferase (*MGMT*) has been established as a promising biomarker, since it was possible to detect glioblastomas with a very high sensitivity and specificity (95% and 60% respectively) in serum samples [130]. Also, researchers were able to predict a lack of cancer progression and overall survival of the patient [130]. A step further, in colorectal cancer (CRC), the Food and Drugs Administration (FDA) has already approved the use of epigenetic markers like *SEPT9* (ColoVantage) and vimentin (ColoSure)[131] in clinical practice, after showing a high sensitivity and a specificity for CRC diagnosis [132 133].

As the field has grown, efforts are made also in discovering candidate biomarkers in the diagnosis and prognosis of other types of diseases such as type 2 diabetes, but this research field is still in its early phases. Twin studies might be a powerful approach as, despite the identical genetic background, discordant twins for T2D have differences in DNA methylation [134 135].

Recently a new classification of diabetes types has been proposed. Using a data-driven cluster analysis approach, newly diagnosed diabetic patients were stratified into five subgroups with differing disease progression and risk of diabetic complications [136]. The study identified five replicable clusters of patients with diabetes, who had significantly different patient characteristics and risk of diabetic complications. Cluster 1 included severe autoimmune diabetes patients; cluster 2 included patients with severe insulin-deficient diabetes; cluster 3 included severe insulin-resistant diabetes patients; cluster 4 was labelled as mild obesity-related diabetes and cluster 5 was labelled as mild age-related diabetes. They found that individuals in cluster 3 had significantly higher risk of diabetic kidney disease than individuals in clusters 4 and

5, but had been prescribed similar diabetes treatment. While, individuals in cluster 2 had the highest risk of retinopathy compared to the other clusters. Moreover, genetic associations in the clusters differed from those seen in traditional T2D [136]. Considering the high heterogeneity of T2D, epigenetics might be a promising path that could help to better disentangle the differences in diabetic phenotypes, which would provide more information on treatment target, as well in prevention strategies individualized by diabetes type, thus evolving towards personalized medicine.

In type 1 diabetes mellitus, recent findings propose the circulating β cell-derived unmethylated insulin (*INS*) DNA as a potential diagnostic tool for the early detection of type 1 diabetes (T1D) [137]. This is based in previous evidence showing that the autoimmune destruction conducted by immune cells in T1D leads to the release of unmethylated *INS* DNA from pancreatic β cells, into the circulation. Thus it is possible the detection of the circulating DNA in blood samples, and the consequent assessment of the unique methylation pattern present in the promotor of the *INS* gene in β cells from patients with early T1D, or in patients with islet transplantation therapy [138].

6.2 Epigenetic based biomarkers for prognosis of T2D

In addition to its diagnostic potential, DNA methylation could be informative for patient prognosis in terms of disease progression/recurrence, complications, treatment and survival. High resolution data usually used for cancer, subtypes-specific profiles can also be used for the identification of powerful single epigenetic biomarker genes and gene combinations at various stages of disease. This advanced screening strategy can be beneficial for many types of clinical applications.

Initially, the potential prognostic profile of DNA methylation has been reported in childhood acute lymphoblastic leukaemia [139]. DNA methylation profiling classified these patients in lymphoblastic leukaemia subtypes and stratified them with high hyperdiploidy and translocation t(12;21) into two subgroups with different probabilities of relapse. DNA methylation profiling therefore resulted in subtype-specific and predictive signatures with potential use as future prognostic biomarkers for this disease [139]. In breast cancer, DNA methylation profiling identified a previously unrecognized subtype associated with T lymphocyte infiltration. Importantly, profiling immune genes determined a prognostic value of the profile. In particular, the hypermethylation at the promoter of lymphocyte transmembrane adaptor 1 (*LAX1*) and *CD3D* significantly correlated with survival in certain breast cancer subtypes [140]. Another study discovered and validated the epigenetic signature of *NEFH* and *HS3ST2* as an independent prognostic factor for type II ovarian cancer [141]. Moreover,

they showed that 3-0 sulfation of HS was important to oncogenic signalling, such as IL-6 and EGF signalling, which could render useless current targeted therapies for ovarian cancer without further patient stratification [141].

However, when it comes to the role of epigenetics used as a prognostic marker for T2D the data are very limited. Recent published studies have suggested that the differentially methylated circulating DNA might be a promising novel biomarker which reflects beta cell death and could predict the progression of diabetes [142]. The novel beta cell death marker unmethylated insulin (*INS*) DNA has been studied in TPIAT (total pancreatectomy with islet autotransplantation) patients before and immediately after islet infusion, and also 90 days post-TPIAT concurrent with metabolic functional assessments [143]. Universal early elevations in the beta cell death marker *INS* DNA after TPAIT were observed, with pronounced elevations in the islet supernatant pre-infusion, likely reflecting beta cell death induced by islet isolation. In addition, persistent post-transplant elevation of INS DNA predicted greater hyperglycaemia at 90 days [143]. Although, more studies are needed to identify the best methylation target sites in the *INS* gene, differentially methylated circulating DNA may be a good method to evaluate progression, diagnosis and prognosis of islet related diseases and in diabetes patients for whom insulin production is impaired.

Aging is also an important risk factor for metabolic disorders, including obesity, impaired glucose tolerance, and T2D [144]. Almost one third of the elderly in the United States of America have diabetes and three quarters have diabetes or prediabetes [145]. Recent attempts have been done in identifying biomarkers of aging, which are thought of as individual-level measures of aging that capture inter-individual differences in the timing of disease onset, functional decline and death over the life course. In this context, DNA methylation has gained a lot of interest as a potential biomarker that could predict mortality [146]. Successfully, a DNA methylation based biomarker of aging (PhenoAge) was developed, which is highly predictive of nearly every morbidity and mortality outcome tested, especially cardiovascular disease and coronary heart disease. In addition, the study observed that higher DNA methylation PhenoAge was associated with increases in the activation of proinflammatory pathways, such as NF-κB, increased interferon signalling and decreases in damage recognition and repair pathways [146]. Given this, one could hypothesize that DNA methylation PhenoAge might have an influence also on T2D. Future studies could test this hypothesis in diabetes patients which would provide some insights on whether methylation age of specific methylation sites would identify T2D patients who are at risk of early death.

7. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

Epigenetics promises an auspicious future in its role as clinical instrument to complement the current practice. Current evidence in the field suggests that study of the role of epigenetic changes in the onset and progression of the diseases is a promising opportunity for the development of prediction, diagnostic and disease progression monitoring tools, as well as novel therapeutic targets (**Figure 2**). Overall, changes in global DNA methylation, CpG islands methylation, and histone acetylation and methylation have been found in relevant genes for aorta, heart, kidney, retina, nerves and glia cells function in samples from diabetic human and rodent donors or after hyperglycaemic stimuli of the same type of tissues, albeit some of them exhibit contrary results. However, since epigenetic changes have shown to be dynamic and vary in response of environmental stimulations, epigenetic mechanisms are important to study also because it is believed that they may play a significant role in the reversibility and treatment of diabetic complications, contributing even as a protective mechanism against them, though which drugs and other therapeutic tools may exert their effect. Furthermore, not solely for risk prediction, epigenetic changes also arise as a promising tool in the diagnosis, stratification of the patients, prognosis and monitoring of therapy response. Various research methodologies and DNA methylation assessment methods have been used to disentangle the epigenetic network; however, the technological advances in DNA methylation arrays have permitted EWAS approach to analyse a higher number of CpGs, a larger samples size and a more standardized method to assess DNA methylation. Nevertheless, identifying optimal methods of detecting possible epigenetic biomarkers and implementing appropriate reference standards are rapidly evolving. The use of common plans in the study designs, methylation assessment methods and statistical models for data analysis with similar strategies for confounding control, would allow comparability and independent validation of the findings, thus better quality evidence to identify biomarkers for T2D precision medicine, reliable for both clinicians and the patients.

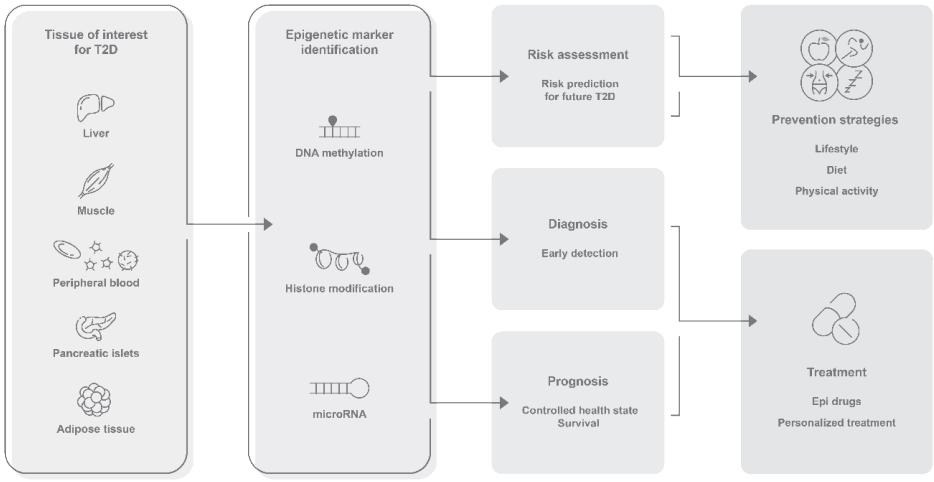


Figure 2. Characteristic epigenetic patterns for T2D and glycemic traits can be investigated in key tissues for glucose homeostasis and energy balance in the body. After the identification of an epigenetic signature, it can be used as a biomarker for the detection of subjects at high risk of developing T2D, before the onset of the disease. Thus, the patient can receive healthcare advice and is still on time to implement strategies in order to prevent the progression towards T2D. A biomarker can also be of use after the onset of the disease, to make a more accurate diagnose, to choose the best treatment strategy and to monitor the progression of T2D and its complications in order to augment patient´s survival and reduce co-morbidities.

Table 1. Overview of genes studied in association of Epigenetics with T2D and glycemic traits (Database URL: www.genecards.org) [146]

Gene	Alias	Chr	Function
IGF1R	Insulin Like Growth Factor 1 Receptor	15	It encodes a receptor which binds insulin-like growth factor 1 (IGF1) with a high affinity, and IGF2 and insulin (INS) with a lower affinity. It is involved in cell growth and survival control, being crucial for tumor transformation and survival of malignant cell. Among its related pathways are Apoptotic Pathways in Synovial Fibroblasts and NFAT and Cardiac Hypertrophy. Associated diseased include Insulin-Like Growth Factor I Deficiency and Ring Chromosome 15 Syndrome.
IGFBP3	Insulin Like Growth Factor Binding Protein 3	7	It encodes a protein with an IGFBP domain and a thyroglobulin type-I domain. The protein forms a ternary complex with insulin-like growth factor acid-labile subunit (IGFALS) and either insulin-like growth factor (IGF) I or II. In this form, it circulates in the plasma, prolonging the half-life of IGFs and altering their interaction with cell surface receptors. Diseases associated with IGFBP3 include Insulin-Like Growth Factor I and Acid-Labile Subunit Deficiency.
IGFBP7	Insulin-like Growth Factor (IGF) Binding protein 7	4	It encodes a protein that binds IGFs to regulate their binding to its receptors. It also stimulates prostacyclin production, which is a potent inhibitor of platelet aggregation and a strong vasodilator that inhibits the growth of vascular smooth muscle cells. Associated diseases include Diabetic Angiopathy.
IGFBP1	Insulin-like growth factor binding protein 1	7	It encodes a protein that binds IGF-I and –II to regulate their binding to their receptors. It is mainly expressed in liver. Low levels of this protein may be associated with impaired glucose tolerance, vascular disease and hypertension in human patients. Associated diseases include Pancreatic Cancer, Childhood and Ovarian Disease.
IGF2	Insulin Like Growth Factor 2	11	It encodes a member of the insulin family of polypeptide growth factors, which are involved in development and growth. It plays a key role in regulating fetoplacental development. In adults, it may be involved in glucose metabolism in adipose tissue, skeletal muscle and liver. It undergoes glucose-mediated co-secretion with insulin, and acts as physiological amplifier of glucose-mediated insulin secretion. Associated diseases include Growth Restriction and Silver-Russell Syndrome.
C8orf31	Chromosome 8 Open Reading Frame 31 (Putative)	8	It is an RNA gene, and it is affiliated with the ncRNA class.
TXNIP	Thioredoxin Interacting Protein	1	It encodes a protein that may act as an oxidative stress mediator by inhibiting thioredoxin activity or by limiting its bioavailability. It also functions as a transcriptional repressor, possibly by acting as a bridge molecule between transcription factors and corepressor complexes, and over-expression will induce G0/G1 cell cycle arrest. It is required for the maturation of natural killer cells. Associated diseases include Hyperglycemia.

Table 1. Overview of genes studied in association of Epigenetics with T2D and glycemic traits (Database URL: www.genecards.org) [146] (*continued*)

Gene	Alias	Chr	Function
MMP-9	Matrix Metallopeptidase 9	20	It encodes an enzyme member of the Metallopeptidase (MMP) family, which are involved in the breakdown of extracellular matrix in normal physiological processes, such as embryonic development, reproduction, and tissue remodeling, as well as in disease processes, such as arthritis and metastasis. Specifically, MMP-9 degrades type IV and V collagens. Associated diseases include Metaphyseal Anadysplasia 2 and Metaphyseal Anadysplasia.
EP300	Histone Acetyltransferase P300	22	It encodes a protein that functions as histone acetyltransferase, which regulates transcription via chromatin remodeling and is important in the processes of cell proliferation and differentiation. It mediates acetylation of histone H3 at Lys-122 (H3K122ac) and at Lys-27 (H3K27ac). It also functions as acetyltransferase for non-histone targets. Associated diseases include Rubinstein-Taybi Syndrome 2 and Colorectal Cancer.
SSTR5-AS1	SSTR5 Antisense RNA 1	16	It is a non-protein coding gene. It is an RNA Gene, and is affiliated with the non-coding RNA class
SSTR5	Somatostatin Receptor 5	16	It encodes a cyclic polypeptide which is an abundant neuropeptide and has a wide range of physiological effects on neurotransmission, secretion and cell proliferation. The activity of this receptor is mediated by G proteins which inhibit adenylyl cyclase, and different regions of this receptor molecule are required for the activation of different signaling pathways. Associated diseases include Pituitary Adenoma
LY86	Lymphocyte Antigen 86	6	It encodes a protein which is involved in the innate immune system. It may cooperate with CD180 and TLR4 to mediate the innate immune response to bacterial lipopolysaccharide (LPS) and cytokine production. Important for efficient CD180 cell surface expression. Associated diseases include Parametritis and Interstitial Emphysema.
TLR2	Toll-like Receptor 2	4	It encodes a cell-surface protein that cooperates with TLR1 or TLR6 to mediate the innate immune response after recognition of the pathogen-associated molecular patterns (PAMPs), such us bacterial lipoproteins or lipopeptides, and it is also thought to promote apoptosis in response to them. It is implicated in the pathogenesis of several autoimmune diseases.
SLC30A8	Solute Carrier Family 30 member 8	8	It encodes a zinc ion efflux transporter that is highly expressed only in the pancreas, particularly in islets of Langerhans. It provides zinc to insulin maturation or storage in the pancreatic beta-cells. Allelic variants of this gene can confer susceptibility to T2D.
GCK	Glucokinase (Hexokinase 4)	7	It encodes an enzyme that phosphorylates glucose to glucose-6-phosphate, the first step in most glucose metabolism pathways. It is expressed in pancreas and liver. In the pancreas, it plays a role in glucose-stimulated insulin secretion. In the liver, it is important in glucose uptake and conversion to glycogen. Mutations in this gene are associated with multiple types of diabetes and hyperinsulinemic hypoglycemia.

Table 1. Overview of genes studied in association of Epigenetics with T2D and glycemic traits (Database URL: www.genecards.org) [146] (*continued*)

Gene	Alias	Chr	Function
PRKCZ	Protein Kinase C Zeta type	1	It encodes an enzyme that plays a role as activator or downstream effector in diverse signaling cascades in different cell types. In adipocytes, upon insulin treatment may contribute to the activation of translocation of the glucose transporter SLC2A4/GLUT4 and subsequent glucose transport.
CTGF	Connective Tissue Growth Factor	6	It encodes a protein that is secreted by vascular endothelial cells, and plays a role in chondrocyte proliferation and differentiation, cell adhesion in many cell types, and it is related to platelet-derived growth factor. Polymorphisms in this gene can be related to a higher incidence of systemic sclerosis.
LEP	Leptin	7	It encodes a protein that is secreted by white adipocytes into the circulation and plays a major role in the regulation of energy homeostasis. Circulating leptin binds to its receptor in the brain, activating downstream signaling pathways that inhibit feeding and promote energy expenditure. It also has endocrine functions, participates in the regulation of immune and inflammatory responses, hematopoiesis, angiogenesis, reproduction, bone formation and wound healing. Mutations in this gene can cause severe obesity, morbid obesity with hypogonadism and T2D.
IRS-1	Insulin Receptor Substrate 1	2	It encodes a protein which, when is phosphorylated by insulin receptor tyrosine kinase, binds specifically to various cellular proteins, thus controlling diverse cellular processes. Mutations in this gene are associated with T2D and susceptibility to insulin resistance.
GIPR	Gastric Inhibitory Polypeptide Receptor	19	It encodes a G-protein coupled receptor for gastric inhibitory polypeptide (GIP), demonstrated to stimulate insulin release in the presence of elevated glucose. Defect in this gene thus may contribute to the pathogenesis of diabetes
CAMK1D	Calcium/Calmodulin Dependent Protein Kinase I Delta	10	It encodes an enzyme that is a component of the calcium-regulated calmodulin-dependent protein kinase cascade. It has been associated with the regulation of granulocyte function, activation of CREB-dependent gene transcription, aldosterone synthesis and secretion, differentiation and activation of neutrophil cells, and apoptosis of erythroleukemia cells.
CRY2	Cryptochrome Circadian Regulator 2	11	It encodes a transcriptional repressor which forms a core component of the circadian clock. It regulates various physiological processes, including metabolism, sleep, body temperature, blood pressure, endocrine, immune, cardiovascular and renal function. I also plays a key role in glucose and lipid metabolism modulation, in part, through the transcriptional regulation of genes involved in these pathways.
CALM2	Calmodulin 2 (Phosphorylase Kinase, Delta)	2	It encodes a calcium binding protein that plays a role in signaling pathways, cell cycle progression and proliferation. It mediates the control of a large number of enzymes, aquaporins and other proteins through calcium-binding and ion channels, such us the calcium-activated potassium channel KCNN2. Diseases associated with CALM2 include Long Qt Syndrome 15 and Long Qt Syndrome 1.

Table 1. Overview of genes studied in association of Epigenetics with T2D and glycemic traits (Database URL: www.genecards.org) [146] (*continued*)

Gene	Alias	Chr	Function
MCP1	Monocyte Chemoattractant Protein 1	17	It encodes a cytokine that displays a chemotactic activity, attracting monocytes and basophils but not neutrophils or eosinophils. It augments monocyte anti-tumor activity and has been implicated in diseases characterized by monocytic infiltrates, like psoriasis, rheumatoid arthritis or atherosclerosis.
TLR4	Toll-Like Receptor 4	9	It encodes a receptor that cooperates with LY96 and CD14 to mediate the innate immune response to bacterial lipopolysaccharide (LPS). It is also involved in LPS-sindependent inflammatory responses triggered by free fatty acids, such as palmitate, and Ni(2+). In complex with TLR6, promotes sterile inflammation in monocytes/ macrophages in response to oxidized low-density lipoprotein (oxLDL) or amyloid-beta 42. Mutations in this gene have been associated with differences in LPS responsiveness.
FFAR3	Free Fatty Acid Receptor 3	19	It encodes a receptor that is activated by a major product of dietary fiber digestion, the short chain fatty acids (SCFAs), for the regulation of whole-body energy homeostasis, glucose homeostasis, intestinal immunity and indirectly LEP/Leptin production.
PP2Ac	Protein Phosphatase 2 Catalytic Subunit Alpha	5	It encodes the alpha isoform of the catalytic subunit of the Protein Phosphatase 2A, which can modulate the activity of some kinase enzymes. It is involved in signal transduction and in the negative control of cell growth and division. Diseases associated include Usher Syndrome, Type I.
PPARG	Peroxisome Proliferator-Activated Receptor Gamma	3	It encodes PPAR-gamma (PPARγ) a member of the peroxisome proliferator-activated receptor (PPAR) subfamily of nuclear receptors which form heterodimers to regulate transcription of various genes. PPARγ is a regulator of adipocyte differentiation and has been implicated in the pathology of numerous diseases including obesity, diabetes, atherosclerosis and cancer. Associated diseases include Familial Partial Lipodystrophy Type 3 and Intimal Medial Thickness of Internal Carotid Artery.
PPARGC1A	Peroxisome Proliferator-Activated Receptor Gamma Coactivator 1-Alpha	4	It encodes a transcriptional coactivator that regulates the genes involved in energy metabolism, such us mitochondrial genes, and the muscle fiber type determination. It also plays a role in metabolic reprogramming in response to dietary availability through coordination of the expression of genes involved in glucose and fatty acid metabolism, and in the integration of the circadian rhythms and energy metabolism. It may be also involved in controlling blood pressure, regulating cellular cholesterol homoeostasis, and the development of obesity.
PDX-1	Pancreatic and duodenal Homeobox 1 / Insulin upstream factor 1	13	It encodes a transcriptional activator of several genes, including insulin, somatostatin, glucokinase, islet amyloid polypeptide and glucose transporter type 2. It is involved in the early development of the pancreas and plays a major role in glucose-dependent regulation of insulin gene expression. A defective gene can lead to early-onset insulin-dependent diabetes mellitus, as well as maturity onset diabetes of the young type 4.

Table 1. Overview of genes studied in association of Epigenetics with T2D and glycemic traits (Database URL: www.genecards.org) [146] (*continued*)

Gene	Alias	Chr	Function
INS	Insulin	11	It encodes a precursor form called proinsulin, which is cleaved to form the A and B chains, and then joined together to form insulin. It binds to the insulin receptor (INSR) to stimulate glucose uptake. It blood glucose concentration and increases cell permeability to monosaccharides, amino acids and fatty acids. It accelerates glycolysis, the pentose phosphate cycle, and glycogen synthesis in liver. Associated diseases include Hyperproinsulinemia and Insulin-Dependent Diabetes Mellitus 2.
GLP1R	Glucagon-Like Peptide 1 Receptor	6	It encodes a 7-transmembrane protein that functions as a receptor for glucagon-like peptide 1 (GLP-1) hormone, which stimulates glucose-induced insulin secretion, and plays an important role in the signaling cascades leading to insulin secretion. The protein is an important drug target for the treatment of type 2 diabetes and stroke. Polymorphisms in this gene are associated with diabetes, insulinoma and fasting hypoglycemia.
UBASH3A	Ubiquitin Associated And SH3 Domain Containing A	21	It encodes one members of the T-cell ubiquitin ligand family, which can negatively regulate T-cell signaling by facilitating the growth factor withdrawal-induced apoptosis in T cells. It can also interfere in the down-regulation and degradation of receptor-type tyrosine kinases and promotes accumulation of activated target receptors, such as T-cell receptors on the cell surface. Diseases associated include Dirofilariasis and Erysipeloid.
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase	12	It encodes an enzyme that catalyzes an important energy-yielding step in carbohydrate metabolism, the reversible oxidative phosphorylation of glyceraldehyde-3-phosphate in the presence of inorganic phosphate and nicotinamide adenine dinucleotide (NAD). It also has uracil DNA glycosylase activity in the nucleus. Related diseased include Fragile X Mental Retardation 1.
TFAM	Transcription Factor A, Mitochondrial	10	It encodes a protein involved in mitochondrial DNA replication and repair. Sequence polymorphisms in this gene are associated with Alzheimer's and Parkinson's diseases.
TRIM3	Tripartite Motif Containing 3	11	The protein encoded by this gene is a member of the tripartite motif (TRIM) family, also called the 'RING-B-box-coiled-coil' (RBCC) subgroup of RING finger proteins. This protein localizes to cytoplasmic filaments. It is similar to a rat protein which is a specific partner for the tail domain of myosin V, a class of myosins which are involved in the targeted transport of organelles. Among its related pathways are Cytokine Signaling in Immune system and Innate Immune System.
TCF7L2	Transcription Factor 7 Like 2	10	This gene encodes a high mobility group (HMG) box-containing transcription factor that plays a key role in the Wnt signaling pathway. The protein has been implicated in blood glucose homeostasis. Genetic variants of this gene are associated with increased risk of type 2 diabetes. Associated diseases include Noninsulin-Dependent Diabetes Mellitus and Colorectal Cancer.

Table 1. Overview of genes studied in association of Epigenetics with T2D and glycemic traits (Database URL: www.genecards.org) [146] (*continued*)

Gene	Alias	Chr	Function
PDK4	Pyruvate Dehydrogenase Kinase 4	7	It encodes a protein located in the matrix of the mitrochondria and plays a key role in the regulation of glucose and fatty acid metabolism and homeostasis via phosphorylation of the pyruvate dehydrogenase subunits PDHA1 and PDHA2. Expression of this gene is regulated by glucocorticoids, retinoic acid and insulin. Diseases associated include non-insulin-dependent diabetes mellitus.
HNF4A	Hepatocyte Nuclear Factor 4 Alpha	20	It encodes a nuclear transcription factor which binds DNA and controls the expression of several genes, including hepatocyte nuclear factor 1 alpha, which regulates the expression of several hepatic genes. It also may play a role in development of the liver, kidney and intestines. Mutations in this gene have been related to with monogenic autosomal dominant non-insulin-dependent diabetes mellitus type I.
KLF11	Kruppel Like Factor 11	2	It encodes a transcription factor that binds to SP1-like sequences in epsilon- and gamma-globin gene promoters. This binding inhibits cell growth and causes apoptosis. Defects in this gene are a cause of maturity-onset diabetes of the young type 7 (MODY7).
DUSP9	Dual Specificity Phosphatase 9	X	It encodes an enzyme that regulates mitogen-activated protein (MAP) kinases by dephosphorylating both the phosphoserine/threonine and phosphotyrosine residues. It shows selectivity for members of the ERK family of MAP kinases and is localized to the cytoplasm and nucleus. Aberrant expression of this gene is associated with type 2 diabetes and cancer progression in several cell types.
HHEX	Hematopoietically Expressed Homeobox	10	It encodes a member of the homeobox family of transcription factors, many of which are involved in developmental processes and it may play a role in hematopoietic differentiation. Related diseases include Gangliosidosis Gm2 and Sandhoff Disease. Among its related pathways are transcriptional misregulation in cancer and mesodermal commitment pathway.
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A	9	It encodes several transcript variants which differ in their first exons. Two of them encode structurally related isoforms known to function as inhibitors of CDK4 kinase. The remaining transcript contains an alternate open reading frame (ARF) that specifies a protein that functions as a stabilizer of the tumor suppressor protein p53. The three share a common functionality in cell cycle G1 control. This gene is frequently mutated or deleted in a wide variety of tumors, and is known to be an important tumor suppressor gene.
KCNQ1	Potassium Voltage-Gated Channel Subfamily Q Member 1	11	It encodes a voltage-gated potassium channel required for repolarization phase of the cardiac action potential, and plays an important role in a number of tissues, including heart, inner ear, stomach and colon. Mutations in this gene are associated with hereditary long QT syndrome 1 (also known as Romano-Ward syndrome), Jervell and Lange-Nielsen syndrome, and familial atrial fibrillation.

Table 1. Overview of genes studied in association of Epigenetics with T2D and glycemic traits (Database URL: www.genecards.org) [146] (*continued*)

Gene	Alias	Chr	Function
CIDEC	Cell Death Inducing DFFA Like Effector C	3	It encodes a member of the cell death-inducing DNA fragmentation factor-like effector family, which plays important roles in apoptosis. The protein binds to lipid droplets in adipocytes and regulates their enlargement, thereby restricting lipolysis and favoring storage. It also may mediate adipocyte apoptosis. This gene is regulated by insulin and its expression is positively correlated with insulin sensitivity. Mutations in this gene may contribute to insulin resistant diabetes. Diseases associated include Familial Partial Lipodystrophytype 5 and Adiposis Dolorosa.
ADCY5	Adenylate Cyclase 5	3	It encodes an enzyme that catalyzes the formation of the signaling molecule cAMP in response to G-protein signaling and regulates the increase of free cytosolic Ca (2+) in response to increased blood glucose levels and contributes to the regulation of Ca (2+)-dependent insulin secretion. Single nucleotide polymorphisms in this gene may be associated with low birth weight and type 2 diabetes. Diseases associated include familial dyskinesia with facial myokymia.
CDKN2B	Cyclin Dependent Kinase Inhibitor 2B	9	It encodes a cyclin-dependent kinase inhibitor, which forms a complex with CDK4 or CDK6, and prevents the activation of the CDK kinases, thus the encoded protein functions as a cell growth regulator that controls cell cycle G1 progression. Diseases associated include adult acute lymphocytic leukemia and scrotal carcinoma.
IDE	Insulin Degrading Enzyme	10	It encodes a zinc metallopeptidase that degrades intracellular insulin, and thereby terminates insulins activity, as well as participating in intercellular peptide signalling by degrading diverse peptides such as glucagon, amylin, bradykinin, and kallidin. The preferential affinity of this enzyme for insulin results in insulin-mediated inhibition of the degradation of other peptides such as beta-amyloid. Deficiencies in this protein's function are associated with Alzheimer's disease and type 2 diabetes mellitus but mutations in this gene have not been shown to be causitive for these diseases.
MTNR1B	Melatonin Receptor 1B	11	It encodes a high affinity form of a receptor for melatonin and it is likely to mediate the reproductive and circadian actions of melatonin. It is widely distributed, with high concentrations in the brain and in the retina. It is thought to participate in light-dependent functions in the retina and may be involved in the neurobiological effects of melatonin. Diseases associated include noninsulin-dependent diabetes mellitus and idiopathic scoliosis.
TSPAN8	Tetraspanin 8	12	It encodes a transmembrane glycoprotein that mediates signal transduction events that play a role in the regulation of cell development, activation, growth and motility. This gene is expressed in different carcinomas. Diseases associated include annular pancreas and autosomal recessive nonsyndromic deafness 3.

Table 1. Overview of genes studied in association of Epigenetics with T2D and glycemic traits (Database URL: www.genecards.org) [146] (*continued*)

Gene	Alias	Chr	Function
APN	Aminopeptidase N	15	It encodes an aminopeptidase of broad specificity which plays a role in the final digestion of peptides generated from hydrolysis of proteins by gastric and pancreatic proteases, as well as in the angiogenesis and promote cholesterol crystallization. It also participates in the processing of various peptides including peptide hormones, such as angiotensin III and IV, neuropeptides, and chemokines. It may be involved the cleavage of peptides bound to major histocompatibility complex class II molecules of antigen presenting cells. Defects in this gene appear to be a cause of various types of leukemia or lymphoma.
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A	9	It encodes a protein that acts as a tumor suppressor, capable of inducing cell cycle arrest in G1 and G2 phases. Its loss has been shown to be a significant event in a number of cancer types. Associated d diseases include Melanoma-Pancreatic Cancer Syndrome and Melanoma-Astrocytoma Syndrome.
CAV1	Caveolin 1	7	It encodes a scaffolding protein which is a main component of the caveolae plasma membranes found in most cell types. It is nvolved in the costimulatory signal essential for T-cell receptor (TCR)-mediated T-cell activation. It is also considered a tumor suppressor gene candidate and a negative regulator of the Ras-p42/44 MAP kinase cascade. Mutations in this gene have been associated with Berardinelli-Seip congenital lipodystrophy, congenital cataracts.
WFS1	Wolfram Syndrome 1 (Wolframin)	4	It encodes a transmembrane glycoprotein, which is located primarily in the endoplasmic reticulum and ubiquitously expressed with highest levels in brain, pancreas, heart, and insulinoma beta-cell lines. It participates in the regulation of cellular Ca(2+) homeostasis, at least partly, by modulating the filling state of the endoplasmic reticulum Ca(2+) store. Mutations in this gene are associated with Wolfram syndrome, also called DIDMOAD (Diabetes Insipidus, Diabetes Mellitus, Optic Atrophy, and Deafness), an autosomal recessive disorder. The disease affects the brain and central nervous system.
MALT1	Mucosa Associated Lymphoid Tissue Lymphoma Translocation Protein 1	18	It encodes an aminopeptidase that catalyzes the removal of amino acids from the amino terminus of proteins and peptides. It may play a role in NF-kappaB activation in a BCL10-induced manner. Diseases associated with MALT1 include Immunodeficiency and Mucosa-Associated Lymphoid Tissue Lymphoma.
FTO	Alpha-Ketoglutarate Dependent Dioxygenase	16	It encodes an enzyme that repairs alkylated DNA and RNA by oxidative demethylation. It contributes to the regulation of the global metabolic rate, energy expenditure and energy homeostasis. In particular, it is involved in the regulation of thermogenesis and the control of adipocyte differentiation into brown or white fat cells. Diseases associated include growth retardation, developmental delay, facial dysmorphism and obesity.

Table 1. Overview of genes studied in association of Epigenetics with T2D and glycemic traits (Database URL: www.genecards.org) [146] (*continued*)

Gene	Alias	Chr	Function
ZNF668	Zinc Finger Protein 668	16	It encodes a protein that may be involved in transcriptional regulation. It is related to DNA binding transcription factor activity.
HSPA2	Heat Shock Protein Family A (Hsp70) Member 2	14	It encodes a chaperone protein implicated in a wide variety of cellular processes, including protection of the proteome from stress, the protein quality control system, by ensuring the correct folding of proteins, the re-folding of misfolded proteins and controlling the targeting of proteins for subsequent degradation, and the transport of newly synthesized polypeptides. It also plays a role in spermatogenesis. Diseases associated include inflammatory bowel disease and papillary cystadenocarcinoma. Heat Shock Protein Family A (Hsp70) Member 2
CD320	CD320 Molecule	19	It encodes a receptor for transcobalamin saturated with cobalamin (TCbl), that is expressed at the cell surface. It mediates the cellular uptake of transcobalamin bound cobalamin (vitamin B12), and is involved in B-cell proliferation and immunoglobulin secretion. Mutations in this gene are associated with methylmalonic aciduria.
HNF1A	SFT2 Domain Containing 3	2	It encodes a protein that may be involved in fusion of retrograde transport vesicles derived from an endocytic compartment with the Golgi complex.
TWIST1	Twist Family BHLH Transcription Factor 1	7	It encodes a transcription factor that plays an important role in embryonic development, regulating the transcription of genes involved in neural crest differentiation and brown fat metabolism. It also represses the expression of proinflammatory cytokines such as TNFA and IL1B, and the activity of the circadian transcriptional activator: NPAS2-ARNTL/BMAL1 heterodimer. it is involved in the osteoblast differentiation. Represses Mutations in this gene cause Saethre-Chotzen syndrome and Craniosynostosis 1.
MYO5A	Myosin VA	15	It encodes a class of actin-based motor proteins involved in cytoplasmic vesicle transport and anchorage, spindle-pole alignment and mRNA translocation, and it is abundant in melanocytes and nerve cells. Mutations in this gene cause Griscelli syndrome type-1 (GS1), Griscelli syndrome type-3 (GS3) and neuroectodermal melanolysosomal disease, or Elejalde disease.
MAPK1	Mitogen-Activated Protein Kinase 1	22	It encodes a member of the MAP kinase family. The MAPK cascade plays also a role in initiation and regulation of meiosis, mitosis, and postmitotic functions in differentiated cells by phosphorylating a number of transcription factors. The activation of this kinase requires its phosphorylation by upstream kinases. Associated diseases include Chromosome 22Q11.2 Deletion Syndrome, Distal and Pertussis.

Table 1. Overview of genes studied in association of Epigenetics with T2D and glycemic traits (Database URL: www.genecards.org) [146] (*continued*)

Gene	Alias	Chr	Function
MYO18B	Myosin XVIIIB	22	It encodes a protein which may be involved in intracellular trafficking of the muscle cell when in the cytoplasm, whereas entering the nucleus, may be involved in the regulation of muscle specific genes. May play a role in the control of tumor development and progression. Mutations in this gene are associated with lung cancer.
HOXC6	Homeobox C6	12	It encodes a member of the homeobox family of transcription factors that play an important role in morphogenesis in all multicellular organisms. This is a sequence-specific transcription factor which is part of a developmental regulatory system that provides cells with specific positional identities on the anterior-posterior axis.
PRKAB1	Protein Kinase, AMP-Activated, Beta 1 Non-Catalytic Subunit	12	It encodes a regulatory subunit of the AMP-activated protein kinase (AMPK), which is an energy sensor protein kinase that plays a key role in regulating cellular energy metabolism. In response to reduction of intracellular ATP levels, AMPK activates energy-producing pathways and inhibits energy-consuming processes: inhibits protein, carbohydrate and lipid biosynthesis, as well as cell growth and proliferation. Associated diseases include Body Mass Index Quantitative Trait Locus 11.
NF-KB	Nuclear Factor Kappa B	4	It encodes a transcription regulator that is activated by various intra- and extra-cellular stimuli such as cytokines, oxidant-free radicals, ultraviolet irradiation, and bacterial or viral products. It translocates into the nucleus and stimulates the expression of genes involved in a wide variety of biological functions. It is a homo- or heterodimeric complex formed by the Rel-like domain-containing proteins RELA/p65, RELB, NFKB1/p105, NFKB1/p50, REL and NFKB2/p52 and the heterodimeric p65-p50 complex. Associated diseases include Immunodeficiency.
RELA	RELA Proto-Oncogene, NF-KB Subunit		It encodes the NF-Kappa-B transcription factor P65. The most abundant form of NF-kappa-B is NFKB1 complexed with the product of this gene, RELA. Associated diseases include Ependymoma and Reticuloendotheliosis. Can modulate chromatin function through deacetylation of histones and can promote alterations in the methylation of histones and DNA, leading to transcriptional repression.
Sirt1	Sirtuin 1	10	It encodes a NAD-dependent protein deacetylase that links transcriptional regulation directly to intracellular energetics and participates in the coordination of several separated cellular functions such as cell cycle, response to DNA damage, metabolism, apoptosis and autophagy. Associated diseases include Aging and Ovarian Endodermal Sinus Tumor.
IL-1A	Interleukin 1 Alpha	2	It encodes cytokine member of the interleukin 1 cytokine family and it is involved in various immune responses, inflammatory processes, and hematopoiesis. It is produced by activated macrophages and released in response to cell injury, and thus induces apoptosis. Polymorphism may be associated with rheumatoid arthritis and Alzheimer's disease, Irritant Dermatitis and Cholesteatoma of middle ear.

Table 1. Overview of genes studied in association of Epigenetics with T2D and glycemic traits (Database URL: www.genecards.org) [146] (*continued*)

Gene	Alias	Chr	Function
PTEN	Phosphatase And Tensin Homolog	10	It encodes a phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase which functions as a tumor suppressor by negatively regulating AKT/PKB signaling pathway, and that is commonly lost in cancer. The isoform alpha plays a role in mitochondrial energetic metabolism by promoting COX activity and ATP production. It may also be a negative regulator of insulin signaling and glucose metabolism in adipose tissue. Related diseases are dysplastic gangliocytoma of the cerebellum, and macrocephaly multiple lipomas and hemangiomata.
TNF-α	Tumor Necrosis Factor	6	It encodes a proinflammatory cytokine that belongs to the tumor necrosis factor (TNF) superfamily. It is mainly secreted by macrophages and can induce cell death of certain tumor cell lines. It is involved in the regulation of a wide spectrum of biological processes including cell proliferation, differentiation, apoptosis, lipid metabolism, and coagulation. It is implicated in a variety of diseases, including autoimmune diseases, insulin resistance, and cancer. Diseases associated include asthma and malaria.
COX-2	Cyclooxygenase 2	1	It encodes cyclooxygenase isoform 2, also known as prostaglandin-endoperoxide synthase 2 (PTGS2). It is an inducible isozyme, which converts arachidonate to prostaglandin H2 (PGH2), during the prostanoid synthesis. It is constitutively expressed in some tissues in physiological conditions, such as the endothelium, kidney and brain, and in pathological conditions, such as in cancer. It is responsible for production of inflammatory prostaglandins. Up-regulation of PTGS2 is also associated with increased cell adhesion, phenotypic changes, resistance to apoptosis and tumor angiogenesis.
IL-8	C-X-C Motif Chemokine Ligand 8 ; interleukin-8	4	It encodes a protein that is secreted primarily by neutrophils, where it serves as a chemotactic factor by guiding the neutrophils to the site of infection. It may also be released from several cell types in response to an inflammatory stimulus. Is also attracts basophils and T-cells, but not monocytes. Associated diseases include Bronchiolitis and Extrinsic Allergic Alveolitis.
UNC13B	Unc-13 Homolog B	9	This gene is expressed in the kidney cortical epithelial cells and it is upregulated by hyperglycemia. It contains three C2 domains and a diacylglycerol-binding C1 domain. Hyperglycemia increases the levels of diacylglycerol, which has been shown to induce apoptosis, thus contributing to the renal cell complications of hyperglycemia. Associated diseases include Hyperglycemia and Hemophagocytic Lymphohistiocytosis
PAI-1	Serpin Peptidase Inhibitor, Clade E (Nexin, Plasminogen Activator Inhibitor Type 1), Member 1	7	It encodes a member of the serine proteinase inhibitor (serpin) superfamily, which is the principal inhibitor of tissue plasminogen activator and urokinase, and hence is an inhibitor of fibrinolysis. Defects in this gene are the cause of plasminogen activator inhibitor-1 deficiency (PAI-1 deficiency), and high concentrations of the gene product are associated with thrombophilia. Associated diseases include Plasminogen Activator Inhibitor-1 Deficiency and Complete Plasminogen Activator Inhibitor 1 Deficiency.

Table 1. Overview of genes studied in association of Epigenetics with T2D and glycemic traits (Database URL: www.genecards.org) [146] (*continued*)

Gene	Alias	Chr	Function
RAGE	Receptor For Advanced Glycation End-Products Variant 20	6	It encodes a receptor for glycosylation end products (AGE). AGE accumulate in vascular tissue in aging and at an accelerated rate in diabetes. Besides AGE, it also interacts with other molecules implicated in homeostasis, development, and inflammation, and certain diseases, such as diabetes and Alzheimer's disease. Acts as a mediator of both acute and chronic vascular inflammation in conditions such as atherosclerosis and in particular as a complication of diabetes. Associated diseases include Diabetic Angiopathy and Thymic Hyperplasia.

REFERENCES

1. Zghebi SS, Steinke DT, Carr MJ, et al. Examining trends in type 2 diabetes incidence, prevalence and mortality in the UK between 2004 and 2014. Diabetes Obes Metab 2017;19(11):1537-45.
2. Wild S, Roglic G, Green A, et al. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 2004;27(5):1047-53.
3. Federation ID. IDF Diabetes Atlas, 8th edn. Brussels, Belgium: International Diabetes Federation, 2017. 2017.
4. Animaw W, Seyoum Y. Increasing prevalence of diabetes mellitus in a developing country and its related factors. PLoS One 2017;12(11):e0187670.
5. Guthrie RA, Guthrie DW. Pathophysiology of diabetes mellitus. Crit Care Nurs Q 2004;27(2):113-25.
6. American Diabetes A. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2018. Diabetes Care 2018;41(Suppl 1):S13-S27.
7. Bonora E, Tuomilehto J. The Pros and Cons of Diagnosing Diabetes With A1C. Diabetes Care 2011;34:S184-S90.
8. Gilbert ER, Liu D. Epigenetics: the missing link to understanding beta-cell dysfunction in the pathogenesis of type 2 diabetes. Epigenetics 2012;7(8):841-52.
9. Volkmar M, Dedeurwaerder S, Cunha DA, et al. DNA methylation profiling identifies epigenetic dysregulation in pancreatic islets from type 2 diabetic patients. Embo J 2012;31(6):1405-26.
10. Bonora E, Tuomilehto J. The pros and cons of diagnosing diabetes with A1C. Diabetes Care 2011;34 Suppl 2:S184-90.
11. International Expert C. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. Diabetes Care 2009;32(7):1327-34.
12. Zhuo X, Zhang P, Selvin E, et al. Alternative HbA1c cutoffs to identify high-risk adults for diabetes prevention: a cost-effectiveness perspective. Am J Prev Med 2012;42(4):374-81.
13. Tabak AG, Jokela M, Akbaraly TN, et al. Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. Lancet 2009;373(9682):2215-21.
14. Weisenberger DJ, Campan M, Long TI, et al. Analysis of repetitive element DNA methylation by MethyLight. Nucleic Acids Res 2005;33(21):6823-36.
15. Muka T, Nano J, Voortman T, et al. The role of global and regional DNA methylation and histone modifications in glycemic traits and type 2 diabetes: A systematic review. Nutr Metab Cardiovas 2016;26(7):553-66.
16. Ribel-Madsen R, Fraga MF, Jacobsen S, et al. Genome-wide analysis of DNA methylation differences in muscle and fat from monozygotic twins discordant for type 2 diabetes. PLoS One 2012;7(12):e51302.
17. Keller M, Kralisch S, Rohde K, et al. Global DNA methylation levels in human adipose tissue are related to fat distribution and glucose homeostasis. Diabetologia 2014;57(11):2374-83.
18. Simar D, Versteyhe S, Donkin I, et al. DNA methylation is altered in B and NK lymphocytes in obese and type 2 diabetic human. Metabolism 2014;63(9):1188-97.
19. Pearce MS, McConnell JC, Potter C, et al. Global LINE-1 DNA methylation is associated with blood glycaemic and lipid profiles. Int J Epidemiol 2012;41(1):210-17.
20. Canivell S, Ruano EG, Siso-Almirall A, et al. Gastric Inhibitory Polypeptide

Receptor Methylation in Newly Diagnosed, Drug-Naive Patients with Type 2 Diabetes: A Case-Control Study. Plos One 2013;8(9).
21. Gillberg L, Jacobsen S, Ribel-Madsen R, et al. Does DNA methylation of PPARGC1A influence insulin action in first degree relatives of patients with type 2 diabetes? Diabetologia 2012;55:S130-S30.
22. Hall E, Dekker Nitert M, Volkov P, et al. The effects of high glucose exposure on global gene expression and DNA methylation in human pancreatic islets. Mol Cell Endocrinol 2017.
23. Su SY, Zhu HD, Xu XJ, et al. DNA Methylation of the LY86 Gene is Associated With Obesity, Insulin Resistance, and Inflammation. Twin Res Hum Genet 2014;17(3):183-91.
24. Bouchard L, Hivert MF, Guay SP, et al. Placental adiponectin gene DNA methylation levels are associated with mothers' blood glucose concentration. Diabetes 2012;61(5):1272-80.
25. Xie XM, Gao HJ, Zeng WJ, et al. Placental DNA methylation of peroxisome-proliferator-activated receptor-gamma co-activator-1 alpha promoter is associated with maternal gestational glucose level. Clin Sci 2015;129(4):385-94.
26. Desgagne V, Hivert MF, St-Pierre J, et al. Epigenetic dysregulation of the IGF system in placenta of newborns exposed to maternal impaired glucose tolerance. Epigenomics 2014;6(2):193-207.
27. Hidalgo B, Irvin MR, Sha J, et al. Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network study. Diabetes 2014;63(2):801-7.
28. Kriebel J, Herder C, Rathmann W, et al. Association between DNA Methylation in Whole Blood and Measures of Glucose Metabolism: KORA F4 Study. Plos One 2016;11(3).
29. Lempradl A, Pospisilik JA, Penninger JM. MODES OF TRANSCRIPTIONAL REGULATION Exploring the emerging complexity in transcriptional regulation of energy homeostasis. Nat Rev Genet 2015;16(11):665-81.
30. Chambers JC, Loh M, Lehne B, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. Lancet Diabetes Endocrinol 2015;3(7):526-34.
31. Cha-Molstad H, Saxena G, Chen JQ, et al. Glucose-stimulated Expression of Txnip Is Mediated by Carbohydrate Response Element-binding Protein, p300, and Histone H4 Acetylation in Pancreatic Beta Cells. J Biol Chem 2009;284(25):16898-905.
32. Bompada P, Atac D, Luan C, et al. Histone acetylation of glucose-induced thioredoxin-interacting protein gene expression in pancreatic islets. Int J Biochem Cell B 2016;81:82-91.
33. Samuel VT, Shulman GI. The pathogenesis of insulin resistance: integrating signaling pathways and substrate flux. J Clin Invest 2016;126(1):12-22.
34. Piyathilake CJ, Badiga S, Alvarez RD, et al. A Lower Degree of PBMC L1 Methylation Is Associated with Excess Body Weight and Higher HOMA-IR in the Presence of Lower Concentrations of Plasma Folate. Plos One 2013;8(1).
35. Zhao JY, Goldberg J, Bremner JD, et al. Global DNA Methylation Is Associated With Insulin Resistance A Monozygotic Twin Study. Diabetes 2012;61(2):542-46.
36. Simar D, Versteyhe S, Donkin I, et al. DNA methylation is altered in B and NK lymphocytes in obese and type 2

diabetic human. Metabolism-Clinical and Experimental 2014;63(9):1188-97.
37. Manning AK, Hivert MF, Scott RA, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet 2012;44(6):659-U81.
38. Kulkarni H, Kos MZ, Neary J, et al. Novel epigenetic determinants of type 2 diabetes in Mexican-American families. Hum Mol Genet 2015;24(18):5330-44.
39. Hidalgo B, Irvin MR, Sha J, et al. Epigenome-Wide Association Study of Fasting Measures of Glucose, Insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network Study. Diabetes 2014;63(2):801-07.
40. Das S, Dey JK, Prabhu N, et al. Association Between 5-HTR2C-759C/T (rs3813929) and-697G/C (rs518147) Gene Polymorphisms and Risperidone-Induced Insulin Resistance Syndrome in an Indian Population (Retraction of, vol 57, 2017) (Retraction of Vol 57, 10.1002/JCPH.1012, 2017). J Clin Pharmacol 2018;58(3):399-99.
41. Gillberg L, Jacobsen S, Ribel-Madsen R, et al. Does DNA Methylation of PPARGC1A Influence Insulin Action in First Degree Relatives of Patients with Type 2 Diabetes? Plos One 2013;8(3).
42. Maier B, Thimme W, Kallischnigg G, et al. Does diabetes mellitus explain the higher hospital mortality of women with acute myocardial infarction? Results from the Berlin Myocardial Infarction Registry. J Investig Med 2006;54(3):143-51.
43. Kabra DG, Gupta J, Tikoo K. Insulin induced alteration in post-translational modifications of histone H3 under a hyperglycemic condition in L6 skeletal muscle myoblasts. Biochim Biophys Acta 2009;1792(6):574-83.
44. Uusitupa M. Gene-diet interaction in relation to the prevention of obesity and type 2 diabetes: evidence from the Finnish Diabetes Prevention Study. Nutr Metab Cardiovasc Dis 2005;15(3):225-33.
45. Ling C, Groop L. Epigenetics: a molecular link between environmental factors and type 2 diabetes. Diabetes 2009;58(12):2718-25.
46. Zhang H, Cai X, Yi B, et al. Correlation of CTGF gene promoter methylation with CTGF expression in type 2 diabetes mellitus with or without nephropathy. Mol Med Rep 2014;9(6):2138-44.
47. Kato S, Lindholm B, Stenvinkel P, et al. DNA hypermethylation and inflammatory markers in incident Japanese dialysis patients. Nephron Extra 2012;2(1):159-68.
48. Luttmer R, Spijkerman AM, Kok RM, et al. Metabolic syndrome components are associated with DNA hypomethylation. Obes Res Clin Pract 2013;7(2):e106-e15.
49. Gu HF, Gu T, Hilding A, et al. Evaluation of IGFBP-7 DNA methylation changes and serum protein variation in Swedish subjects with and without type 2 diabetes. Clin Epigenetics 2013;5(1):20.
50. Gu T, Gu HF, Hilding A, et al. Increased DNA methylation levels of the insulin-like growth factor binding protein 1 gene are associated with type 2 diabetes in Swedish men. Clin Epigenetics 2013;5(1):21.
51. Remely M, Aumueller E, Jahn D, et al. Microbiota and epigenetic regulation of inflammatory mediators in type 2 diabetes and obesity. Benef Microbes 2014;5(1):33-43.
52. Seman NA, Mohamud WN, Ostenson CG, et al. Increased DNA methylation of the SLC30A8 gene promoter is associated with type 2 diabetes in a

Malay population. Clin Epigenetics 2015;7:30.
53. Tang L, Ye H, Hong Q, et al. Elevated CpG island methylation of GCK gene predicts the risk of type 2 diabetes in Chinese males. Gene 2014;547(2):329-33.
54. Zou L, Yan S, Guan X, et al. Hypermethylation of the PRKCZ Gene in Type 2 Diabetes Mellitus. J Diabetes Res 2013;2013:721493.
55. Ling C, Del Guerra S, Lupi R, et al. Epigenetic regulation of PPARGC1A in human type 2 diabetic islets and effect on insulin secretion. Diabetologia 2008;51(4):615-22.
56. Yang BT, Dayeh TA, Volkov PA, et al. Increased DNA methylation and decreased expression of PDX-1 in pancreatic islets from patients with type 2 diabetes. Mol Endocrinol 2012;26(7):1203-12.
57. Yang BT, Dayeh TA, Kirkpatrick CL, et al. Insulin promoter DNA methylation correlates negatively with insulin gene expression and positively with HbA(1c) levels in human pancreatic islets. Diabetologia 2011;54(2):360-7.
58. Jun Z, Wangqiang Z, Yulei D, et al. [Correlation between type 2 diabetes and DNA methylation and mRNA expression of APN in abdominal adipose tissues in Xinjiang Uygur population]. Yi Chuan 2015;37(3):269-75.
59. Canivell S, Ruano EG, Siso-Almirall A, et al. Gastric inhibitory polypeptide receptor methylation in newly diagnosed, drug-naive patients with type 2 diabetes: a case-control study. PLoS One 2013;8(9):e75474.
60. Cheng J, Tang L, Hong Q, et al. Investigation into the promoter DNA methylation of three genes (CAMK1D, CRY2 and CALM2) in the peripheral blood of patients with type 2 diabetes. Exp Ther Med 2014;8(2):579-84.
61. Liu ZH, Chen LL, Deng XL, et al. Methylation status of CpG sites in the MCP-1 promoter is correlated to serum MCP-1 in Type 2 diabetes. J Endocrinol Invest 2012;35(6):585-9.
62. Tros F, Meirhaeghe A, Hadjadj S, et al. Hypomethylation of the promoter of the catalytic subunit of protein phosphatase 2A in response to hyperglycemia. Physiol Rep 2014;2(7).
63. Kulkarni SS, Salehzadeh F, Fritz T, et al. Mitochondrial regulators of fatty acid metabolism reflect metabolic dysfunction in type 2 diabetes mellitus. Metabolism 2012;61(2):175-85.
64. Canivell S, Ruano EG, Siso-Almirall A, et al. Differential methylation of TCF7L2 promoter in peripheral blood DNA in newly diagnosed, drug-naive patients with type 2 diabetes. PLoS One 2014;9(6):e99310.
65. Ma J, Cheng J, Wang L, et al. No association between IRS1 promoter methylation and type 2 diabetes. Mol Med Rep 2013;8(3):949-53.
66. Hall E, Dayeh T, Kirkpatrick CL, et al. DNA methylation of the glucagon-like peptide 1 receptor (GLP1R) in human pancreatic islets. BMC Med Genet 2013;14:76.
67. Gillberg L, Jacobsen SC, Ribel-Madsen R, et al. Does DNA methylation of PPARGC1A influence insulin action in first degree relatives of patients with type 2 diabetes? PLoS One 2013;8(3):e58384.
68. Dayeh T, Volkov P, Salo S, et al. Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. PLoS Genet 2014;10(3):e1004160.
69. Kulkarni H, Kos MZ, Neary J, et al. Novel epigenetic determinants of type 2 diabetes in Mexican-American fami-

lies. Hum Mol Genet 2015;24(18):5330-44.
70. Toperoff G, Aran D, Kark JD, et al. Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood. Hum Mol Genet 2012;21(2):371-83.
71. Strahl BD, Allis CD. The language of covalent histone modifications. Nature 2000;403(6765):41-5.
72. Sims RJ, 3rd, Nishioka K, Reinberg D. Histone lysine methylation: a signature for chromatin function. Trends Genet 2003;19(11):629-39.
73. Paneni F, Costantino S, Battista R, et al. Adverse epigenetic signatures by histone methyltransferase Set7 contribute to vascular dysfunction in patients with type 2 diabetes mellitus. Circ Cardiovasc Genet 2015;8(1):150-8.
74. Miao F, Wu X, Zhang L, et al. Genome-wide analysis of histone lysine methylation variations caused by diabetic conditions in human monocytes. J Biol Chem 2007;282(18):13854-63.
75. Hou C, Zhao M, Li X, et al. [Histone H3 acetylation of tumor necrosis factor-alpha and cyclooxygenase-2 in patients with type 2 diabetes]. Zhonghua Yi Xue Za Zhi 2011;91(26):1805-8.
76. Roglic G. WHO Global report on diabetes: A summary. International Journal of Noncommunicable Diseases 2016;1(1):3-8.
77. Muka T, Imo D, Jaspers L, et al. The global impact of non-communicable diseases on healthcare spending and national income: a systematic review. European Journal of Epidemiology 2015;30(4):251-77.
78. Noble D, Mathur R, Dent T, et al. Risk models and scores for type 2 diabetes: systematic review. BMJ 2011;343:d7163.
79. Meigs JB, Shrader P, Sullivan LM, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med 2008;359(21):2208-19.
80. Talmud PJ, Hingorani AD, Cooper JA, et al. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. BMJ 2010;340:b4838.
81. Stancakova A, Laakso M. Genetics of Type 2 Diabetes. Endocr Dev 2016;31:203-20.
82. Martin-Nunez GM, Rubio-Martin E, Cabrera-Mulero R, et al. Type 2 diabetes mellitus in relation to global LINE-1 DNA methylation in peripheral blood: a cohort study. Epigenetics 2014;9(10):1322-8.
83. Dayeh T, Tuomi T, Almgren P, et al. DNA methylation of loci within ABCG1 and PHOSPHO1 in blood DNA is associated with future type 2 diabetes risk. Epigenetics 2016;11(7):482-8.
84. Andersen GS, Thybo T, Cederberg H, et al. The DEXLIFE study methods: identifying novel candidate biomarkers that predict progression to type 2 diabetes in high risk individuals. Diabetes Res Clin Pract 2014;106(2):383-9.
85. Kannel WB, McGee DL. Diabetes and cardiovascular disease. The Framingham study. JAMA 1979;241(19):2035-8.
86. Kim JA, Koh KK, Quon MJ. The union of vascular and metabolic actions of insulin in sickness and in health. Arterioscl Throm Vas 2005;25(5):889-91.
87. Paneni F, Mocharla P, Akhmedov A, et al. Gene silencing of the mitochondrial adaptor p66(Shc) suppresses vascular hyperglycemic memory in diabetes. Circ Res 2012;111(3):278-89.
88. Abi Khalil C. The emerging role of epigenetics in cardiovascular disease. Ther Adv Chronic Dis 2014;5(4):178-87.

89. Monkemann H, De Vriese AS, Blom HJ, et al. Early molecular events in the development of the diabetic cardiomyopathy. Amino Acids 2002;23(1-3):331-6.
90. Deodati A, Inzaghi E, Liguori A, et al. IGF2 Methylation Is Associated with Lipid Profile in Obese Children. Horm Res Paediat 2013;79(6):361-67.
91. Arner P, Sahlqvist AS, Sinha I, et al. The epigenetic signature of systemic insulin resistance in obese women (vol 59, pg 2393, 2016). Diabetologia 2016;59(12):2728-28.
92. Lou XD, Wang HD, Xia SJ, et al. Effects of Resveratrol on the Expression and DNA Methylation of Cytokine Genes in Diabetic Rat Aortas. Arch Immunol Ther Ex 2014;62(4):329-40.
93. Madsbad S. Liraglutide Effect and Action in Diabetes (LEAD™) trial. Expert Review of Endocrinology & Metabolism 2009;4(2):119-29.
94. Hall E, Dayeh T, Kirkpatrick CL, et al. DNA methylation of the glucagon-like peptide 1 receptor (GLP1R) in human pancreatic islets. Bmc Med Genet 2013;14.
95. Gaikwad AB, Gupta J, Tikoo K. Epigenetic changes and alteration of Fbn1 and Col3A1 gene expression under hyperglycaemic and hyperinsulinaemic conditions. Biochem J 2010;432(2):333-41.
96. Sathishkumar C, Prabu P, Balakumar M, et al. Augmentation of histone deacetylase 3 (HDAC3) epigenetic signature at the interface of proinflammation and insulin resistance in patients with type 2 diabetes. Clin Epigenetics 2016;8.
97. Pirola L, Balcerczyk A, Tothill RW, et al. Genome-wide analysis distinguishes hyperglycemia regulated epigenetic signatures of primary vascular cells. Genome Research 2011;21(10):1601-15.
98. Brasacchio D, Okabe J, Tikellis C, et al. Hyperglycemia Induces a Dynamic Cooperativity of Histone Methylase and Demethylase Enzymes Associated With Gene-Activating Epigenetic Marks That Coexist on the Lysine Tail. Diabetes 2009;58(5):1229-36.
99. Miao F, Chen Z, Genuth S, et al. Evaluating the role of epigenetic histone modifications in the metabolic memory of type 1 diabetes. Diabetes 2014;63(5):1748-62.
100. Jones CA, Krolewski AS, Rogus J, et al. Epidemic of end-stage renal disease in people with diabetes in the United States population: Do we know the cause? Kidney Int 2005;67(5):1684-91.
101. Kato M, Natarajan R. Diabetic nephropathy-emerging epigenetic mechanisms. Nat Rev Nephrol 2014;10(9):517-30.
102. Reddy MA, Park JT, Natarajan R. Epigenetic Modifications in the Pathogenesis of Diabetic Nephropathy. Semin Nephrol 2013;33(4):341-53.
103. Bell CG, Teschendorff AE, Rakyan VK, et al. Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. Bmc Med Genomics 2010;3.
104. Brennan EP, Ehrich M, O'Donovan H, et al. DNA methylation profiling in cell models of diabetic nephropathy. Epigenetics 2010;5(5):396-401.
105. Sapienza C, Lee J, Powell J, et al. DNA methylation profiling identifies epigenetic differences between diabetes patients with ESRD and diabetes patients without nephropathy. Epigenetics 2011;6(1):20-28.
106. Smyth LJ, McKay GJ, Maxwell AP, et al. DNA hypermethylation and DNA hypomethylation is present at different loci in chronic kidney disease. Epigenetics 2014;9(3):366-76.

107. Yuan H, Reddy MA, Sun G, et al. Involvement of p300/CBP and epigenetic histone acetylation in TGF-beta1-mediated gene transcription in mesangial cells. Am J Physiol Renal Physiol 2013;304(5):F601-13.
108. Li Y, Reddy MA, Miao F, et al. Role of the histone H3 lysine 4 methyltransferase, SET7/9, in the regulation of NF-kappaB-dependent inflammatory genes. Relevance to diabetes and inflammation. J Biol Chem 2008;283(39):26771-81.
109. Reddy MA, Sumanth P, Lanting LT, et al. Losartan reverses permissive epigenetic changes in renal glomeruli of diabetic db/db mice. Kidney International 2014;85(2):362-73.
110. Kowluru RA. Diabetic retinopathy, metabolic memory and epigenetic modifications. Vision Res 2017;139:30-38.
111. Diabetes C, Complications Trial/ Epidemiology of Diabetes I, Complications Research G, et al. Retinopathy and nephropathy in patients with type 1 diabetes four years after a trial of intensive therapy. N Engl J Med 2000;342(6):381-9.
112. Tewari S, Zhong Q, Santos JM, et al. Mitochondria DNA replication and DNA methylation in the metabolic memory associated with continued progression of diabetic retinopathy. Invest Ophthalmol Vis Sci 2012;53(8):4881-8.
113. Mishra M, Kowluru RA. Epigenetic Modification of Mitochondrial DNA in the Development of Diabetic Retinopathy. Invest Ophthalmol Vis Sci 2015;56(9):5133-42.
114. Kowluru RA, Shan Y, Mishra M. Dynamic DNA methylation of matrix metalloproteinase-9 in the development of diabetic retinopathy. Lab Invest 2016;96(10):1040-9.
115. Mishra M, Zhong Q, Kowluru RA. Epigenetic modifications of Nrf2-mediated glutamate-cysteine ligase: implications for the development of diabetic retinopathy and the metabolic memory phenomenon associated with its continued progression. Free Radic Biol Med 2014;75:129-39.
116. Zhong Q, Kowluru RA. Epigenetic modification of Sod2 in the development of diabetic retinopathy and in the metabolic memory: role of histone methylation. Invest Ophthalmol Vis Sci 2013;54(1):244-50.
117. Kadiyala CS, Zheng L, Du Y, et al. Acetylation of retinal histones in diabetes increases inflammatory proteins: effects of minocycline and manipulation of histone acetyltransferase (HAT) and histone deacetylase (HDAC). J Biol Chem 2012;287(31):25869-80.
118. Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet 2009;10(5):295-304.
119. Zhong Q, Kowluru RA. Regulation of Matrix Metalloproteinase-9 by Epigenetic Modifications and the Development of Diabetic Retinopathy. Diabetes 2013;62(7):2559-68.
120. Reddy MA, Zhang EL, Natarajan R. Epigenetic mechanisms in diabetic complications and metabolic memory. Diabetologia 2015;58(3):443-55.
121. Zochodne DW. Diabetes mellitus and the peripheral nervous system: Manifestations and mechanisms. Muscle Nerve 2007;36(2):144-66.
122. Kim ES, Isoda F, Kurland I, et al. Glucose-Induced Metabolic Memory in Schwann Cells: Prevention by PPAR Agonists. Endocrinology 2013;154(9):3054-66.
123. Pande M, Hur J, Hong Y, et al. Transcriptional Profiling of Diabetic Neuropathy in the BKS db/db Mouse A Model of Type 2 Diabetes. Diabetes 2011;60(7):1981-89.

124. El-Osta A, Brasacchio D, Yao D, et al. Transient high glucose causes persistent epigenetic changes and altered gene expression during subsequent normoglycemia (vol 205, pg 2409, 2008). J Exp Med 2008;205(11):2683-83.
125. Yagihashi S, Mizukami H, Sugimoto K. Mechanism of diabetic neuropathy: Where are we now and where to go? J Diabetes Invest 2011;2(1):18-32.
126. Nakayama M, Bennett CJ, Hicks JL, et al. Hypermethylation of the human glutathione S-transferase-pi gene (GSTP1) CpG island is present in a subset of proliferative inflammatory atrophy lesions but not in normal or hyperplastic epithelium of the prostate - A detailed study using laser-capture-microdissection. Am J Pathol 2003;163(3):923-33.
127. Van Neste L, Herman JG, Otto G, et al. The epigenetic promise for prostate cancer diagnosis. Prostate 2012;72(11):1248-61.
128. Yegnasubramanian S, Kowalski J, Gonzalgo ML, et al. Hypermethylation of CpG islands in primary and metastatic human prostate cancer. Cancer Res 2004;64(6):1975-86.
129. Ellinger J, Bastian PJ, Jurgan T, et al. CpG island hypermethylation at multiple gene sites in diagnosis and prognosis of prostate cancer. Urology 2008;71(1):161-67.
130. Balana C, Carrato C, Ramirez JL, et al. Tumour and serum MGMT promoter methylation and protein expression in glioblastoma patients. Clin Transl Oncol 2011;13(9):677-85.
131. Gyparaki MT, Basdra EK, Papavassiliou AG. DNA methylation biomarkers as diagnostic and prognostic tools in colorectal cancer. J Mol Med 2013;91(11):1249-56.
132. Warren JD, Xiong W, Bunker AM, et al. Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. Bmc Medicine 2011;9.
133. Lofton-Day C, Model F, DeVos T, et al. DNA methylation biomarkers for blood-based colorectal cancer screening. Clinical Chemistry 2008;54(2):414-23.
134. Fraga MF, Ballestar E, Paz MF, et al. Epigenetic differences arise during the lifetime of monozygotic twins. P Natl Acad Sci USA 2005;102(30):10604-09.
135. Rakyan VK, Beyan H, Down TA, et al. Identification of Type 1 Diabetes-Associated DNA Methylation Variable Positions That Precede Disease Diagnosis. Plos Genetics 2011;7(9).
136. Ahlqvist E, Storm P, Karajamaki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol 2018;6(5):361-69.
137. Zhang K, Lin G, Han Y, et al. Circulating unmethylated insulin DNA as a potential non-invasive biomarker of beta cell death in type 1 Diabetes: a review and future prospect. Clin Epigenetics 2017;9:44.
138. Husseiny MI, Kaye A, Zebadua E, et al. Tissue-specific methylation of human insulin gene and PCR assay for monitoring beta cell death. PLoS One 2014;9(4):e94591.
139. Milani L, Lundmark A, Kiialainen A, et al. DNA methylation for subtype classification and prediction of treatment outcome in patients with childhood acute lymphoblastic leukemia. Blood 2010;115(6):1214-25.
140. Dedeurwaerder S, Desmedt C, Calonne E, et al. DNA methylation profiling reveals a predominant immune component in breast cancers. Embo Mol Med 2011;3(12):726-41.
141. Huang RL, Chen HJ, Chen LY, et al. Epigenetic loss of heparan sulfate 3-O-

sulfation sensitizes ovarian carcinoma to oncogenic signals and predicts prognosis. Int J Cancer 2018.
142. Liu Y, Tan Q, Liu F. Differentially methylated circulating DNA: A novel biomarker to monitor beta cell death. J Diabetes Complications 2018;32(3):349-53.
143. Bellin MD, Clark P, Usmani-Brown S, et al. Unmethylated Insulin DNA Is Elevated After Total Pancreatectomy With Islet Autotransplantation: Assessment of a Novel Beta Cell Marker. Am J Transplant 2017;17(4):1112-18.
144. Gong Z, Muzumdar RH. Pancreatic function, type 2 diabetes, and metabolism in aging. Int J Endocrinol 2012;2012:320482.
145. Record Owner NLM. Full accounting of diabetes and pre-diabetes in the U.S. population in 1988-1994 and 2005-2006.
146. Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging-Us 2018;10(4):573-91.
147. Safran M, Dalah I, Alexander J, et al. GeneCards Version 3: the human gene integrator. Database (Oxford) 2010;2010:baq020.

Chapter 2.2

A Systematic review of sexually dimorphic DNA-methylation in cardiometabolic health

Eralda Asllanaj, **Carolina Ochoa-Rosales***, Xiaofang Zhang*, Jana Nano, Wichor Bramer, Eliana Portilla, Kim Braun, Valentina González-Jaramillo, Wolfgang Ahrens, Arfan Ikram, Mohsen Ghanbari, Trudy Voortman, Oscar Franco, Taulant Muka, Marija Glisic.

Sexually dimorphic DNA-methylation in cardiometabolic health: A systematic review. *Maturitas 2020;135:6-26*.

ABSTRACT

Sex is a major determinant of cardiometabolic risk. DNA methylation (DNAm), an important epigenetic mechanism that differs between sexes, has been associated with cardiometabolic diseases. Therefore, we aimed to systematically review studies in adults investigating sex-specific associations of DNAm with intermediate cardiometabolic traits and incident cardiovascular disease including stroke, myocardial infarction (MI) and coronary heart disease (CHD). Five bibliographic databases were searched from inception to 15 July 2019. We selected 35 articles (based on 30 unique studies) from 17,023 references identified; including a total of 14,020 participants from European, North American and Asian ancestries. Four studies reported sex differences between global DNAm and blood lipid levels and stroke risk. In 25 genome wide and candidate gene approach studies, DNAm at 31 gene sites were associated with sex differences in cardiometabolic diseases. The identified genes were *PLA2G7, BCL11A, KDM6A, LIPC, ABCG1, PLTP, CETP, ADD1, CNN1B, HOOK2, GFBP-7, PTPN1, GCK, PTX3, ABCG1, GALNT2, CDKN2B, APOE, CTH, GNASAS, INS, PON1, TCN2, CBS, AMT, KDMA6A, FTO, MAP3K13, CCDC8, MMP-2* and *ER-α*. Prioritized pathway connectivity analysis associated these genes with biological pathways such as vitamin B12 metabolism, statin pathway, plasma lipoprotein, plasma lipoprotein assembly, remodeling and clearance and cholesterol metabolism. Our findings suggest that DNAm might be a promising molecular strategy for understanding sex differences in the pathophysiology of cardiometabolic diseases and that future studies should investigate the effects of sex on epigenetic mechanisms in cardiometabolic risk. In addition, we emphasize the gap between the translational potential and the clinical utilization of cardiometabolic epigenetics.

INTRODUCTION

Cardiometabolic diseases include cardiovascular diseases (CVD), type 2 diabetes (T2D) and their associated risk factors including components of the metabolic syndrome and obesity[1]. Aging is associated with development of unfavorable cardiometabolic profile which in large contributes to increased incidence of major cardiovascular events and mortality[2]. Intermediate cardiometabolic risk factors are unequally distributed among sexes, and sex differences are also described in cardiometabolic diseases prevalence, clinical characteristics and prognosis[3, 4]·[5]. Generally, before menopause women have better cardiometabolic risk profiles than same aged men; however, this sex advantage gradually disappears with advancing age, particularly after menopause[6]. Mechanisms underlying sex differences in CVD have been extensively studied in the past two decades and signaling pathways including epigenetic modifications of the genome emerged as possible pathways leading to sexual dimorphism in cardiometabolic diseases [7].

Epigenetic modifications comprise dynamic changes in the genome engaged in the modification of important cellular processes such as gene expression, chromosomal stability and genomic imprinting [8, 9]. DNA methylation (DNAm) is the best understood and most extensively studied epigenetic mechanism in regard to CVD risk [8, 9]. DNAm refers to the transfer of a methyl group into the C5 position of the cytosine to form 5-methylcytosine (5mC) and increases or decreases in genomic 5mC are referred as DNA hyper- and hypomethylation respectively [8, 9]. The global DNAm assessed at long-interspersed nuclear element (LINE-1) has been inversely associated with intermediate CVD risk factors and higher risk of metabolic status worsening [10]. Conversely, a higher degree of global DNA methylation measured at Alu repeats or by the LUMA method was associated with the presence of CVD [10]. Also, gene specific hyper- or DNA hypomethilation was associated with changes in gene expression and was shown to affect cardiometabolic risk including atherosclerosis, inflammation, blood pressure, serum lipid and glucose levels, subsequently leading to increased risk of developing T2D, stroke and myocardial infarction [11]. Also, sex-specific differences in DNAm have been found in brain, human pancreatic islets and blood [12, 13]. Men in general seem to have lower levels of methylation in their genome as compared to women [14, 15], indicating that similarly to sex chromosomes, methylation at the autosomes might be subject to sex differences. Despite this, only a relatively small number of studies in the ample field of cardiometabolic epigenetics stratify according to sex or focus in sex differences. Although a few studies [11, 16, 17] have summarized the existing literature on this complex topic, they did not focus on epigenetically induced sex differences in CVD, intermediate CVD

risk factors or T2D, with the exception of a recent review that focused only on major CVD events[16].

Therefore, we aimed to systematically review the available evidence in human studies exploring the association between sex-specific DNAm and cardiometabolic diseases.

METHODS

Data Sources and Searches

This review was conducted using a predefined protocol (which was not registered at online platforms) and following a recently published guideline on how to perform systematic reviews[18] and in accordance with PRISMA guidelines[19]. A literature search was done using 5 electronic databases (Medline ALL via Ovid (1946-current), EMBASE (1974-current) via embase.com, Web of Science Core Collection (1900-current), Cochrane CENTRAL registry of trials (issue 7 2019) via Wiley and Google Scholar) until 15 July 2019 (date last searched) with the help of an experienced medical information specialist (WMB). All references were imported in an EndNote library and deduplicated with the algorithm developed by Bramer et al[20]. Additionally, we searched the reference lists of the included studies and relevant reviews. Two independent reviewers screened the titles and abstracts of all studies identified initially, according to the selection criteria. Full texts were retrieved from studies that satisfied all selection criteria. All disagreements were resolved through consensus or consultation with two other independent reviewers.

Study Selection

Observational (cross-sectional, case-control, prospective) studies conducted in adults and investigating the associations of global or gene-specific DNAm with cardiometabolic outcomes were selected. Studies were included if they investigated sex-stratified associations between DNAm and intermediate cardio-metabolic traits (blood lipids, glucose, blood pressure, inflammatory markers, atherosclerosis, T2D) and CVD (MI, CHD, stroke). Also, we included studies that reported a significant interaction term with sex but did not stratify by sex in their analysis. Furthermore, studies conducted only in men/women were not included in the current review.

Data Extraction

A predesigned electronic data abstraction form was used to extract relevant information. This included questions on study location, percentage of men and women

included in the study, participants' age, cardio-metabolic outcome, tissue type, DNAm technique used and general and sex-specific findings. Two authors independently extracted the data and a consensus was reached in case of any inconsistency with involvement of an additional author.

Assessing the Risk of Bias

Two independent investigators used the Newcastle-Ottawa Scale[21] to assess the risk of bias of the included observational studies. The Newcastle-Ottawa Scale uses a star system (maximum of nine stars) to evaluate three domains: selection of participants; comparability of study groups; and the ascertainment of outcomes and exposures of interest. Studies that received a score of nine stars were judged to be of low risk of bias; a score of seven or eight stars was medium risk and those that scored six or less were considered at high risk of bias.

Pathway connectivity analysis

To identify biological pathways of the differentially methylated genes previously linked to CVD, we used the CPDB (ConsensusPathDB-human) tool from the Max Plank Institute for Molecular Genetics[22]. This tool integrates interaction networks in humans (in this study) and includes information on binary and complex protein-protein, genetic, metabolic, signaling, gene regulatory and drug-target interactions, as well as biochemical pathways[22]. Data that explains interactions was derived from 32 public resources.

RESULTS

Search Results and Study Characteristics

The search strategy identified 17,023 potentially relevant studies; after titles and abstracts were screened 16, 814 references were excluded (Figure 1). For the remaining 209 references, full-text articles were reviewed, 174 of which were excluded for various reasons outlined in Figure 1. A total of 35 articles based on 30 unique studies were included in this systematic review including a total of 14,020 non-overlapping participants, of whom approximately 53% were women. The studies included population with European (n=13), North American (n=3) and Asian (n=14) ancestries with age ranging from 32 to 75 years. Due to differences in the epigenetics marks and outcomes investigated, as well as different study designs of the included studies, we were not able to quantitatively pool the estimates from various studies. Therefore, we report in this review a detailed descriptive summary of the available published literature. The characteristics of the included studies are described in Table 1. The

median Newcastle-Ottawa quality score for the included studies was 7 of 9 possible points. The Table 1 depicts the methodologic quality of all included studies.

The Role of Sex-Specific DNAm in Intermediate Cardio-metabolic Traits

Blood Lipids

Global DNA methylation is a frequent used marker for epigenetic screening since it captures the DNA methylation also at unknown genetic locations while the results of average DNA methylation correlate with the methylation of some trait-relevant genes[23, 24]. Ten articles[14, 25-33] investigated sex-specific associations between DNAm and blood lipid concentrations applying global (n=2), epi-genome wide (n=1) and candidate gene (n=7) approaches. In total 2,443 non-overlapping participants (1,174 women and 1,269 men) from USA, Canada, Finland, UK and China were included in these studies.

Two cross-sectional studies[14, 25] investigating global DNAm and blood lipid levels reported sex differences. In the study conducted by Cash et al, LINE-1 methylation was significantly higher in men than in women, and among the entire sample, lower levels of *LINE-1* methylation was associated with higher levels of fasting low-density lipoprotein-cholesterol (LDL) and lower levels of fasting high-density lipoprotein-cholesterol (HDL)[14]. However, when stratifying by sex, the inverse association between global LINE-1 methylation, LDL and HDL remained significant only in men[14]. Malipatil et al, in their study reported that an increase of 10% in LINE-1 methylation was associated with decreased cholesterol/HDL ratio by 0.4 mmol/L in the overall sample of men and women. However, when stratifying by sex, the inverse association remained significant only in women[25]. In an EWAS performed by Garcia-Calzon et al, female samples displayed on average higher methylation in the X-chromosome, whereas males presented higher methylation in the autosomes. Further, women showed higher HDL levels, which were associated with higher *KDM6A* expression and epigenetic differences in human liver[31]. The results were not replicated. Further, the authors integrated DNA regulatory regions and other epigenetic factors for CpGs in the autosomal and X-chromosome based on sex ($q < 0.05$) for only four liver donors. Particularly, 42% of the autosomal CpG sites (13,817 CpGs) and 27% of the X-chromosome sites (2601 CpGs) differentially methylated by sex overlapped with histone marks related to active chromatin and enhancer regions (H3K4me1), whereas 14% of the autosomal sites (4760 CpGs) and 11% of the X-chromosome sites overlapped with histone marks related to heterochromatin (H3K27me3)[31].

In three candidate gene-studies *PTPN1*[32], *PLA2G7*[26] and *BCL11A*[27] DNAm was positively associated with serum lipids in women, but not in men. Another study reported that methylation at *ABCG1* was negatively associated with triglycerides in women only[28], whereas methylation at *LIPC* was negatively associated with triglycerides only in men[27]. This latter study also reported sex-specific associations for total cholesterol: whereas *ABCG1* was associated with triglycerides only in women, methylation at this CpG site was inversely associated with total cholesterol only in men[28]. For some CpG sites there were specific associations with HDL for males; methylation at *PLTP*[28], *CETP*[29], and *LIPC*[28] were negatively associated with HDL in men, but not in women. Moreover, a male-specific association was found between *GCK* CpG4 methylation at *GCK* and total cholesterol concentration[33]. However, a single study performed among 739 African Americans in the Genetic Epidemiology Network of Arteriopathy (GENOA) did not find overall or sex-specific significant associations between DNA methylation and lipid levels[30].

Blood pressure

Seven articles[25, 34-39] based on five unique studies reported associations between sex-specific DNAm and essential hypertension (EH), while one study investigated the cross-sectional association between DNAm and blood pressure[25]. Among these, five studies investigated candidate gene methylation[34-36, 38], while another study investigated epi-genome wide methylation[40] in regard to hypertension. Studies included 3,561 non-overlapping participants (2,029 women and 1,373 men) from China and UK.

In the study conducted by Malipatil et al, in the overall sample of men and women, a 10% increase in LINE-1 methylation was associated with a 2.5 mmHg lower baseline diastolic blood pressure. The stratified analysis by sex did not show any significant influence of sex on this association[25]. One candidate gene study reported higher methylation levels of the two CpG sites at *SCNN18* in women compared with men as controls (CpG1: t=-2.283, P=0.025; CpG2: t=-2.568, P=0.012) and incident EH cases (CpG1: t=-2.694, P=0.009; CpG2: t=-2.583, P=0.011)[35]. However, for these two CpG sites no significant difference was observed between males and females in the prevalent cases group (CpG1: t=0.409, P=0.068; CpG2: t=0.621, P=0.536)[35]. These results indicated a significant association between EH and *SCNN1B* methylation, which was affected by age, sex and antihypertensive therapy. Similarly, in one other study, higher *ADD1* DNAm levels were observed in females as compared to males (CpG1: P = 0.016; CpG2-5: P = 0.021)[34]. Further, the study showed that lower *ADD1* CpG1 methylation levels were significantly associated with EH in females (cases versus controls (%, SD): 10.00±1.41 versus 11.36±3.63, adjusted P =0.042) but not in males

(adjusted P = 0.133). In contrast, lower levels of *ADD1* CpG2-5 methylation were associated with an increased risk of EH in males (cases versus controls: 22.48% versus 31.86%, adjusted P = 0.008) but not in females[34]. The prediction potential of EH for *ADD1* CpG1 and CpG2-5 methylation levels was assessed by the ROC curves. CpG2-5 methylation was reported as a significant predictor of EH in males (area under curve (AUC) = 0.855, P = 0.001), while CpG1 methylation showed a trend toward being an EH indicator in females (AUC= 0.699, P =0.054)[34]. In the same population, *AGTR1* CpG1 methylation was a significant predictor of EH in both sexes[36] and hypomethylation of CpG3 site at *IL-6* promotor was significantly associated with EH risk in both, men and women. Further, sex-specific DNAm levels were observed only at CpG1 and CpG2 sites of *IL-6* promoter (males were hypomethylated as compared to females)[38]. Another study by Han et al, investigated the interactions between alcohol consumption and DNA methylation of the *ADD1* gene promoter and its association with EH, involving 2040 cases and controls[36]. The researchers concluded that CpG1 methylation was associated with EH in females while CpG2-5 methylation was significantly associated with EH in males, suggesting that these interactions in the *ADD1* gene promoter might play a role in modifying EH susceptibility[36]. Finally, Bao et al. reported that hypomethylation of the *IFNG* promotor region was associated with a higher risk of EH. However, the authors did not observe any sex differences overall, except that in the control group DNAm levels were found to be higher in males when compared to females[39].

Inflammation and Atherosclerosis

Three articles[41-43], investigated the associations between epi-genome wide DNAm (n=2), candidate gene methylation and inflammatory markers. Also, we did not identify any study investigating the sex-specific role of DNAm in atherosclerosis. Studies in inflammatory markers included 2,771 non-overlapping participants (713 women and 317 men and one study did not report the number of men and women separately[41]) from Germany, China and USA. None of the epi-genome wide studies reported sex specific associations between global DNAm and inflammatory markers[41, 42]. Nevertheless, Guo et al, reported men-specific association between lower *PTX3* promoter methylation levels and higher neutrophil to lymphocyte ratio. Also, the level of *PTX3* promoter methylation in the coronary artery disease group (mean, SD: 62.69% ± 20.57%) was significantly lower than that of the group free of coronary artery disease (mean, SD: 72.45% ±11.84%), suggesting a role of this gene in developing coronary artery disease[43].

Glycemic Traits and T2D

Eight articles[12, 27, 32, 33, 44-47] reported sex-specific associations between DNAm and glycemic traits and T2D. Five studies were candidate gene studies and three were epi-genome wide studies. Among them, six studies focused on DNAm and T2D, one investigated the association between DNAm and metabolic syndrome and another investigated insulin metabolism. Studies included 2,239 non-overlapping participants (353 women and 554 men, with one study not specifying the number of men and women[46]) from Israel, Spain, Sweden, China and USA.

In a case-control study including 1,169 individuals, individual methylation levels at the *FTO* gene showed that CpG sites in the first intron were slightly (3.35%) hypomethylated in T2D cases relative to controls[46]. The odds of developing T2D increased by 6.1% for every 1% decrease in DNAm. Men were hypomethylated relatively to women and the effect of DNAm was stronger in males compared to females (P =0.034 for sex interaction, AUC = 0.675 among men and 0.609 among women)[46]. Also, in another case-control study association between *PTPN1* promoter methylation and the risk of T2D was observed in the overall population and in females[32] .

In the study by Burghardt et al., a significant increase in *CDH22* gene methylation in subjects with metabolic syndrome was identified in the overall sample[44]. However, when investigating males and females separately; differential methylation levels were observed within the *MAP3K13* gene in females and the *CCDC8* gene in males with metabolic syndrome. In the validation sample a significant difference in methylation was again observed for the *CDH22* and *MAP3K13* genes, but not for *CCDC8* gene[44]. Another study investigated the impact of sex on the genome-wide DNAm pattern in human pancreatic islets from 53 males and 34 females, and relate the methylome to changes in expression and insulin secretion[12]. The study identified both chromosome-wide and site-specific sex differences in DNA methylation at the X chromosome of human pancreatic islets. However, the autosomal chromosomes showed differences in DNA methylation only on the level of individual CpG sites between sexes. Importantly, they found higher insulin secretion in pancreatic islets from females compared with males, as well as sex differences in gene expression[12]. Additionally, the authors did not find any difference in β-cell number between females and males. This could suggest that the DNA methylation differences seen between males and females might not be due to altered β-cell composition in the islets[12].

In a case-control study conducted by Rodriguez-Rodero et al. [45], hypermethylation at CpG sites annotated to the HOOK2 gene was associated with the presence of

T2D. Interestingly, when these results were analyzed by sex, female T2D samples were found hypermethylated at the cg04657146-region and the cg11738485-region of the *HOOK2* gene, whilst male samples were found hypomethylated in this latter region only[45]. Tang et al. reported a significant association only in males when investigating the overall *BCL11A* methylation in T2D patients[27]. While in another study among the same population, significantly elevated methylation levels of *GCK* CpG4 were observed in T2D patients than in the healthy controls. Also, this association was characteristic to males only[33]. Further, serum IGFBP-7 protein levels were similar among newly diagnosed and treated T2D patients and were not correlated with *IGFBP7* DNAm overall, but solely in males[47].

The Role of Sex-Specific DNAm in CVD

Coronary Heart Disease

Eight articles[48-55] investigated the associations between DNAm and CHD and MI. All studies applied a candidate-gene approach and included a total of 2,353 participants (1,010 women and 1,343 men) from China, Italy and the Netherlands.

One study reported that a higher DNAm at the imprinted loci of *INS* and *GNASAS* was associated with the incidence of MI in women (INS: +2.5%, P=0.002; GNASAS: +4.2%, P=0.001)[48]. Hypermethylation at one locus and at both loci was associated with odds ratios (ORs) of 2.8 and 8.6, respectively (Ptrend=3.0×10−4) while no association was observed among men. Similarly, one study revealed a female-specific significant association between methylation at *PLA2G7* promoter and risk of CHD[49]. Another study reported a female specific association of *CDKN2B* methylation with CHD [women with CHD (mean, SD: 7.21±2.40%) compared with women without CHD. In contrast, four studies reported men-specific associations between DNAm of various genes and CHD[43, 50, 51]. Peng et al, reported significant associations of the methylated promoter of the *ABCG1* and *GALNT2* genes with an increased risk of CHD overall and among men only[50]. Also, CHD patients had significantly lower levels of *APOE* methylation than non-CHD controls. In addition, rs7412-T and rs7259620-A were protective factors for CHD in males (rs7412-T: OR=0.527, allele P=0.004; rs7259620-A: OR=0.743, allele P=0.029)[54]. Giannakopoulou et al. reported a sex specific increased methylation in the *CTH* promoter gene in 34 patients who had coronary artery bypass surgery (CABG) (19.1%) compared to 16 control subjects (10.3%). Increased methylation levels were observed in male CABG patients compared to male control subjects while in females this was not observed[51]. Furthermore, Xu at al., showed that CHD cases had a significantly lower methylation level of the GCK gene (mean, SD: 49.77 ± 6.43%) compared with controls, while there was no

difference of *GCK* methylation level between males and females and no significant interaction between gender and disease[52]. However, a significant difference of the CpG2 methylation level with CHD was observed in males only[52]. On the other hand, one study evaluated the association between the DNA methylation profiles of genes involved in One-Carbon Metabolism (OCM) and the homocysteine (Hcy) pathway, with the myocardial infraction risk due to the low B-vitamins intake[53], based in the rationale that B-vitamins and folates pathway may modulate DNA methylation[53]. The results from this study showed an inverse association between B-vitamins intake and the hypermethylation in three genes (*TCN2* promoter, *CBS* 50UTR, *AMT* gene-body) in male cases, as well as two genes (*PON1* gene-body, *CBS* 50UTR) in female cases[53].

Stroke

Four articles[15, 56-58] reported sex-specific associations between DNAm and stroke. Among these, three articles[56-58] used data from the same population for their analyses. Two studies applied global and two others candidate gene approaches. These studies applied a cross-sectional design, used blood samples and included a total of 1,045 non-overlapping participants (459 women and 586 men) from diverse ethnic backgrounds, such as Chinese-Taiwanese[56-58] and Spanish[15].

Two of the studies used a candidate gene approach and performed pyrosequencing to assess methylation of the targeted regions in the gene promoter[57, 58], while the two other studies investigated the global DNAm of the LINE-1 elements using pyrosequencing[56] or luminometric methylation assay (LUMA)[15]. Further, two studies found a significant decrease in LINE-1 methylation in men compared to women in cases of stroke[15, 56]. One of them also reported that cases of ischemic stroke presented a lower methylation level compared to controls. In addition, this hypomethylation of LINE-1 in men was associated with an increased stroke risk by 1.2-fold after adjusting for risk factors, while no significant association was observed in women[56]. On the other hand, among the two studies investigating the estrogen receptor alpha *ERα*[57] and the matrix metalloproteinase-2 (*MMP-2*)[58] respectively, the methylation levels were lower in individuals with stroke compared to controls, in all the CpG sites analyzed in both studies. Further, when exploring the sex-specific associations, the two studies obtained contrary results. One study found a significant difference in methylation levels in all the investigated CpGs annotated to the gene *ERα* only between female cases and controls[57]. Whereas the other study reported a significant difference between the methylation levels in one out of eight CpGs annotated to *MMP-2* only between male cases and controls[58]. None of these studies investigated the difference between male and female cases.

Genes, pathways and cardiovascular disease

Studies included in this systematic review report that methylated CpG sites annotated to *KDM6A, PLA2G7, CETP, ABCG1, LIPC, BCL11A, ADD1, CNN1B, HOOK2, PLTP, GALNT2, PON1, TCN2, CBS, AMT, CTH, INS, GNAS-AS1, MMP2, CCDC8, MAP3K13, FTO, ESR1, CDKN2B, APOE, IGFBP7, PTPN1, GCK* and *PTX3* were differently methylated for men and women. An overview of these genes, function and their sex specific methylations is provided in **Table 1**. In addition, **Figure 2** shows the prioritized pathway connectivity between cardio-metabolic genes that were found to be differentially methylated in men and women. The most significant pathways, that in **Figure 2** are shown with darker red nodes with more representative enrichment include the Vitamin B12 Metabolism, Statin Pathway, Plasma lipoprotein, Plasma lipoprotein assembly, remodeling and clearance and Cholesterol metabolism. Hence, the most important genes connecting these pathways that merit further consideration were: *ABCG1, APOE, PLTP, LIPC, CETP, CTH* and *INS*. Overall, the majority of the genes reported in this review were previously known to be associated with CVD risk factors or CVD outcomes.

DISCUSSION

In this review, we systematically summarized the current evidence on sex differences in DNA methylation in relation to cardiometabolic diseases. We included 30 unique studies that had either stratified their analyses by sex and/or specified that their results did not differ among sexes by testing for statistical sex-interaction. Overall, our findings indicate that global DNAm might influence cardiometabolic risk in a sex-specific manner, and that DNAm at 31 gene sites could be differentially associated with various cardiometabolic traits in men and women.

Global DNA methylation

We identified four studies suggesting that altered LINE-1 DNAm may play a role in CVD risk in a sex-specific manner: (i) DNAm measured in LINE-1 repeats was inversely associated with different serum lipids in men and women[14, 25], (ii) decreased LINE-1 was associated with higher stroke risk in the overall sample and in men[56] and (iii) DNAm measured using genomic 5-methyl cytosine content and LUMA indicated hypomethylation in male as compared to female stroke cases[15].

In the current review, global DNA hypomethylation was associated with poorer outcomes. In particular, global DNA hypomethylation was associated with higher LDL and lower HDL levels in the overall sample and in Samoan men, but not in

women[14] and increased stroke risk in Chinese men but not in women[53]. These findings are in line with observations in healthy Caucasian men where subjects with decreased LINE-1 methylation were more likely to develop ischemic heart disease and stroke (women were not included in this study)[59]. Global DNA hypomethylation has been previously reported in stroke patients as compared with healthy[59] and in a large study with participants from European ancestry, decreased global DNAm was observed in male as compared to female stroke cases[56]. In experimental studies, global DNA hypomethylation has been shown to precede the formation of atherosclerosis in Apoe-/- mice, and has been associated with hyperhomocysteinemia and aortic lipid deposition in mutant mice deficient in methylenetetrahydrofolate reductase[60]. While in humans global loss of DNAm has been previously associated with atherosclerosis in both, atherosclerotic lesions[61] and peripheral blood leukocytes[62] but also with blood lipids, inflammation and blood pressure[10] implying that LINE-1 hypomethylation could be associated with an unfavorable cardiovascular risk profile. Global DNAm is considered a robust measurement of the overall genomic methylation which is reported to be one of the earliest molecular changes in the transition of a cell from a normal to a diseased state[63]. Blood DNA hypomethylation might be easily measured and could be used to identify people at risk of cardiovascular events. Our findings emphasize the need of sex-specific approaches when further exploring the possible role of global DNAm as a biomarker and potential intervention target in cardiometabolic diseases.

Epigenome wide-association studies and candidate gene approach

We identified only six studies[12, 31, 41, 42, 44, 46] that investigated differentially methylated regions in the genome with cardiometabolic diseases in a hypothesis-free approach. Among these, three EWAS[12, 31, 44] reported sex-specific associations, and *KDMA6A, FTO, MAP3K13* and *CCDC8* were some of the important genes that were found to be methylated in a sex-specific manner with blood lipids and glycaemia traits. Among 25 candidate gene studies, 22 studies reported sex-specific associations between DNAm and cardiometabolic diseases at the following genes *PLA2G7, BCL11A, KDM6A, LIPC, ABCG1, PLTP, CETP, ADD1, CNN1B, HOOK2, GFBP-7,PTPN1, GCK, PTX3, ABCG1, GALNT2, CDKN2B, APOE, CTH, GNASAS, INS, PON1, TCN2, CBS, AMT, MMP-2* and *ER-α*. Based on the prioritized pathway connectivity analysis, although, limited, the evidence suggests an involvement of biological pathways related to vitamin B12 metabolism, statin pathway, plasma lipoprotein, plasma lipoprotein assembly, remodeling and clearance and cholesterol metabolism, with sex differences in cardiometabolic diseases (Figure 2). Some of the most relevant genes from the pathway analysis were *ABCG1, APOE, PLTP, LIPC, CETP, CTH* and *INS*. Overall, these genes have

been associated to cardiometabolic outcomes, however little evidence links them to epigenetics and cardiometabolic diseases and even less to sex differences in cardiometabolic diseases.

ABCG1 gene is a transmembrane cholesterol transporter that effluxes cellular cholesterol from macrophages by delivering cholesterol to mature high-density lipoprotein particles. Beyond a role in cellular lipid homeostasis, ABCG1 equally participates to glucose and lipid metabolism by controlling the secretion and activity of insulin and lipoprotein lipase. Moreover, there is a growing body of evidence suggesting that modulation of *ABCG1* expression might contribute to the development of diabetes and obesity[64], which are major risk factors of CVD. The *ABCG1, GALNT2* and *HMGCR* genes have been previously associated with pathogenesis and progression of CHD through manipulating the various lipid pathways[65, 66]. In the current review, hypermethylation of these three genes was associated with higher levels of total cholesterol and LDL , and increased CHD risk in men[50], while it was linked to higher levels of TG in women but not with risk of CHD[7]. The expression of *ABCG1* gene reduces cholesterol accumulation in macrophages by promoting the transfer of intracellular cholesterol into HDL pathway[67]. In contrast to this, hypermethylation at *PLA2G7* was associated with levels of total cholesterol, triglycerides and ApoB in females but not in males, and also only female CHD cases were hypermethylated as compared to controls[5]. *PLA2G7* is the coding gene for Lp-PLA2 whose abnormal activity can cause high risk of CHD and may serve as a diagnostic marker for CHD[68]. Therefore, the sex disparities in the *ABCG1* and *PLA2G7* methylation may have an effect in the molecular mechanisms underlying the sex-specific pathophysiology of CHD and may provide epigenetic clues to explain the inconsistency in the epidemiological studies. However, both studies were done in Han Chinese population, and sample size was rather small (only 85 CHD patients and 54 participants without CHD[50] and 36 CHD cases and 36 controls). Hence, further replication studies with larger sample size and in different ethnic populations are required to confirm these findings.

The *APOE* gene encodes apolipoprotein ε (ApoE), a protein that associates with lipids to form lipoproteins that package and traffic cholesterol and lipids through the bloodstream and has been linked with numerous physiological conditions, including healthy aging[69], cardiovascular disease[70], diabetes[71] and cognitive function[72]. One study using samples of 563 blood-bank donors, found that 1% of the inter-individual variation in plasma ApoE levels was attributable to variation of age and sex[73]. Researchers from the ApoEurope project, reported a sex-differential

effect of age on mean levels of ApoE[74]. In men, the levels of ApoE levelled off after the age of 45 years, whereas they continued to increase in women[74].

PLTP (phospholipid transfer protein) is essential for the transfer of excess surface lipids from TG-rich lipoproteins to HDL particles. PLTP-mediated phospholipid transport among HDL particles is also known to be associated with HDL particle size and lipid composition. Sex disparities for HDL levels associated with *PLTP* have been previously reported[75]. In the PAGE study, the major allele of rs7679 was associated with higher HDL levels in women only. The locus with the most consistent evidence for sex disparities across the studies is *LPL*, or lipoprotein lipase. Different SNPs in this gene exhibited sex disparities for HDL levels in two prior studies, with a larger effect in males[76, 77]. In two other studies, *LPL* exhibited sex differences for TG levels, also with a stronger effect in males[75, 78]. LPL is the rate-limiting enzyme for hydrolysis of triglycerides in lipoproteins and polymorphisms and mutations in LPL have been associated with lipid metabolism disorders. Hormone levels have been shown to affect regulation of LPL, including thyroid hormone, estrogen, and testosterone[79], which could possibly and partially explain the observed association in cardiometabolic diseases.

Although these pathways and the respective reported genes need further investigation, confirmation and translational research, the current evidence suggests they could be biologically relevant and could hold the key for future drug discovery, diagnosis and treatment of cardiometabolic diseases overall and in a sex specific manner. Determining the relationships between genes is essential for molecular biology and medicine. These relationships often cluster together into functional and disease pathways, and the characterization of these pathways is necessary to improve disease classification, patient stratification and, ideally, personalized treatment[80].

Epigenetic mechanisms in biological processes of sex differences

Sex differences in pathophysiology of cardiometabolic diseases could be attributed to several gender and sex-specific factors[81]. Lifestyle factors (smoking, diet, stress) can determine gender differences by modifying cardiometabolic risk directly, and they can also modify the epigenetic marks in a sex-specific manner leading to sex differences in cardiometabolic diseases[81](**Figure 2**). Sex differences may also be driven by biological dissimilarities rather than different environmental exposures among men and women[81]. In particular, the major mechanisms by which sex might influence cardiometabolic diseases epigenetics may include: (i) the genomic and non-genomic effects of steroid hormones and their receptors on DNAm en-

zymes, histone modifiers and miRNAs, (ii) genomic imprinting, leading to DNAm of either maternal or paternal alleles and (iii) increased expression of X-chromosomal escape genes in women targeting epigenetic modifications and the expression of non-pseudo-autosomal Y-chromosomal epigenetic modifiers in men[16] .

Sex hormones have been extensively studied in the past decade in regard to cardiometabolic diseases due to the better cardiometabolic profiles in women as compared to their male counterparts. In the current review, we found some implications for the interactions between sex hormones and methylation in modifying sex differences cardiometabolic diseases. Three[15, 57, 58] of the studies included in the current review, investigating epigenetic changes and stroke and reporting differences between men and women, highlight the importance of sex hormones and their receptors. Using a global DNA methylation approach, Soriano-Tarraga et al[15], found that global hypomethylation was independently associated with stroke subtypes only in females. Moreover, there was an association between lower *ERα* methylation levels and large-artery and cardio-embolic stroke subtypes in women, while in men this association was not observed. It might be that women suffering from a major ischemic stroke may cause a more significant change in *ERα* methylation levels to reduce the brain injury[57]. In line with this, differential DNAm profiles mediated sex differences in the endogenous neuroprotective response to middle cerebral artery occlusion (MCAO) in mice were reported. In female mice, MACO induced selective demethylation of the ERα gene promoter, leading to the increase in ERα expression[82]. Also, sex differences were observed in *MMP-2* methylation, with expression of *MMP-2* being closely related to sex hormones[58]. Males with small-vessel ischemic stroke had lower methylation levels at all *MMP-2* CpG sites, while no association was observed in women[58]. Further exploration of the underlying mechanisms is needed. Sex- and stroke-subtype-specific effects must be taken into consideration when investigating potential strategies to alter the activity of *MMP-2* in patients with ischemic stroke. Steroid hormones can induce, among others, modification of histones. Androgen or estrogen receptors act by binding to hormone response elements in the DNA and attract various cofactors that have inherent histone acetyltransferase or methyl transferase activity. This is particularly known for the *CREB* binding protein and E1A binding protein p300[83]. The histone-modifying enzymes alter the epigenetic state of gene promoters to which the nuclear receptors bind, thereby changing gene expression.

It is known that DNAm contributes to X-chromosome inactivation in females[84], and findings by Garcia-Calzon et al, demonstrated that DNA methylation in the X-chromosome in human liver mirrors the methylome in other human tissues[31].

They reported higher average degree of X-chromosome methylation in females than in males with 37% of the significant sites on the X-chromosome having higher methylation in males[31]. Around 95% of the CpG sites in the X-chromosome that had differential DNAm in human liver between sexes also had different methylation levels between males and females in pancreatic islets and brain independently of the clinical characteristics of the population[12, 85]. Further, they identified four genes on the X-chromosome with large differences in DNA methylation between males and females and being more expressed in liver from females than males: *XIST*, *ARSE*, *RPS4X*, and *KDM6A*[31]. Also, higher *ARSE* and *RPS4X* mRNA expression has been found in pancreatic islets, and higher *XIST* and *RPS4X* mRNA expression was also found in brain from females compared with males[15,][85]. These differences in gene expression in several tissues may explain some metabolic differences between males and females. Interestingly, these four genes are known to escape X-chromosome inactivation[86-88]. In this study serum HDL levels were positively associated with *KDM6A* mRNA expression in human liver in addition to higher serum HDL levels and higher *KDM6A* expression in females. Also, silencing *KDM6A* in hepatocytes resulted in lower HDL levels and lower expression of key genes encoding proteins that regulate HDL levels, supporting the direct contribution of *KDM6A* in the differences found in HDL levels between males and females[31].

Potential clinical implications and recommendations for future research

Although the clinical use of epigenetic marks in the field of cardiometabolic diseases is still in its infancy this is not the case with cancer research. Molecular risk stratification using (epi)genetic marks have been focused on identifying molecular features associated with clinical outcome and have applied them to patients' risk stratification and treatment guidance[89, 90]. Such results indicated that a gene expression score that incorporates prognostic genetic and epigenetic information could be used as a model for treatment response but also for risk stratification and early disease detection. In particular, sex-specific epigenetic marks against or as a supplement to existing risk scores (such as the Framingham Risk Score[91]) may be an added value when predicting the risk of cardiometabolic diseases. This is also supported by our findings suggesting that for cardiometabolic traits epigenetic markers may not be equally good predictors in men and women, emphasizing the role of sex in epigenetic patterns of cardiometabolic diseases. Further, as sex is one of the strongest predictors of treatment response, the epigenetic signatures may be used as markers to indicate the successfulness of pharmacological or dietary/lifestyle interventions in cardiometabolic diseases among sexes. Given the lack of sex-stratification in studies focusing on epigenetic mechanisms and the fact

that the majority of the studies were focused to epigenetic changes in autosome chromosomes in regard of cardiometabolic diseases, our review underscores the emerging need for future studies to investigate the influence of sex on epigenetic mechanisms in cardiometabolic diseases. In complex phenotypes such as cardiometabolic diseases, the collection of high-quality blood samples and metabolically active tissues could provide the basis for the creation of large data sets that should accurately incorporate the many sources of variability (age, sex/gender, race/ethnicity). In particular, future prospective observational studies should aim to explore the role of sex when studying the associations between epigenetic marks and mechanistic pathways of cardiometabolic diseases by stratifying their analyses by sex and comparing male and female participants. Second, studying the associations between DNAm and intermediate CVD risk factors is valuable, however, from the clinical perspective, the value of DNAm as a biomarker of the risk factor is as good as the intermediate risk factor itself. Therefore, as we did not identify studies focusing on outcomes such as stroke or myocardial infarction, it might be of great value for future research to investigate the role of sex on the epigenetic determinants of stroke and myocardial infarction. Third, potential biological mechanisms underlying sex-specific associations should be further explored in an experimental setting. It is now known that sex differences in morphology and in response to stress exist also in cellular levels[92-94]. Therefore, when translating the observational findings into experimental settings a clear distinction between male and female animal models or cell cultures is of high importance in order to obtain non-biased results on the sex-specific pathophysiology of cardiometabolic diseases.

Strengths and limitations

In this systematic review on sexually dimorphic DNAm, we critically appraised the literature following an a priori designed protocol with clearly defined inclusion and exclusion criteria using a comprehensive literature search in five databases. While previous systematic reviews on the topic are limited only to major CVD outcomes[16], our study took in consideration a broad range of cardiometabolic risk factors and diseases. However, the limitations of the findings from this study merit careful consideration. The included studies in this review were limited in sample size and the majority of studies included were cross-sectional assessments, making it difficult to conclude whether DNAm patterns are a cause or consequence of cardiometabolic changes. In addition, the results of some of the studies need cautious interpretation when it comes to the biological or functional relevance of their findings. Even though a study may report a significant difference in DNAm the biological relevance of small differences could be likely minimal and unknown. Studies investigating associations between metabolic syndrome and DNAm also

need to be interpreted with caution given the heterogeneity of metabolic syndrome and that the subjects may or may not have dyslipidemia, elevated BP, and hyperglycemia. Therefore, interpreting associations between changes in DNAm and subjects classified as having metabolic syndrome is Moreover, although individual studies attempted to adjust for established CVD risk factors, adjustment levels were inconsistent across the studies. Also, DNAm patterns reported in blood samples may not mirror the methylation patterns in the relevant targeted tissues. Further, we did not perform the search for non-coding microRNAs and histone modifications because the scope of our search was DNA methylation. Given the importance of microRNAs and histone modifications as epigenetic mechanisms, future systematic reviews and meta- analyses on microRNAs, histone modifications and sex differences in different types of cardiovascular tissues would be an added value on the topic. Moreover, we hand searched relevant reviews and references of studies included in the current review in order to minimize the possibility of missing important studies. Also, we cannot exclude the possibility of publication bias from underreporting negative findings. Lastly, a meaningful quantitative pooling of the existing data was unfeasible due to the heterogeneity in the input parameters, assumptions and study design.

CONCLUSIONS

Although a growing body of evidence suggests biological, genetic and epigenetic sex differences in cardiometabolic diseases, only a small number of studies in the field stratify or present their results by sex. Nevertheless, the cumulative evidence from the studies that reported sex-based results, suggest that epigenetic changes in specific individual genes might be differently associated with cardiometabolic traits in males and females, encouraging further and larger-scale investigation. Robust, replicable results from carefully designed studies have the potential to uncover the molecular biological processes involved in disease onset and progression. In addition, future studies should help characterize gene regulatory effects of non-coding genetic variations, and, hopefully, give indications into disease-relevant biological pathways which could be addressed by preventive or therapeutic interventions. Clearly, a considerable amount of functional work is required in the future to expand our field of view beyond the classic biological mechanisms involved in sex differences of cardiometabolic diseases, and that could be important to design new drugs that target sex-specific mechanisms and permit more precise and efficient care.

REFERENCES

1. Manach, C., et al., *Addressing the inter-individual variation in response to consumption of plant food bioactives: Towards a better understanding of their role in healthy aging and cardiometabolic risk reduction.* Mol Nutr Food Res, 2017. 61(6).
2. Park, Y.W., et al., *The metabolic syndrome: prevalence and associated risk factor findings in the US population from the Third National Health and Nutrition Examination Survey, 1988-1994.* Arch Intern Med, 2003. 163(4): p. 427-36.
3. Bjornerem, A., et al., *Endogenous sex hormones in relation to age, sex, lifestyle factors, and chronic diseases in a general population: the Tromso Study.* J Clin Endocrinol Metab, 2004. 89(12): p. 6039-47.
4. Humphries, K.H., et al., *Sex differences in cardiovascular disease - Impact on care and outcomes.* Front Neuroendocrinol, 2017. 46: p. 46-70.
5. Appelman, Y., et al., *Sex differences in cardiovascular risk factors and disease prevention.* Atherosclerosis, 2015. 241(1): p. 211-8.
6. Mehta, L.S., et al., *Acute Myocardial Infarction in Women: A Scientific Statement From the American Heart Association.* Circulation, 2016. 133(9): p. 916-47.
7. Boyne, D.J., et al., *Endogenous sex hormone exposure and repetitive element DNA methylation in healthy postmenopausal women.* Cancer Causes Control, 2017. 28(12): p. 1369-1379.
8. Feinberg, A.P., *Epigenetics at the epicenter of modern medicine.* JAMA, 2008. 299(11): p. 1345-50.
9. Robertson, K.D., *DNA methylation and human disease.* Nat Rev Genet, 2005. 6(8): p. 597-610.
10. Muka, T., et al., *The role of epigenetic modifications in cardiovascular disease: A systematic review.* Int J Cardiol, 2016. 212: p. 174-83.
11. Braun, K.V., et al., *The role of DNA methylation in dyslipidaemia: A systematic review.* Prog Lipid Res, 2016. 64: p. 178-191.
12. Hall, E., et al., *Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets.* Genome Biol, 2014. 15(12): p. 522.
13. Yousefi, P., et al., *Sex differences in DNA methylation assessed by 450 K BeadChip in newborns.* BMC Genomics, 2015. 16: p. 911.
14. Cash, H.L., et al., *Cardiovascular disease risk factors and DNA methylation at the LINE-1 repeat region in peripheral blood from Samoan Islanders.* Epigenetics, 2011. 6(10): p. 1257-1264.
15. Soriano-Tarraga, C., et al., *Global DNA methylation of ischemic stroke subtypes.* PLoS One, 2014. 9(4): p. e96543.
16. Hartman, R.J.G., S.E. Huisman, and H.M. den Ruijter, *Sex differences in cardiovascular epigenetics-a systematic review.* Biol Sex Differ, 2018. 9(1): p. 19.
17. Stoll, S., C. Wang, and H. Qiu, *DNA methylation and histone modification in hypertension.* Int J Mol Sci, 2018. 19(4).
18. Muka, T., et al., *A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in medical research.* Eur J Epidemiol, 2019.
19. Moher, D., et al., *Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.* Int J Surg, 2010. 8(5): p. 336-41.
20. Bremer, W.M., *Reference checking for systematic reviews using Endnote.* Journal of the Medical Library Association, 2018. 106(4): p. 542-546.
21. Lewis, J.E., et al., *A randomized controlled trial of the effect of dietary soy and flaxseed*

muffins on quality of life and hot flashes during menopause. Menopause, 2006. 13(4): p. 631-642.
22. Kamburov, A., et al., *The ConsensusPathDB interaction database: 2013 update.* Nucleic Acids Res, 2013. 41(Database issue): p. D793-800.
23. Ohka, F., et al., *The Global DNA Methylation Surrogate LINE-1 Methylation Is Correlated with MGMT Promoter Methylation and Is a Better Prognostic Factor for Glioma.* Plos One, 2011. 6(8).
24. Knothe, C., et al., *Disagreement between two common biomarkers of global DNA methylation.* Clinical Epigenetics, 2016. 8.
25. Malipatil, N., et al., *Assessment of global long interspersed nucleotide element-1 (LINE-1) DNA methylation in a longitudinal cohort of type 2 diabetes mellitus (T2DM) individuals.* Int J Clin Pract, 2018: p. e13270.
26. Jiang, D., et al., *Elevated PLA2G7 Gene Promoter Methylation as a Gender-Specific Marker of Aging Increases the Risk of Coronary Heart Disease in Females.* PLoS One, 2013. 8(3).
27. Tang, L., et al., *BCL11A gene DNA methylation contributes to the risk of type 2 diabetes in males.* Exp Ther Med, 2014. 8(2): p. 459-463.
28. Guay, S.P., et al., *Epipolymorphisms within lipoprotein genes contribute independently to plasma lipid levels in familial hypercholesterolemia.* Epigenetics, 2014. 9(5): p. 718-729.
29. Guay, S.P., et al., *DNA methylation variations at CETP and LPL gene promoter loci: New molecular biomarkers associated with blood lipid profile variability.* Atherosclerosis, 2013. 228(2): p. 413-420.
30. Wright, M.L., et al., *Joint Influence of SNPs and DNA Methylation on Lipids in African Americans From Hypertensive Sibships.* Biol Res Nurs, 2018. 20(2): p. 161-167.
31. Garcia-Calzon, S., et al., *Sex Differences in the Methylome and Transcriptome of the Human Liver and Circulating HDL-Cholesterol Levels.* J Clin Endocrinol Metab, 2018. 103(12): p. 4395-4408.
32. Huang, Q., et al., *Elevation of PTPN1 promoter methylation is a significant risk factor of type 2 diabetes in the Chinese population.* Exp Ther Med, 2017. 14(4): p. 2976-2982.
33. Tang, L., et al., *Elevated CpG island methylation of GCK gene predicts the risk of type 2 diabetes in Chinese males.* Gene, 2014. 547(2): p. 329-33.
34. Zhang, L.-N., et al., *Lower ADD1 gene promoter DNA methylation increases the risk of essential hypertension.* PLoS One, 2013. 8(5): p. e63455.
35. Zhong, Q., et al., *Association of SCNN1B promoter methylation with essential hypertension.* Mol Med Rep, 2016. 14(6): p. 5422-5428.
36. Han, L., et al., *The interactions between alcohol consumption and DNA methylation of the ADD1 gene promoter modulate essential hypertension susceptibility in a population-based, case-control study.* Hypertens Res, 2015. 38(4): p. 284-90.
37. Fan, R., et al., *Association of AGTR1 Promoter Methylation Levels with Essential Hypertension Risk: A Matched Case-Control Study.* Cytogenet Genome Res, 2015. 147(2-3): p. 95-102.
38. Mao, S.Q., et al., *Hypomethylation of interleukin-6 (IL-6) gene increases the risk of essential hypertension: a matched case-control study.* J Hum Hypertens, 2017. 31(8): p. 530-536.
39. Bao, X.J., et al., *Hypomethylation of the Interferon gamma Gene as a Potential Risk Factor for Essential Hypertension: A Case-Control Study.* Tohoku J Exp Med, 2018. 244(4): p. 283-290.
40. Bostrom, A.E., et al., *Longitudinal genome-wide methylation study of Roux-en-Y gastric bypass patients reveals novel CpG*

sites associated with essential hypertension. BMC Med Genomics, 2016. 9: p. 20.

41. Marzi, C., et al., *Epigenetic Signatures at AQP3 and SOCS3 Engage in Low-Grade Inflammation across Different Tissues.* PLoS One, 2016. 11(11): p. e0166015.
42. Sun, Y.V., et al., *Gene-specific DNA methylation association with serum levels of C-reactive protein in African Americans.* PLoS One, 2013. 8(8): p. e73480.
43. Guo, T.M., et al., *Pentraxin 3 (PTX3) promoter methylation associated with PTX3 plasma levels and neutrophil to lymphocyte ratio in coronary artery disease.* J Geriatr Cardiol, 2016. 13(8): p. 712-717.
44. Burghardt, K.J., et al., *The Influence of Metabolic Syndrome and Sex on the DNA Methylome in Schizophrenia.* Int J Genomics, 2018. 2018: p. 8076397.
45. Rodriguez-Rodero, S., et al., *Altered intragenic DNA methylation of HOOK2 gene in adipose tissue from individuals with obesity and type 2 diabetes.* PLoS One, 2017. 12(12): p. e0189153.
46. Toperoff, G., et al., *Genome-wide survey reveals predisposing diabetes type 2-related DNA methylation variations in human peripheral blood.* Hum Mol Genet, 2012. 21(2): p. 371-83.
47. Gu, H.F., et al., *Evaluation of IGFBP-7 DNA methylation changes and serum protein variation in Swedish subjects with and without type 2 diabetes.* Clin Epigenetics, 2013. 5(1): p. 20.
48. Talens, R.P., et al., *Hypermethylation at loci sensitive to the prenatal environment is associated with increased incidence of myocardial infarction.* Int J Epidemiol, 2012. 41(1): p. 106-15.
49. Jiang, D., et al., *Elevated PLA2G7 gene promoter methylation as a gender-specific marker of aging increases the risk of coronary heart disease in females.* PLoS One, 2013. 8(3): p. e59752.
50. Peng, P., et al., *A preliminary study of the relationship between promoter methylation of the ABCG1, GALNT2 and HMGCR genes and coronary heart disease.* PLoS One, 2014. 9(8): p. e102265.
51. Giannakopoulou, E., et al., *Epigenetics-by-Sex Interaction for Coronary Artery Disease Risk Conferred by the Cystathionine gamma-Lyase Gene Promoter Methylation.* OMICS, 2017. 21(12): p. 741-748.
52. Xu, L., et al., *GCK gene-body hypomethylation is associated with the risk of coronary heart disease.* Biomed Res Int, 2014. 2014: p. 151723.
53. Fiorito, G., et al., *B-vitamins intake, DNA-methylation of One Carbon Metabolism and homocysteine pathway genes and myocardial infarction risk: the EPICOR study.* Nutr Metab Cardiovasc Dis, 2014. 24(5): p. 483-8.
54. Ji, H., et al., *APOE hypermethylation is significantly associated with coronary heart disease in males.* Gene, 2019. 689: p. 84-89.
55. Chen, X., et al., *Elevated methylation of cyclin dependent kinase inhibitor 2B contributes to the risk of coronary heart disease in women.* Exp Ther Med, 2019. 17(1): p. 205-213.
56. Lin, R.T., et al., *LINE-1 methylation is associated with an increased risk of ischemic stroke in men.* Curr Neurovasc Res, 2014. 11(1): p. 4-9.
57. Lin, H.F., et al., *Demethylation of Circulating Estrogen Receptor Alpha Gene in Cerebral Ischemic Stroke.* PLoS One, 2015. 10(9): p. e0139608.
58. Lin, H.F., et al., *Methylation in the matrix metalloproteinase-2 gene is associated with cerebral ischemic stroke.* J Investig Med, 2017. 65(4): p. 794-799.
59. Baccarelli, A., et al., *Ischemic heart disease and stroke in relation to blood DNA methylation.* Epidemiology, 2010. 21(6): p. 819-28.
60. Chen, Z., et al., *Mice deficient in methylenetetrahydrofolate reductase exhibit hyperhomocysteinemia and decreased meth-*

ylation capacity, with neuropathology and aortic lipid deposition. Hum Mol Genet, 2001. 10(5): p. 433-43.
61. Hiltunen, M.O., et al., *DNA hypomethylation and methyltransferase expression in atherosclerotic lesions.* Vasc Med, 2002. 7(1): p. 5-11.
62. Castro, R., et al., *Increased homocysteine and S-adenosylhomocysteine concentrations and DNA hypomethylation in vascular disease.* Clin Chem, 2003. 49(8): p. 1292-6.
63. Mikeska, T. and J.M. Craig, *DNA methylation biomarkers: cancer and beyond.* Genes (Basel), 2014. 5(3): p. 821-
64. Tavoosi, Z., et al., *Cholesterol Transporters ABCA1 and ABCG1 Gene Expression in Peripheral Blood Mononuclear Cells in Patients with Metabolic Syndrome.* Cholesterol, 2015. 2015: p. 682904.
65. Jeemon, P., et al., *Implications of discoveries from genome-wide association studies in current cardiovascular practice.* World J Cardiol, 2011. 3(7): p. 230-47.
66. Tietjen, I., et al., *Segregation of LIPG, CETP, and GALNT2 mutations in Caucasian families with extremely high HDL cholesterol.* PLoS One, 2012. 7(8): p. e37437.
67. Oram, J.F. and A.M. Vaughan, *ATP-Binding cassette cholesterol transporters and cardiovascular disease.* Circ Res, 2006. 99(10): p. 1031-43.
68. Grallert, H., et al., *Eight genetic loci associated with variation in lipoprotein-associated phospholipase A2 mass and activity and coronary heart disease: meta-analysis of genome-wide association studies from five community-based studies.* Eur Heart J, 2012. 33(2): p. 238-51.
69. Garatachea, N., et al., *ApoE gene and exceptional longevity: Insights from three independent cohorts.* Exp Gerontol, 2014. 53: p. 16-23.
70. Kathiresan, S., et al., *Polymorphisms associated with cholesterol and risk of cardiovascular events.* N Engl J Med, 2008. 358(12): p. 1240-9.
71. Yin, Y.W., et al., *Influence of apolipoprotein E gene polymorphism on development of type 2 diabetes mellitus in Chinese Han population: a meta-analysis of 29 studies.* Metabolism, 2014. 63(4): p. 532-41.
72. Rubino, E., et al., *Apolipoprotein E polymorphisms in frontotemporal lobar degeneration: a meta-analysis.* Alzheimers Dement, 2013. 9(6): p. 706-13.
73. Boerwinkle, E. and G. Utermann, *Simultaneous effects of the apolipoprotein E polymorphism on apolipoprotein E, apolipoprotein B, and cholesterol metabolism.* Am J Hum Genet, 1988. 42(1): p. 104-12.
74. Haddy, N., et al., *The importance of plasma apolipoprotein E concentration in addition to its common polymorphism on inter-individual variation in lipid levels: results from Apo Europe.* Eur J Hum Genet, 2002. 10(12): p. 841-50.
75. Asselbergs, F.W., et al., *Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci.* Am J Hum Genet, 2012. 91(5): p. 823-38.
76. Teslovich, T.M., et al., *Biological, clinical and population relevance of 95 loci for blood lipids.* Nature, 2010. 466(7307): p. 707-13.
77. Aulchenko, Y.S., et al., *Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts.* Nat Genet, 2009. 41(1): p. 47-55.
78. Taylor, K.C., et al., *Investigation of gene-by-sex interactions for lipid traits in diverse populations from the population architecture using genomics and epidemiology study.* BMC Genet, 2013. 14: p. 33.
79. Wang, H. and R.H. Eckel, *Lipoprotein lipase: from gene to obesity.* Am J Physiol Endocrinol Metab, 2009. 297(2): p. E271-88.
80. Barabasi, A.L., N. Gulbahce, and J. Loscalzo, *Network medicine: a network-based approach to human disease.* Nat Rev Genet, 2011. 12(1): p. 56-68.

81. Jaenisch, R. and A. Bird, *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.* Nature Genetics, 2003. 33: p. 245-254.
82. Westberry, J.M., A.K. Prewitt, and M.E. Wilson, *Epigenetic regulation of the estrogen receptor alpha promoter in the cerebral cortex following ischemia in male and female rats.* Neuroscience, 2008. 152(4): p. 982-9.
83. Leader, J.E., et al., *Epigenetic regulation of nuclear steroid receptors.* Biochemical Pharmacology, 2006. 72(11): p. 1589-1596.
84. Carrel, L. and H.F. Willard, *X-inactivation profile reveals extensive variability in X-linked gene expression in females.* Nature, 2005. 434(7031): p. 400-4.
85. Xu, H., et al., *Sex-biased methylome and transcriptome in human prefrontal cortex.* Hum Mol Genet, 2014. 23(5): p. 1260-70.
86. Brown, C.J., et al., *A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome.* Nature, 1991. 349(6304): p. 38-44.
87. Lan, F., et al., *A histone H3 lysine 27 demethylase regulates animal posterior development.* Nature, 2007. 449(7163): p. 689-94.
88. Lee, M.G., et al., *Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination.* Science, 2007. 318(5849): p. 447-50.
89. Figueroa, M.E., et al., *DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia.* Cancer Cell, 2010. 17(1): p. 13-27.
90. Marcucci, G., et al., *Epigenetics meets genetics in acute myeloid leukemia: clinical impact of a novel seven-gene score.* J Clin Oncol, 2014. 32(6): p. 548-56.
91. Wannamethee, S.G., et al., *Metabolic syndrome vs Framingham Risk Score for prediction of coronary heart disease, stroke, and type 2 diabetes mellitus.* Arch Intern Med, 2005. 165(22): p. 2644-50.
92. Addis, R., et al., *Human umbilical endothelial cells (HUVECs) have a sex: characterisation of the phenotype of male and female cells.* Biol Sex Differ, 2014. 5(1): p. 18.
93. Lorenz M, K.J., Kaufmann K, Kreye C, Mertens M, Kuebler WM, Baumann G, Gossing G, Marki A, Zakrzewicz A, Miéville C, Benn A, Horbelt D, Wratil PR, Stangl K, Stangl V, *Does cellular sex matter? Dimorphic transcriptional differences between female and male endothelial cells.* Atherosclerosis, 2015. 240(1): p. 61-72.
94. Du L, H.R., Bayir H, Watkins SC, Tyurin VA, Guo F, Kochanek PM, Jenkins LW, Ren J, Gibson G, Chu CT, Kagan VE, Clark RS, *Starving neurons show sex difference in autophagy.* J Biol Chem, 2009. 284(4): p. 2383-96.

Chapter 3.

Dissecting the association between statin use and risk of type 2 diabetes

Chapter 3.1.

Associations of statin use with glycaemic traits and incident type 2 diabetes

Fariba Ahmadizar, **Carolina Ochoa-Rosales**, Marja Glisic, Oscar Franco, Taulant Muka, Bruno Stricker.

Associations of statin use with glycaemic traits and incident type 2 diabetes. *Br J Clin Pharmacol 2019;85:993-1002*

ABSTRACT

There are several epidemiological studies on the association between statins and incident diabetes, but most of them lack details. In this study, we aimed to investigate the association of statin use with glycaemic traits and incident type 2 diabetes. Using the prospective population-based Rotterdam Study, we included 9,535 individuals free from diabetes at baseline (>45 years) during the study period between 1997 and 2012. Linear regression analysis was applied to examine the cross-sectional associations between statin use and glycaemic traits including fasting blood serum of glucose and insulin concentrations, and insulin resistance. In a longitudinal follow-up study, we applied a Cox regression analysis to determine adjusted hazard ratios (HR) for incident type 2 diabetes in new users of statins. The mean age at baseline was 64.3±10.1 year and 41.7% were men. In the fully adjusted model, compared to never users of statins, baseline use of statins was associated with higher concentrations of serum fasting insulin (β, 0.07; 95% CI: 0.02-0.13) and insulin resistance (β, 0.09; 95% CI: 0.03-0.14). Ever use of statins was associated with a 38% higher risk of incident type 2 diabetes (HR, 1.38; 95% CI: 1.09-1.74). This risk was more prominent in subjects with impaired glucose homeostasis and in overweight/obese individuals. Individuals using statins may be at higher risk for hyperglycaemia, insulin resistance and eventually type 2 diabetes. Rigorous preventive strategies such as glucose control and weight reduction in patients when initiating statin therapy might help minimizing the risk of diabetes.

INTRODUCTION

Although it is well known that statins significantly reduce the risk of cardiovascular disease (CVD) and CVD related mortality [1, 2], statin therapy may lead to increased risk of type 2 diabetes [3, 4]. A recent meta-analysis of 29 randomized clinical trials (RCTs) reported a 12% significantly increased risk of developing type 2 diabetes in the treated arm of statin therapy [5]. However, use under everyday circumstances differs from use in a clinical trial setting with its homogeneous population and short follow-up [6]. Experimental studies suggest that statins may have a diabetogenic effect through beta-cell dysfunction and glucose and insulin secretion/sensitivity [7, 8]; however, the data linking statins with glucose and insulin resistance to show the underlying mechanisms is limited. The literature so far on statin use and the risk of developing type 2 diabetes has been limited by several drawbacks: i) underestimated cases of incident type 2 diabetes due to the inclusion of questionnaire-based data, ii) short follow-up time not adequately taking into account the long-term effects of statins, and iii) lack of a direct comparison between different statin types, dosages and duration of use with respect to diabetes-related outcomes [3, 9, 10].

We hypothesized that statins have influence on serum blood glucose and insulin concentrations and insulin resistance, and lead to hyperglycaemia and hyperinsulinemia as early markers of type 2 diabetes. In this study, exploring different aspects of statin therapy including type, dosages and durations of use, our objective was to assess the population-based association of statins with concentrations of serum fasting glucose and insulin, insulin resistance, and incident type 2 diabetes.

METHODS

Study setting.

The Rotterdam Study is a prospective population-based cohort study in Ommoord, a district of Rotterdam, the Netherlands. The design of the Rotterdam Study has been described in more detail elsewhere [11]. Briefly, in 1989 all residents aged 45 years or older within the well-defined study area were invited to participate, of whom 78% (7,983 out of 10,275) agreed. In 2000, an additional 3,011 participants were enrolled (RS-II), consisting of all persons living in the study district who had become 55 years of age. A second extension of the cohort was initiated in 2006, in which 3,932 participants aged 45 years or older were included (RSIII). The Rotterdam Study has been approved by the medical ethics committee according to the Population Screening Act: Rotterdam Study, executed by the Ministry of Health, Welfare and

Sports of the Netherlands. All participants in the present analysis provided written informed consent to participate and to obtain information from their physicians.

Study Design.

We performed two analyses. First, a cross-sectional analysis to test the association between statins and glycaemic traits and second, a longitudinal follow-up study to test the association of statins with incident type 2 diabetes.

Population for analysis.

The present study used data from the third visit of the first cohort, RSI-3 (March 1997- December 1999) and the baseline examinations of the second cohort, RSII-1 (February 2000-December 2001) and third cohort, RSIII-1 (February 2006-December 2008) (Supplementary **Figure S1**). To study the association between statins and the outcomes of interest, we excluded prevalent type 2 diabetes cases at baseline (n=1,565), and included only participants with available data on statin use and baseline glycaemic values (n=9,535). To study the association of statins with incident type 2 diabetes, patients who used statins at baseline and prevalent cases of cardiovascular diseases (n=691) were also excluded from the analysis. The end of follow-up for incidence of type 2 diabetes was January 1st 2012 (**Supplementary figure 1**).

Exposure measurement.

All community pharmacies in the Rotterdam Study area store all information on drug dispensing on one common computer network. Information on statin treatment was obtained from the dispensing data using the Anatomical Therapeutical Chemical code (ATC code) (**Supplementary table S1**). On the baseline date, an individual was considered as current statin user if this date fell within a prescription episode. During follow-up, each participant was classified into one of the following mutually exclusive categories on the event date of type 2 diabetes (in non-cases at the same day of follow-up): 'current use', 'past use', and 'never use'. If the outcome measurement occurred within a prescription episode of statins, a person was classified as a current user and stratified by statin duration into three mutually exclusive groups: <30 days; 31-365 days; and >365 days. If the participant had previously used statins but was no longer a current user on the event date, this was defined as past use. If the participant had not used statin during the study period, this was defined as 'never use' whereas any use ('current' or 'past use') during the study period was defined as 'ever use'.

OUTCOMES' MEASUREMENT

Fasting serum glucose concentrations, insulin and insulin resistance.

Serum glucose (mmol/l) and insulin (pmol/l) concentrations were measured at the research centre after an overnight fast. For collecting serum specimens from the whole blood, fasting blood samples were centrifuged at 4 °C at 3500 rpm, then serums were immediately frozen at – 140 °C and stored at −80°C in the clinical chemistry laboratory at Erasmus Medical Centre in Rotterdam. Glucose concentration was measured using glucose hexokinase method within 1 week after sampling [12] and insulin concentration by metric assay (Biosource Diagnostics, Camarillo, CA). Homeostasis model assessment-insulin resistance (HOMA-IR) index as a surrogate marker for the degree of insulin resistance was calculated by the following formula: (fasting serum insulin (mU/l) × fasting serum glucose (mmol/l))/22.5 [13].

Type 2 diabetes diagnosis.

From the date of baseline centre visit and during follow-up, first incident type 2 diabetes cases were identified through active follow-up and according to general practitioners' records, hospital discharge letters, and glucose measurements. Type 2 diabetes was defined as a fasting serum glucose concentration of ≥7.0 mmol/l or a non-fasting serum glucose concentration of ≥11.1 mmol/l (in case fasting serum samples were not available), or the use of blood glucose-lowering medications [14].

Covariates assessment.

At baseline and during follow-up, information was obtained on individuals' characteristics, health status, clinical data including medical and medication history, and behavioural/lifestyle factors. Physical activity levels assessed using the Longitudinal Aging Study Amsterdam Physical Activity Questionnaire expressed in MET-h/week [15] and defined as number of hours per week that participants spent in each activity in the past year. Body mass index (BMI) was calculated as body weight (in kilogram) divided by the square of body height (in metres) included as a continuous variable. Baseline and postbaseline values (at the closest time to diabetes diagnosis) for BMIs were categorized as overweight/obesity defined as BMI ≥25 kg/m^2 versus normal (BMI<25 kg/m^2). Hypertension was defined as an average value (of two measurements) of ≥140 mmHg for systolic- and ≥90 mmHg for diastolic blood pressure after two measurements and/or a current prescription for an antihypertensive agent. Serum total cholesterol (TC) (mmol/l), high density lipoproteins-cholesterol (HDL-C) (mmol/l), and triglycerides (TG) (mmol/l) were all measured on the COBAS 8000 Modular Analyzer (Roche Diagnostics GmbH). We calculated low density

lipoproteins-cholesterol (LDL-C) levels (mmol/l) indirectly from the measurements of TC, HDL-C and TG using the Friedewald equation (TC $_{minus}$ HDL-C $_{minus}$ TG/5) [16]. A family history of diabetes was defined as having a parent, sibling, or both with type 2 diabetes; the information on family history of diabetes was collected during home visits at RSI and RSII. All biochemical variables were assessed in serum samples taken after an overnight fasting.

Statistical analyses.

For comparison of baseline characteristics between statin users and never users in the first analysis, we used chi-square statistics for dichotomous variables and independent sample t-tests for normally distributed continuous variables. To achieve a symmetric distribution, non-normally distributed data on BMI, TG, fasting serum insulin concentrations and the HOMA-IR index were natural log transformed. Univariable and multivariable linear regression analyses were used to study the association between statins and glycaemic traits including serum fasting glucose and insulin concentrations and HOMA-IR. Correlation coefficients to identify the magnitude of the linear association between serum fasting glucose and insulin concentrations and the HOMA-IR index were studied with the Pearson test.

To study the association between statin use as a time-varying exposure during the follow-up and incident type 2 diabetes, we used multiple Cox regression analyses [17]. The association was estimated by calculating hazard ratio (HR) and 95% confidence intervals (CI). All analyses show the effect estimates for ever, past or current versus never statin users as the reference group. Among ever statin users, in subgroup analyses, we analysed the risk of different outcomes in relation to statin type, dose, and duration of use.

Because statins are often started in patients with type 2 diabetes, we performed two subgroup analyses to avoid confounding by indication. First, we subtracted one year from the event date in type 2 diabetes cases (in non-cases one year from the same date of follow-up) and studied the association with statin use until that date (cumulative statin use). Second, for those participants for whom we had an incident date of impaired fasting glucose (defined as a serum fasting glucose concentration between 6.1 and 6.9 mmol/l), we studied the association between statin use and incident impaired fasting glucose during follow-up.

All regression analyses were adjusted for potential baseline confounding factors using two models including: model 1 adjusted for age and gender, cohort (RSI, II and

III), smoking status, alcohol consumption, physical activity, and education level, and model 2 additionally adjusted for BMI and hypertension.

Sensitivity analyses for the association between statin and incident type 2 diabetes.

In a sensitivity analysis, we examined the association between statins and risk of incident type 2 diabetes only in a subset of patients (n=6,787) with normal baseline fasting glucose concentrations <6.1 mmol/l by excluding cases of impaired fasting glucose cases (n=2,748). To exclude the likelihood of residual confounding, the association was further adjusted by postbaseline values of BMI. To explore if biological risk factors could partly explain this association, the model 2 was further adjusted for (i) TC, or (ii) HDL-C, LDL-C and TG. In another sensitivity analysis, the model 2 was further adjusted for family history of diabetes, and use of proton pump inhibitors (PPIs) at baseline [18]. To test whether the association between statins and incident type 2 diabetes would be explained by the potential intermediate factors, serum glucose and insulin concentrations were added to the model 2. Using stratified analyses, we also checked the potential effect modification from age at baseline (≤65 years versus >65 years), gender and BMI at baseline (overweight/obese (BMI>25) versus normal BMI individuals). We further tested an interaction between statin ever use and serum fasting glucose concentrations using a Cox regression model containing statin ever use, glucose concentrations and the product term of statin ever multiplied by glucose concentrations adjusted for potential confounders in the model 1.

A *p*-value of 0.05 was used to assess the significance of main effect associations. All statistical analyses were carried out using IBM SPSS Statistics software (version 24.0, Chicago, IL, USA).

RESULTS

Baseline characteristics.

A total of 9,535 diabetes-free individuals at baseline were included after exclusion of patients with prevalent type 2 diabetes (**Table 1**). The median follow-up for the entire study population was 4 years (a follow-up of up to 15 years.) The mean age at baseline was 64.3±10.1 years and the majority of patients included in this study were females (58.3%). 64.5% of individuals at baseline were overweight/obese. Median serum fasting glucose, insulin concentrations and HOMA-IR index (interquartile range (IQR)) were 5.4 (5.0-5.8), 67.0 (46.0-97.0), and 2.3 (1.5-3.4), respectively.

Characteristics of the baseline statin users and never users are shown in Table 1. Compared to never users, statin users had significantly higher BMIs, TGs, fasting serum glucose and insulin concentrations and insulin resistance and higher prevalence rates of hypertension. Statin users had significantly lower TC, HDL-C and LDL-C as compared to never users. Serum glucose concentrations were weakly but significantly (*p*-value<0.001) correlated with both (ln) insulin concentrations (r=0.14) and (ln) HOMA-IR (r=0.25).

Table 1. Baseline characteristics of the study population

	Total (n=9,535)	Statin users (n=968)	Never statin users (n=8,567)	*P*-value
Age, years, mean ±SD	64.3±10.1	64.9±8.2	64.3±10.2	**0.04**
Gender (male), n (%)	3977 (41.7)	481 (49.7)	3496 (40.8)	**<0.001**
BMI, kg/m^2, mean ±SD	27.0±4.1	27.5±3.9	26.9±4.1	**<0.001**
Family history of diabetes, n (%)	524 (8.7)	53 (5.5)	471 (5.5)	0.98
Hypertension, n (%)	3472 (36.5)	651 (67.3)	2821 (32.9)	**<0.001**
Total cholesterol, mmol/l, median (IQR)	5.7 (5.1-6.4)	5.1 (4.4-5.7)	5.8 (5.2-6.5)	**<0.001**
HDL-C, mmol/l, median (IQR)	1.4 (1.1-1.7)	1.3 (1.1-1.5)	1.4 (1.2-1.7)	**<0.001**
LDL-C, mmol/l, median (IQR)	4.0 (3.4-4.6)	3.4 (2.8-3.9)	4.1 (3.5-4.7)	**<0.001**
Triglycerides, mmol/l, median (IQR)	1.3 (1.0-1.8)	1.5 (1.1-2.1)	1.3 (0.97-1.7)	**<0.001**
Physical activity, MET-h/week, mean ±SD	75.6±49.6	75.4±47.2	75.7±49.8	0.44
Education status (High), n (%)	1717 (18.0)	153 (15.8)	1564 (18.3)	0.05
Smoking status (ever), n (%)	2040 (21.4)	222 (22.9)	1818 (21.2)	0.28
Alcohol consumption (ever), n (%)	5443 (57.1)	539 (55.7)	4904 (57.2)	0.32
Glucose, mmol/l, median, IQR	5.4 (5.0-5.8)	5.5 (5.1-5.9)	5.4 (5.0-5.8)	**0.02**
Insulin, pmol/l, median, IQR	67.0 (46.0-97.0)	78.0 (54.0-115.0)	66.0 (46.0-96.0)	**<0.001**
HOMA-IR index, median, IQR	2.3 (1.5-3.4)	2.7 (1.8-4.1)	2.3 (1.5-3.4)	**<0.001**
Type of statins				
Simvastatin (C10AA01)	529 (57.0)	-	-	-
Atorvastatin (C10AA05)	237 (25.5)	-	-	-
Pravastatin (C10AA03)	96 (10.3)	-	-	-
Fluvastatin (C10AA04)	66 (7.1)	-	-	-
Average statin dose, mg				
Simvastatin (C10AA01)	36	-	-	-
Atorvastatin (C10AA05)	42	-	-	-
Pravastatin (C10AA03)	33	-	-	-
Fluvastatin (C10AA04)	54	-	-	-

Abbreviations: SD, standard deviation; IQR, interquartile range; HDL-C, high density lipoproteins-cholesterol; LDL-C, low density lipoproteins-cholesterol; HOMA-IR, homeostasis model assessment-Insulin resistance

After excluding patients who used statins at baseline (n=968), the remaining study participants were subsequently followed for a period up to 15 years for incident type 2 diabetes. There were 716 cases of incident type 2 diabetes. In approximately 12.4% of the study population, statins were used during follow-up. In the majority of current statin users, the duration of use was longer than 365 days (75.5%). The most frequently dispensed statins were simvastatin (57.0%), atorvastatin (25.5%), and pravastatin (10.3%).

The association of statin use with glycaemic traits.

As shown in Table 2, in a cross-sectional analysis, baseline statin use was statistically significantly associated with increased serum fasting insulin concentrations (β, 0.07; 95%CI: 0.02-0.13) and HOMA-IR index (β, 0.09; 95%CI: 0.03-0.14). However, the association with serum fasting glucose concentrations was no longer significant after adding BMI and hypertension to the model 1 (β, 0.08; 95%CI: -0.03-0.18). Subgroup multiple linear regression analyses showed no role of statin type and dosage on these associations (**Table 2**).

The association of statin use with incident type 2 diabetes.

In a longitudinal follow-up study, compared with never statin users, ever statin use was associated with incident type 2 diabetes (crude HR, 1.64; 95%CI: 1.37-1.97). The observed association remained statistically significant even after adjusting for a range of potential confounders; model 2: HR, 1.38; 95%CI: 1.09-1.74 (**Table 3**). Current but not past use of statins was associated with a greater risk of type 2 diabetes; model 2: HR, 1.52; 95%CI: 1.15-2.00 vs. HR, 1.18; 95%CI: 0.83-1.67.

Among ever statin users, no statistically significant effect of statin type and dosage on risk of incident type 2 diabetes was found. However, we found a significant association of statin duration with incident type 2 diabetes in which longer duration of statin use was associated with a statistically significantly increased risk of incident type 2 diabetes (**Table 3**).

SUBGROUP ANALYSES

The association of cumulative statin use with incident type 2 diabetes.

In a fully adjusted model, cumulative exposure to statins was associated with a 35% higher risk of type 2 diabetes (HR, 1.35; 95% CI: 1.01-1.82). Among statin users, statin dose did not modify the association; HR, 1.12; 95% CI: 0.87-1.44.

Table 2. Multivariate linear regression analysis on the association between statin therapy and serum fasting glucose/insulin concentrations and HOMA-IR

	Crude effect	Model 1	Model 2
Glucose concentration	**β coefficient (SEM); 95% CI**	**β coefficient (SEM); 95% CI**	**β coefficient (SEM); 95% CI**
Baseline statin users	0.12 (0.07); -0.02-0.25	**0.14 (0.06); 0.04-0.27**	0.08 (0.06); -0.03-0.18
Subgroup analyses			
Statin type			
Simvastatin	0.03 (0.09); -0.14-0.20	-0.01 (0.06); -0.14-0.11	0.01 (0.05); -0.08-0.11
Atorvastatin	-0.10 (0.10); -0.29-0.10	0.02 (0.08); -0.13-0.18	-0.06 (0.06); -0.17-0.06
Others*	Ref		
Statin dose	-0.13 (0.07); -0.27-0.01	0.02 (0.06); -0.09-0.14	-0.07 (0.04); -0.16-0.01
(ln) Insulin concentration			
Baseline statin users	**0.15 (0.02); 0.11-0.20**	**0.13 (0.03); 0.08-0.19**	**0.07 (0.03); 0.02-0.13**
Subgroup analyses			
Statin type			
Simvastatin	0.01 (0.04); -0.07-0.09	0.08 (0.05); -0.02-0.18	0.05 (0.05); -0.04-0.15
Atorvastatin	**0.11 (0.05); 0.01-0.20**	0.05 (0.06); -0.08-0.17	0.05 (0.06); -0.07-0.16
Others*	Ref		
Statin dose	**0.19 (0.03); 0.13-0.25**	0.03 (0.05); -0.06-0.12	-0.01 (0.04); -0.09-0.08
(ln) HOMA-IR			
Baseline statin users	**0.19 (0.02); 0.14-0.23**	**0.16 (0.03); 0.10-0.21**	**0.09 (0.03); 0.03-0.14**
Subgroup analyses			
Statin type			
Simvastatin	0.01 (0.04); -0.07-0.10	0.08 (0.05); -0.03-0.19	0.05 (0.05); -0.04-0.15
Atorvastatin	0.10 (0.05); -0.02-0.19	0.04 (0.07); -0.09-0.17	0.04 (0.06); -0.08-0.16
Others*	Ref		
Statin dose	**0.19 (0.04); 0.12-0.25**	0.03 (0.05); -0.07-0.13	-0.02 (0.05); -0.11-0.07

Abbreviation: SEM, standard error of the mean; Ln, natural logarithm; CI, confidence interval; HOMA-IR, homeostasis model assessment-insulin resistance.
Model 1 adjusted for age, gender, cohort (I, II and III), smoking status, alcohol consumption, physical activity, and education level, and model 2 included all variables in model 1 addition to body mass index and hypertension.
*Others including pravastatin and fluvastatin.

The association of cumulative statins with incident impaired fasting glucose.

In a Cox regression analysis, cumulative statin use was associated with a 9% higher risk of incident impaired fasting glucose; HR, 1.09; 95% CI: 1.08-1.10 after adjusting for multiple confounders (model 2). Among statin users, in a fully adjusted model, the association between statin dose and incident impaired fasting glucose was

statistically significant in which higher dosages were related to an increased risk of incident impaired fasting glucose; HR, 2.34; 95% CI: 2.32-2.36.

Table 3. Univariate and multivariate Cox regression analysis on the associations of statin use with incident type 2 diabetes

	Crude effect: HR; 95% CI	Model 1: HR; 95% CI	Model 2: HR; 95% CI
Type 2 diabetes (n=8,844)			
Ever statin users	**1.64; 1.37-1.97**	**1.49; 1.19-1.86**	**1.38; 1.09-1.74**
Past statin users	**1.44; 1.10-1.89**	1.20; 0.85-1.68	1.18; 0.83-1.67
Current statin users	**1.63; 1.32-2.02**	**1.61; 1.23-2.10**	**1.52; 1.15-2.00**
Subgroup analyses-among ever statin users			
Statin type			
Simvastatin	0.98; 0.65-1.50	0.92; 0.55-1.53	0.91; 0.54-1.55
Atorvastatin	1.09; 0.69-1.72	1.50; 0.88-2.56	1.73; 0.99-3.01
Others*	Ref		
Statin dose	1.04; 0.92-1.17	1.02; 0.85-1.23	1.01; 0.79-1.28
Statin duration			
<30 days	1.04; 0.43-2.51	0.31; 0.04-2.21	0.29; 0.04-2.04
31-365	**2.02; 1.48-2.74**	**1.66; 1.11-2.49**	**1.70; 1.13-2.56**
>365 days	**1.51; 1.21-1.89**	**1.51; 1.15-1.98**	**1.37; 1.04-1.81**
Never statin users	Ref		

Abbreviation: HR, hazard ratio; CI, confidence interval
Model 1 adjusted for age, gender, cohort (I, II and III), smoking status, alcohol consumption, physical activity, and education level, and model 2 included all variables in model I addition to body mass index and hypertension.
*Others including pravastatin and fluvastatin.

Sensitivity analyses for the association between statin use and incident type 2 diabetes

In a series of sensitivity analyses, when we excluded cases with impaired fasting glucose at baseline, the positive association did not change (crude HR, 1.58; 95% CI: 1.23-2.04); however, when we added BMI and hypertension to the model 1, the association was not statistically significant anymore (HR, 1.28; 95% CI: 1.00-1.76). Using postbaseline values for BMIs did not change the effect estimates. Further adjustment for TC, PPIs, fasting serum glucose and insulin concentrations did not affect the association (**Table 4**). Substituting TC with other blood lipids (HDL-C, LDL-C and TG) attenuated the observed associations. Stratification by age resulted in a statistically significant association only in individuals >65 years (HR, 1.36; 95% CI: 1.02-1.83) but not in those younger than 65 years (HR, 1.25; 95% CI: 0.85-1.84). In gender-stratified analyses, the association between statin use and risk of incident type 2 diabetes was

only statistically significant in males (HR, 1.52; 95% CI: 1.07-2.16) but not in females (HR, 1.28; 95% CI: 0.93-1.74). When we stratified the analysis by baseline BMI, our findings showed a statistically significant association only in a subset of overweight/obese subjects (HR, 1.42; 95% CI: 1.10-1.83) but not in those with normal BMI (HR, 1.18; 95% CI: 0.69-2.02). The interaction term of statin ever use with serum fasting glucose concentrations was highly significant exerting the highest odds of incident type 2 diabetes (HR, 3.51; 95% CI: 2.51-4.91).

Table 4. Sensitivity analyses on the associations of statin ever use with incident type 2 diabetes

	Model 2: HR; 95% CI
Excluding cases with impaired fasting glucose at baseline	1.28; 1.00-1.76
Model 2+TC	**1.42; 1.12-1.80**
Model 2+HDL-C, LDL-C and TG	1.26; 0.99-1.61
Model 2+F.H.D	1.26; 0.98-1.62
Model 2+ PPIs	**1.30; 1.03-1.65**
Model 2+ fasting serum glucose and insulin concentrations	**1.41; 1.12-1.78**
Interaction between statin ever use and serum fasting glucose concentrations	**3.51; 2.51-4.91**

Abbreviations: HR, hazard ratio; CI, confidence interval; TC, total cholesterol; HDL-C, high density lipoproteins-cholesterol; TG, triglycerides; F.H.D, family history of diabetes; PPIs; proton pump inhibitors.
Model 2 adjusted for age, gender, cohort (I, II and III), smoking status, alcohol consumption, physical activity, education level, body mass index and hypertension.

DISCUSSION

To the best of our knowledge, this is the first detailed population-based study to show the effect of statin use on incident type 2 diabetes in those free from diabetes at baseline. At baseline, statin use was associated with elevated values for several glycaemic traits. After excluding these baseline statin users, new ever statin users during a follow-up of up to 15 years showed a 38% increased risk of incident type 2 diabetes independent of statin type and dosage. All observed associations between statin use and incident type 2 diabetes remained significant even after adjusting for several potential risk factors for diabetes, e.g. age, physical activity and education level. However, there was a steep decrease in the effect estimates corresponding to the adjusted models.

Impaired glycaemic traits are important hallmarks of incident diabetes in which impaired glucose metabolism and decreased insulin sensitivity are the two major pathophysiological disturbances required for the conversion to overt type 2 diabetes [19]. Currently, limited evidence is available concerning the impact of statin use on

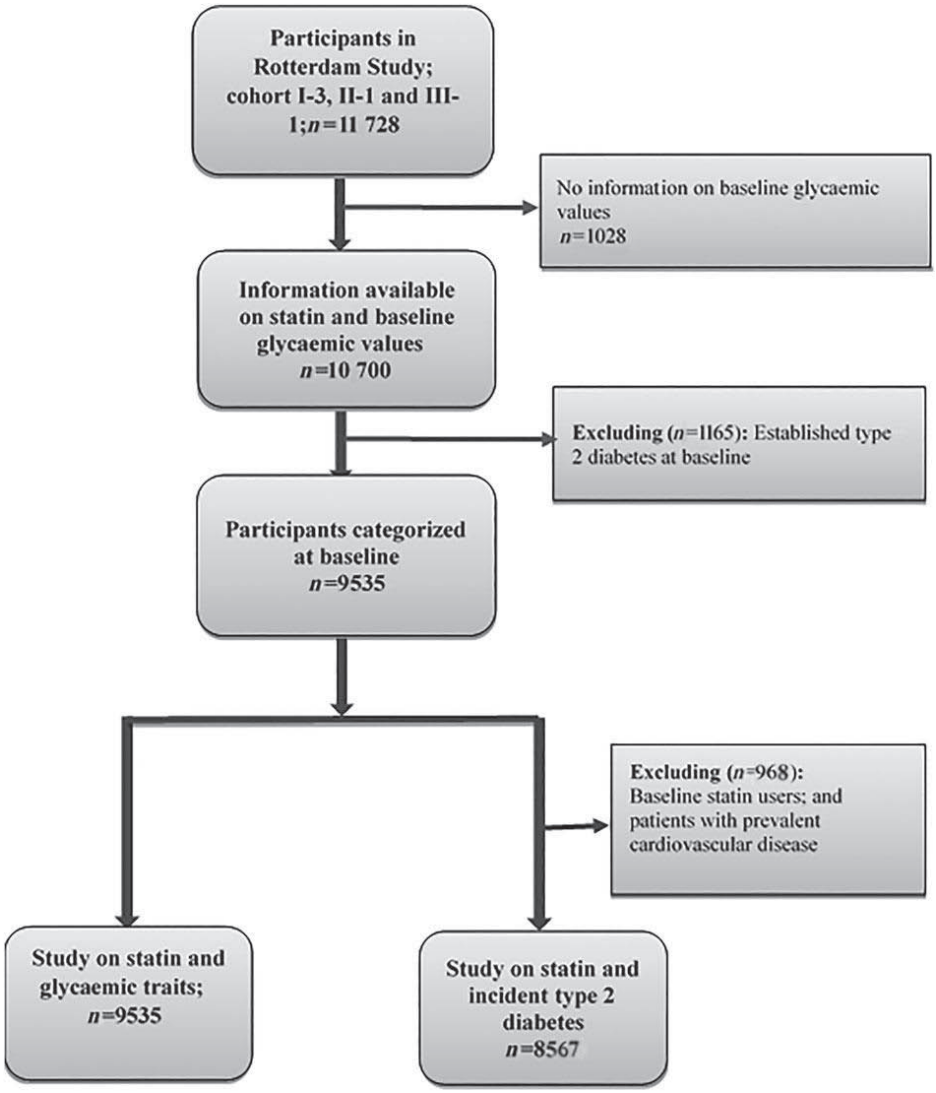


Fig 1. Flowchart of study population for the current study

glycaemic traits [20, 21]. Elevated baseline serum glucose and insulin concentrations might be explained by a diabetogenic effect of statins. However, conclusions from previous studies on this effect is inconsistent [22-24]. Theoretically, some potential underlying mechanisms for this effect in diabetes-free individuals consist of i) decreased insulin-mediated cellular glucose uptake leading to glucose intolerance, ii) decreased isoprenoids synthesis which causes downregulation of glucose transporter 4 and leads to hyperglycaemia and hyperinsulinemia, and iii) reduced factors such as coenzyme Q10, farnesyl pyrophosphate, and dolichol leading to altered insulin

secretion/resistance [25-28]. Since the association of statin use and serum fasting glucose concentrations was no longer significant after including BMI and hypertension, our finding suggests that the association between statins and diabetes could be through insulin secretion/resistance.

In our study, the significantly higher risk of incident type 2 diabetes among statin users confirms and extends the results of previous studies [6, 24, 29]. Several observational studies and RCTs have already reported an increased risk of incident type 2 diabetes in subjects treated with statins; where the increased risk reported by observational studies is much higher (44%) compared with RCTs (9–13%) [6, 24]. Discrepant results might be explained by the much longer follow-up time in observational studies than in RCTs and a relatively less healthy population as trials preferably enrol participants without multimorbidity. Longer time could result in an increased probability of developing adverse drug events such as diabetes [29]. According to the findings of a previous study including over 2 million subjects in the UK, the statin-associated diabetes risk (HR=1.22) over a period of 3 years was increased with longer statin duration in which the strongest significant association was observed when patients were followed-up for 15–20 years (HR=3.63) [30]. The magnitude of the association between statins and incident type 2 diabetes is also partly dependent on different confounders such as lipid profile and family history of diabetes [31]. In the 3 large RCTs, baseline fasting glucose concentrations and features of the metabolic syndrome e.g. TG were shown as predictive factors of incident type 2 diabetes among statin users [32]. Similarly, we showed that the association between statin therapy and incident type 2 diabetes was partly dependent on lipid profiles including HDL-C, LDL-C and TG, though, independent of several diabetes risk factors such as hypertension. Moreover, our result of the interaction between statin use and serum fasting glucose concentrations was highly significant implying the greatest statin-associated diabetes risk in presence of baseline impaired glucose concentrations. Our finding here is consistent with the results of JUPITER trial where the majority of statin users who developed diabetes during 5 years of follow-up had already impaired fasting glucose. However, this study was restricted to rosuvastatin at a single dosage of 20 mg daily [33].

In our study, the majority of patients were on lipid-soluble statins (simvastatin, atorvastatin and fluvastatin) while the number of patients on water-soluble statins (pravastatin) was low (10.3%). Therefore, due to different actions of specific molecules of statins on insulin resistance and metabolic risk profiles, further studies are required. Our findings concerning a significantly increased risk of incident type 2 diabetes independent of statin types and dosages are in line with a previous meta-

analysis [34]. To date, limited evidence suggests that different statin therapy including different types, dose and duration might exert distinct effects on statin related outcomes and the results reported by previous studies are conflicting [6, 34]. In line with our hypothesis of a cumulative effect of statins on impaired glucose and insulin resistance, the risk of incident diabetes was significantly higher in intermediate (31-365 days)- and long-term (>365 days) users of statins.

We showed a significantly increased risk of statin-related type 2 diabetes only among overweight/obese but not in individuals with a normal BMI. Given the concern that BMI values above normal and dyslipidaemia are both associated with the onset of type 2 diabetes [35], it is important to focus on more effective prevention in individuals with a high risk taking this medication. This could also help improving their lipid profile and thereby reducing the risk of cardiovascular diseases.

Our study has some strengths and limitations. First, the population-based and prospectively gathered information on disease outcome make selection- and information bias unlikely. The long follow-up, the availability of data on statin type, dose and duration of use along with the use of a series of adjusted regression models with a broad range of potential confounders are among the most important strengths in this study. Moreover, we were able to provide a comprehensive overview of the associations between statins and glycaemic traits and incident type 2 diabetes. However, our study has some potential limitations that should be acknowledged. Most importantly, to study the association between statins and incident type 2 diabetes, we excluded prevalent type 2 diabetes cases at baseline (n=1,565), and included only participants with normal fasting glucose levels (<6.1 mmol/l) who were not on hypoglycaemic medication. Unfortunately, we did not have data on HbA1c at baseline, a marker with the ability of reflecting the long-term glycemic history. Also, the association between statins and type 2 diabetes could be caused by reverse causation in which high risk individuals e.g. patients with obesity and fatty liver were prescribed statins to prevent CVD. However, to decrease the effect of reverse causality on our findings, we performed two subgroup analyses including the associations of type 2 diabetes with cumulative statin use until one year before diabetes and an impaired fasting glucose, where the results still showed significant associations. Cumulative statin users during follow-up showed a 9% increased risk of incident impaired fasting glucose in which the magnitude of increased risk was directly proportional to the statin dosage. Though, we were able to adjust the associations for many measured confounding variables, the possibility remains that some factors e.g. genetic factors which we have not measured still modify the associations [36]. Additionally, compared with never statin users, patients treated

with statins tend to be sicker and are more prone to enhanced glucose and insulin concentrations, and to eventually developing diabetes [37]. We cannot rule out the effect of detection bias in our study in which individuals treated with statins are more likely to be clinically evaluated. Finally, since our study included a population consisting of roughly 95% Caucasian ethnicity and race might be an important factor for the association between statin use and susceptibility of diabetes, our findings cannot be easily generalized to other populations.

In summary, using multiple subgroup and sensitivity analyses, we found a consistent significant association between statin use and incident type 2 diabetes. Our analysis also highlights the greatest effect of statins in presence of hyperglycaemia and overweight/obesity. This suggests that it is necessary to take statin diabetogenicity into consideration in clinical practice when statin is indicated, emphasising the concomitant need for dietary measures and exercise.

REFERENCES

1. Baigent C, Landray M, Reith C, et al. Study of heart and renal protection (SHARP): Randomized trial to assess the effects of lowering low-density lipoprotein cholesterol among 9,438 patients with chronic kidney disease. Am Heart J. 2010;160(5):785-U28.
2. Cholesterol Treatment Trialists' (CTT) Collaborators, Kearney PM, Blackwell L, Collins R, Keech A, Simes J, Peto R, Armitage J, Baigent C. Efficacy of cholesterol-lowering therapy in 18,686 people with diabetes in 14 randomised trials of statins: A meta-analysis. Lancet 2008 Jan 12;371(9607):117-125
3. Culver AL, Ockene IS, Balasubramanian R, et al. Statin use and risk of diabetes mellitus in postmenopausal women in the women's health initiative. Arch Intern Med 2012 Jan 23;172(2):144-152
4. Cederberg H, Stancakova A, Yaluri N, Modi S, Kuusisto J, Laakso M. Increased risk of diabetes with statin treatment is associated with impaired insulin sensitivity and insulin secretion: A 6 year follow-up study of the METSIM cohort. Diabetologia 2015 May;58(5):1109-1117
5. Thakker D, Nair S, Pagada A, Jamdade V, Malik A. Statin use and the risk of developing diabetes: A network meta-analysis. Pharmacoepidemiol Drug Saf 2016 Oct;25(10):1131-1149
6. Casula M, Mozzanica F, Scotti L, et al. Statin use and risk of new-onset diabetes: A meta-analysis of observational studies. Nutr Metab Cardiovasc Dis 2017 May;27(5):396-406
7. Chen YH, Chen YC, Liu CS, Hsieh MC. The different effects of atorvastatin and pravastatin on cell death and PARP activity in pancreatic NIT-1 cells. J Diabetes Res 2016;2016:1828071
8. Lorza-Gil E, Salerno AG, Wanschel AC, et al. Chronic use of pravastatin reduces insulin exocytosis and increases beta-cell death in hypercholesterolemic mice. Toxicology 2016 Feb 17;344-346:42-52
9. Rajpathak SN, Kumbhani DJ, Crandall J, Barzilai N, Alderman M, Ridker PM. Statin therapy and risk of developing type 2 diabetes: A meta-analysis. Diabetes Care 2009 Oct;32(10):1924-1929
10. Sattar N, Preiss D, Murray HM, et al. Statins and risk of incident diabetes: A collaborative meta-analysis of randomised statin trials. Lancet 2010 Feb 27;375(9716):735-742
11. Ikram MA, Brusselle GGO, Murad SD, et al. The rotterdam study: 2018 update on objectives, design and main results. Eur J Epidemiol 2017 Oct 24
12. Neeley WE. Simple automated determination of serum or plasma glucose by a hexokinase-glucose-6 -phosphate dehydrogenase method. Clin Chem 1972 Jun;18(6):509-515
13. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: Insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia 1985 Jul;28(7):412-419
14. Ligthart S, van Herpt TT, Leening MJ, et al. Lifetime risk of developing impaired glucose metabolism and eventual progression from prediabetes to type 2 diabetes: A prospective cohort study. Lancet Diabetes Endocrinol 2016 Jan;4(1):44-51
15. Stel VS, Smit JH, Pluijm SM, Visser M, Deeg DJ, Lips P. Comparison of the LASA physical activity questionnaire with a 7-day diary and pedometer. J Clin Epidemiol 2004 Mar;57(3):252-258

16. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 1972 Jun;18(6):499-502
17. Stricker BH, Stijnen T. Analysis of individual drug use as a time-varying determinant of exposure in prospective population-based cohort studies. Eur J Epidemiol 2010 Apr;25(4):245-251
18. Lin HC, Hsiao YT, Lin HL, et al. The use of proton pump inhibitors decreases the risk of diabetes mellitus in patients with upper gastrointestinal disease: A population-based retrospective cohort study. Medicine (Baltimore) 2016 Jul;95(28):e4195
19. Singh B, Saxena A. Surrogate markers of insulin resistance: A review. World J Diabetes 2010 May 15;1(2):36-47
20. Sukhija R, Prayaga S, Marashdeh M, et al. Effect of statins on fasting plasma glucose in diabetic and non-diabetic patients. J Investig Med 2009 Mar;57(3):495-499
21. Parida S, Swain TR, Routray SN, Maiti R. Effect of atorvastatin on glycaemic parameters in normoglycaemic and prediabetic subjects: A prospective, panel study. J Clin Diagn Res 2017 Feb;11(2):FC04-FC09
22. Kostapanos MS, Milionis HJ, Agouridis AD, Rizos CV, Elisaf MS. Rosuvastatin treatment is associated with an increase in insulin resistance in hyperlipidaemic patients with impaired fasting glucose. Int J Clin Pract 2009 Sep;63(9):1308-1313
23. Chan DC, Pang J, Watts GF. Pathogenesis and management of the diabetogenic effect of statins: A role for adiponectin and coenzyme Q10?. Curr Atheroscler Rep 2015 Jan;17(1):472-014-0472-7
24. Millan Nunez-Cortes J, Cases Amenos A, Ascaso Gimilio JF, et al. Consensus on the statin of choice in patients with impaired glucose metabolism: Results of the DIANA study. Am J Cardiovasc Drugs 2017 Apr;17(2):135-142
25. Wang S, Cai R, Yuan Y, Varghese Z, Moorhead J, Ruan XZ. Association between reductions in low-density lipoprotein cholesterol with statin therapy and the risk of new-onset diabetes: A meta-analysis. Sci Rep 2017 Jan 10;7:39982
26. Kanda M, Satoh K, Ichihara K. Effects of atorvastatin and pravastatin on glucose tolerance in diabetic rats mildly induced by streptozotocin. Biol Pharm Bull 2003 Dec;26(12):1681-1684
27. Andersson C, Lyass A, Larson MG, Robins SJ, Vasan RS. Low-density-lipoprotein cholesterol concentrations and risk of incident diabetes: Epidemiological and genetic insights from the framingham heart study. Diabetologia 2015 Dec;58(12):2774-2780
28. Chen WM, Sheu WH, Tseng PC, et al. Modulation of microRNA expression in subjects with metabolic syndrome and decrease of cholesterol efflux from macrophages via microRNA-33-mediated attenuation of ATP-binding cassette transporter A1 expression by statins. PLoS One 2016 May 3;11(5):e0154672
29. Mansi I, Frei CR, Wang CP, Mortensen EM. Statins and new-onset diabetes mellitus and diabetic complications: A retrospective cohort study of US healthy adults. J Gen Intern Med 2015 Nov;30(11):1599-1610
30. Macedo AF, Douglas I, Smeeth L, Forbes H, Ebrahim S. Statins and the risk of type 2 diabetes mellitus: Cohort study using the UK clinical practice pesearch datalink. BMC Cardiovasc Disord 2014 Jul 15;14:85-2261-14-85

31. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB S. Prediction of incident diabetes mellitus in middle-aged adults: The framingham offspring study. Arch Intern Med 2007 May 28;167(10):1068-1074
32. Waters DD, Ho JE, DeMicco DA, et al. Predictors of new-onset diabetes in patients treated with atorvastatin: Results from 3 large randomized clinical trials. J Am Coll Cardiol 2011 Apr 5;57(14):1535-1545
33. Ridker PM, Pradhan A, MacFadyen JG, Libby P, Glynn RJ. Cardiovascular benefits and diabetes risks of statin therapy in primary prevention: An analysis from the JUPITER trial. Lancet 2012 Aug 11;380(9841):565-571
34. Navarese EP, Buffon A, Andreotti F, et al. Meta-analysis of impact of different types and doses of statins on new-onset diabetes mellitus. Am J Cardiol 2013 Apr 15;111(8):1123-1130
35. Abhay kumar pandey , deepti pandey , abha pandit. obesity and lipid profile study in type 2 diabetes patients with auditory and reaction time deficits and non-diabetic control subjects. advances in diabetes and metabolism. 2017; 5 , 1 - 5. doi: 10.13189/adm.2017.050101.
36. Stancakova A, Kuulasmaa T, Kuusisto J, et al. Genetic risk scores in the prediction of plasma glucose, impaired insulin secretion, insulin resistance and incident type 2 diabetes in the METSIM study. Diabetologia 2017 Sep;60(9):1722-1730
37. Schrom JR, Caraballo PJ, Castro MR, Simon GJ. Quantifying the effect of statin use in pre-diabetic phenotypes discovered through association rule mining. AMIA Annu Symp Proc 2013 Nov 16;2013:1249-1257

Chapter 3.2.

Epigenetic Link Between Statin Therapy and Type 2 Diabetes

Carolina Ochoa-Rosales, Eliana Portilla, Jana Nano, Rory Wilson, Benjamin Lehne, Pashupati Mishra, Xu Gao, Mohsen Ghanbari, Oscar Rueda-Ochoa, Diana Juvinao-Quintero, Marie Loh, Weihua Zhang, Jaspal Kooner, Hans Grabe, Stephan Felix, Ben Schöttker, Yan Zhang, Christian Gieger, Martina Müller-Nurasyid, Margit Heier, Annette Peters, Terho Lehtimäki, Alexander Teumer, Hermann Brenner, Melanie Waldenberger, M. Arfan Ikram, Joyce B.J. van Meurs, Oscar H. Franco, Trudy Voortman, John Chambers, Bruno H. Stricker, Taulant Muka.

Epigenetic Link Between Statin Therapy and Type 2 Diabetes. *Diabetes Care 2020;43:875-84.*

ABSTRACT

We aimed to investigate the role of epigenetics on statins' diabetogenic effect comparing DNA methylation (DNAm) between statin users and non-users, in an epigenome-wide association study (EWAS) in blood.

Five cohort studies' participants (*n*=8,270) were classified as statin users when they were on statin therapy at the time of DNAm assessment with Illumina 450K or EPIC array, or non-current users otherwise. Associations between DNAm and various outcomes like incident type 2 diabetes, plasma glucose, insulin and insulin resistance (HOMA-IR) as well as with gene expression were investigated.

Discovery (*n*=6,820) and replication phases (*n*=1,450) associated five DNAm sites to statin use: cg17901584 ($1.12x10^{-25}$; *DHCR24*), cg10177197 ($3.94x10^{-08}$; *DHCR24*), cg06500161 ($2.67x10^{-23}$; *ABCG1*), cg27243685 ($6.01x10^{-09}$; *ABCG1*), cg05119988 ($7.26x10^{-12}$; *SC4MOL*). Two sites were associated with at least one glycemic trait or type 2 diabetes. Higher cg06500161 methylation was associated with higher fasting glucose, insulin, HOMA-IR and type 2 diabetes (odds ratio 1.34 (0.20, 0.39)). Mediation analyses suggested *ABCG1* methylation to partially mediate the effect of statins on high insulin and HOMA-IR. Gene expression analyses showed that statins exposure and *ABCG1* methylation were associated with *ABCG1* downregulation, suggesting epigenetic regulations of *ABCG1* expression. Further, outcomes insulin and HOMA-IR were significantly associated with *ABCG1* expression.

This study sheds light on potential mechanisms linking statins with type 2 diabetes risk, providing evidence on DNAm partially mediating statins effects on insulin traits. Further efforts shall disentangle the molecular mechanisms through which statins may induce DNAm changes, potentially leading to *ABCG1* epigenetic regulation.

INTRODUCTION

Statins effectively reduce the risk of cardiovascular disease.[1] However, clinical trials and observational studies show that statins lead to insulin resistance and type 2 diabetes.[2,3] The underlying mechanisms remain unclear.

Statins are associated with epigenetic changes, including histone acetylation, microRNA regulation[4] and DNA methylation (DNAm), particularly at genes related to lipid and insulin metabolism.[5] DNAm is linked to type 2 diabetes pathophysiology,[6] thus it may be a potential mechanism contributing to the increased risk of type 2 diabetes observed in statin therapy. Nevertheless, this hypothesis has not been investigated.[4]

We conducted an Epigenome-wide Association Study in blood investigating the association between statin use and changes in DNAm at sites in the genome called CpGs. Further, we sought to replicate the findings, study the associations of the statin-related CpGs with gene expression, glycemic traits and type 2 diabetes. Finally, we explored he potential role of the statin-related CpGs mediating the association of statins with glycemic traits and type 2 diabetes.

METHODS

Study design and population

Overall, 8,270 participants from five prospective cohort studies were included. In EWAS analyses, the discovery panel included 6,820 Caucasian individuals from: Epidemiologische Studie zu Chancen der Verhütung, Früherkennung und optimierten Therapie chronischer Erkrankungen in der älteren Bevölkerung (ESTHER),[7] an population-based study recruited in Saarland, Germany, with participants aged 50 to 75 years; Kooperative Gesundheitsforschung in der Region Augsburg-F4 (KORA-F4),[8] a follow-up study to KORA-S4 cohort, established in 1996 in Augsburg, Germany; Study of Health Pomerania-Trend (SHIP-Trend),[9] a population-based study of participants from northeast Germany; and a population-based cohort of South-Asians residing in London, United Kingdom, aged 35 to 75 years: The London Life Sciences Prospective Population Study (LOLIPOP).[10] Our study included 998 individuals from ESTHER; 1,726 from KORA-F4; 247 from SHIP-Trend and 3,848 from LOLIPOP. Hence, the discovery panel was composed by 56% South-Asians and 44% Caucasians.

The replication panel was selected based on ethnical homogeneity and availability of data for post-EWAS analyses. It included 1,450 Caucasians from the Rotterdam Study (RS; subcohorts RS-III-1 and RS-BIOS),[11] a prospective population-based cohort study of participants living in Ommoord, Rotterdam, the Netherlands, aged 45 years and over. Only RS had information on former statin users (*n*=74). Additional associations analyses of DNAm with glycemic traits (*n*=1,165) and gene expression (*n*=731) included RS data, whereas associations with incident type 2 diabetes used RS, KORA and LOLIPOP (*n*=4,422). See Figure 1 for study design.

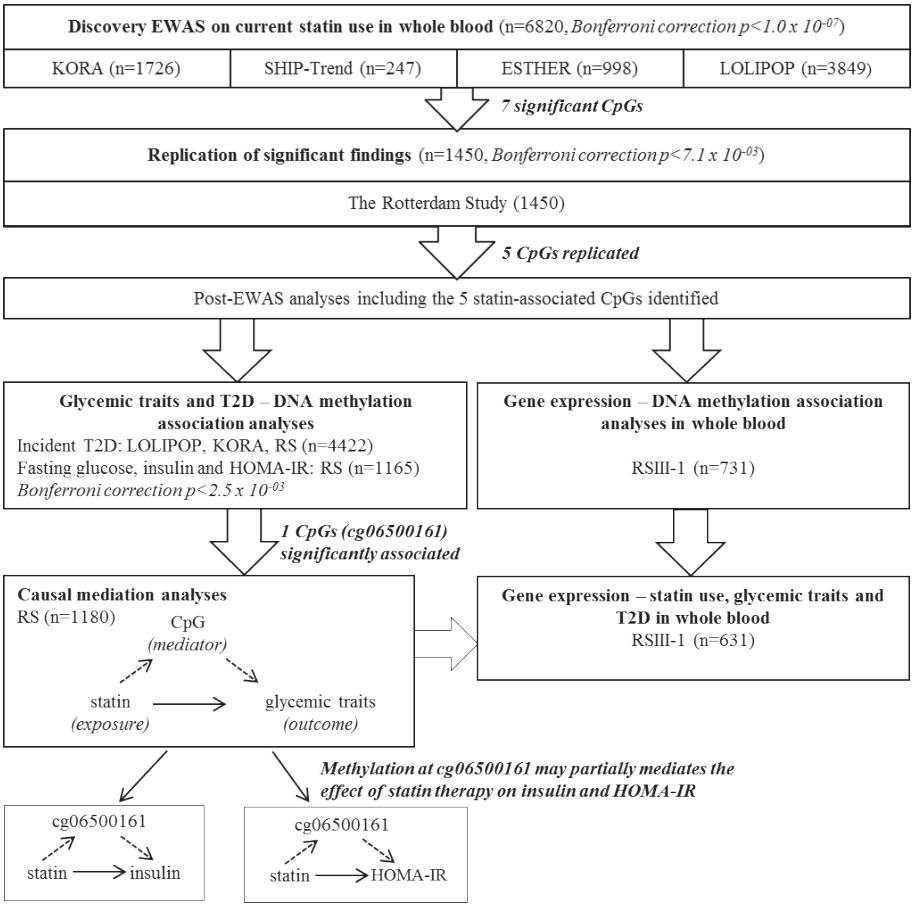


Figure 1. Study design. An epigenome-wide association study was conducted to identify DNA methylation changes at CpG sites related to current statin use, compared to non-current statin use. Four cohorts were used as a discovery panel and finding were replicated in an independent cohort. The replicated CpG sites were investigated in their association with glycemic traits and incident type 2 diabetes. The significant associations from this analysis were tested in a causal mediation analysis. Association analyses of the identified CpGs, current statin use and glycemic traits with gene expression were performed.

Statin use

Information on medication was obtained from pharmacy records in RS, self-declaration in KORA-F4, LOLIPOP and SHIP-Trend, and general practitioner's records in ESTHER.

Statin-use status was categorized in current users and non-current users. Current users were defined if the statins prescription occurred at the time of blood drawing for DNAm assessment, or as non-current users otherwise. RS additionally identified former statin users as participants who had previously used statins but were no longer users on the blood draw date.

DNA methylation data

DNA was extracted from whole peripheral blood (stored in EDTA tubes) by standardized salting out methods. Genome-wide DNAm was measured in bisulfite-converted genomic DNA using the Illumina-Infinium Methylation 450K or Illumina-Infinium EPIC BeadChip, according to manufacturer's protocols. Samples showing inadequate hybridization, incomplete bisulfite treatment and gender swaps were excluded. Details on cohort-specific methods in eTable 1. CpG methylation proportion was reported as a normalized β-value ranging from 0 to 1, where 1 represents 100% methylation. To avoid sex bias, CpGs annotated to genes in sex chromosomes were excluded. After normalization and quality control, overlapping CpGs across studies were meta-analyzed. CpGs missing in one cohort were also included, while CpGs missing in two or more cohorts were excluded, leaving 332,357 CpGs for final analysis.

Messenger rna expression data

Messenger RNA data included a subset of the replication panel (n=731, RSIII-1). Total RNA was isolated (PAXgene Blood RNA kits—Qiagen) from whole blood (PAXgene Tubes—Becton Dickinson). Samples were analysed using the LabChip GX (Caliper) according to the manufacturer's instructions. Samples with an RNA quality score >7 were amplified, labelled (Ambion TotalPrep RNA) and hybridized to the Illumina HumanHT-12-v4 Expression BeadChips, according to the manufacturer's protocol. Expression data was quantile normalized to the median distribution and $\log^2$ transformed. Probe and sample means were centred to zero. Genes were considered significantly expressed when detection P-values calculated by Genome Studio were <0.05 in more than 10% of all discovery samples. Thus, 21,238 probes were available to study. Quality control used eQTL mapping pipeline (https://github.com/molgenis/systemsgenetics/tree/master/eqtl-mapping-pipeline).[12] Only probes that uniquely mapped to the human genome build 37 and represented gene mRNA expression

were analyzed.[13] Samples were processed at the Genetic Laboratory of Internal Medicine, Erasmus Medical Center, Rotterdam. Expression data is available for 881 samples at Gene Expression Omnibus public repository (accession GSE338828).

Covariates, glycemic traits and type 2 diabetes

Covariates were selected based on previous literature on DNAm studies, including: age, sex, BMI, smoking status, history of coronary heart disease (CHD), anti-hypertensive medication, systolic blood pressure (SBP), total cholesterol, HDL, triglycerides and LDL. BMI was calculated by dividing weight in kilograms by height in meters squared. Smoking history was self-reported and classified as current, former and never smokers. SBP was the mean of 2 consecutive measurements at the right brachial artery with a random-zero sphygmomanometer with the participant in sitting position. Anti-hypertensive medication use was defined as the use of diuretics, β-blockers, angiotensin-converting-enzyme inhibitors and calcium channel blockers. CHD was defined as the occurrence of myocardial infarction, revascularization, coronary artery bypass graft surgery or percutaneous coronary intervention. Triglycerides, HDL, and total cholesterol concentrations were measured in blood using an automated enzymatic method.

Glycemic traits examined were plasma fasting glucose, insulin, homeostatic model assessment-insulin resistance (HOMA-IR) and type 2 diabetes incidence. Type 2 diabetes cases were defined as the presence of either serum glucose level ≥7.0 mmol/L or glucose-lowering medication use in RS; self-report in KORA; physician diagnosis or use of glucose-lowering medication in ESTHER; diabetes medication use, HbA1c ≥6.5%, serum glucose level (non-fasting) ≥11,1 mmol/l or self-report in SHIP-Trend; diabetes history or fasting glucose ≥7.0 mmo/L or HbA1c >6.5% in LOLIPOP. The interassay coefficients of variations are: insulin <8%, glucose <1.4% and lipids <2.1%. LDL was calculated using the Friedewald formula (total cholesterol − HDL − triglycerides/5).

STATISTICAL ANALYSES

Epigenome wide association study (EWAS)

The association between CpGs methylation and current versus non-current statin users was investigated in a two-step EWAS approach: discovery and replication. Normalized DNAm β-values at each CpG were the dependent variable and current statins use (yes=1, no=0) the predictor. Per individual CpG, participants with methylation levels higher than three times the interquartiles range (IQR) were excluded. The

regression analysis used the 'lme4' package in R v.3.4.2 (https://cran.r-project.org/web/packages/lme4/index.html) to perform linear mixed-effect models. The random part of the model included adjustment for array-number and position-on-array as random terms to account for batch effects in the DNAm measurement. The fixed part included leukocyte proportions (to account for cell mixture), sex, age, smoking status, BMI, SBP, anti-hypertensive medication, prevalent CHD and type 2 diabetes. Leukocyte proportions (B-cells, CD4, T-cells, CD8, T-cells, granulocytes, monocytes and NK-cells) were estimated in the R 'minfi' package, as described by Houseman.[14,15] Statistical significance for discovery was determined by a Bonferroni-corrected threshold (1.0×10^{-7}). Heterogeneity (I^2) of effect estimates was assessed to account for differences between cohorts. Results across the discovery cohorts were combined in a meta-analysis by inverse variance weighted method in METAL v.2011-03-25.

The identified CpGs were replicated in an independent panel using identical models. Bonferroni correction for replication was 0.05 divided by the number of discovery findings. Results of discovery and replication panels were combined in a meta-analysis as aforementioned.

Pearson test was used to assess the correlation among the replicated CpGs in the replication panel (RS).

Associations between statin-related CpGs and glycemic traits and incident type 2 diabetes

Standardized methylation values at the replicated CpGs as predictors of fasting glucose, insulin and HOMA-IR (RS, *n*=1,165) were tested in a linear mixed-effect models. To explore reverse causation, prevalent type 2 diabetes cases and former statin users were excluded. Logistic regression was used to assess the association with incident type 2 diabetes (RS, KORA, LOLIPOP, *n*=4,422, including 1,173 cases). Results across cohorts were meta-analyzed as aforementioned. The adjusting covariates included leukocyte proportions, batch effects, sex, age, smoking status, SBP, anti-hypertensive medication, CHD and statin use. Bonferroni-correction was applied (<0.0025). Variables with right-skewed distributions were transformed to the natural logarithmic scale (glucose, insulin, HOMA-IR, HDL).

Causal mediation analysis

To investigate whether DNAm changes are part of the pathway though which statins exert their diabetogenic effect, we explored the significant associations in the previous stage in a non-parametric causal mediation analysis using the R 'mediation' package,[16] as elsewhere.[17] The mediation analysis dissects the total effect of the

exposure (statins treatment) on the outcome (glycemic traits and incident type 2 diabetes) into the (i) indirect effect acting via the mediator of interest (DNAm), and (ii) direct effect acting directly or via a mediator other than the one under study. The mediation package compares the effect estimates of each association between statin use and traits, in subjects with different methylation levels. Non-parametric bootstrap with 20,000 resamples was performed. Data from the RS were used (n=1,180), excluding prevalent type 2 diabetes cases and former statin users. The normalized CpG methylation β-value was modeled as mediator and the CpG-associated trait as outcome. eFigure 1 shows an schematic representation. Sex, age and cohort were considered confounders. To investigate potential unmeasured confounding the sequential ignorability assumption (no confounder assumption) was tested in a sensitivity analysis using the correlated residuals method,[18] which estimates the correlation that quantifies how strong the confounder would have to be to change the conclusion.

Gene expression-association analysis

To investigate possible biological pathways, relations between the identified CpGs and transcriptome-wide gene expression were explored (RS, n=731). Statin-related CpGs were regressed against methylation batch variables as random terms; similar approach was applied for 21,238 transcription probes and the expression batch variables. Next, linear regressions between the residuals from the CpGs regression (predictors) and the expression probes residuals (dependent variable) were performed. The Bonferroni-correction was 0.05/number of probes times number of CpGs tested.

Additionally, the significant expression probes in the aforementioned analysis were tested in association with statin use exposure (n=631), and as predictors of the significant glycemic traits in the mediation analyses (n=616). Type 2 diabetes cases and former statin users were excluded.

Dose effects

In 303 current statin users from SHIP-Trend and RS with available data on statin dose, we examined the association between the replicated CpGs and the defined daily dose (DDD, WHO ATC/DDD-classification, https://www.whocc.no/atc_ddd_index/), using the EWAS model. Next, the association between gene expression and statin dose was studied in the RS.

Confounding and Ethnicity

To investigate whether the associations of statin use with the replicated CpGs were independent of blood lipids, we additionally adjusted the EWAS for serum total

cholesterol or for individual blood lipids instead (HDL, LDL and triglycerides) in both discovery and replication.

In an effort for ruling out confounding by indication, we restricted the EWAS to subjects with LDL ≥70 mg/dL or ≥100 mg/dL, cut-offs used to advice statins use.[19,20] Further, we reran the EWAS excluding prevalent type 2 diabetes and pre-diabetes cases.

To account for ethnicity heterogeneity in the discovery panel (South-Asians and Caucasians) we reran the EWAS, taking Caucasians (*n*=4,349) for discovery and South-Asians (*n*=3,849) for replication.

Post-hoc power analysis

To calculate the range of power for replication of the findings we utilized GPower 3.1 tool[21] with 6 predictors representing all discovered CpGs except for cg03467813 which was not included due to its high heterogeneity (I^2=99%). CpGs' smallest beta estimate (0.0039, cg10177197) and probability error 0.008, or the largest estimate (0.0219, cg17901584) were used.

RESULTS

Association between statin use and DNA methylation

The mean [SD] age in the discovery panel ranged from 51.5 [13.8] years (SHIP-Trend) to 62.1 [6.5] (KORA-F4). Of the total discovery panel, 2,969 (43.5%) participants were women. eTable 2a-b shows other baseline characteristics.

In the discovery panel, seven CpGs passed Bonferroni threshold for significance ($P<1.0\times10^{-7}$), being differentially methylated in current statins users compared to non-current users. Of them, current users had lower methylation at four CpGs, annotated to genes *DHCR24* (cg17901584, located at TSS1500), *FAM50B* (cg03467813, located at gene body), *SC4MOL* (cg05119988, located at 5´ untranslated region (5'UTR)) and *AHRR* (cg05575921, located at gene body). Three CpGs annotated to *ABCG1* (cg06500161, cg27243685, located at gene body) and *DHCR24* (cg10177197, located at gene body) showed higher methylation among current users (Table 1; eFigure 1). All associations were independently replicated, except for two CpGs annotated to *AHRR* and *FAM50B*, which failed to reach statistical significance after Bonferroni-correction for replication ($P<7.14\times10^{-3}$). The post-hoc power analysis showed a power ranging from 39% to 99% for our replication study. In the replication panel, correlations

among CpGs ranged from r=-0.011 between cg06500161 (*ABCG1*) and cg17901584 (*DHCR24*) to r=0.477 for cg06500161 and cg27243685 (both annotated to *ABCG1*) (eTable 3).

The meta-analysis across the discovery and replication panels revealed one new CpGs at *DHCR24* (cg17475467). The heterogeneity in the associations between statin use and the replicated CpGs across cohorts varied from I^2=0 for cg10177197 (*DHCR24*) to I^2=74.8 for cg17901584 (*DHCR24*) (Table 1). Unreplicated CpGs were not further investigated.

The sensitivity analysis using RS data with never statin users as reference group (former users excluded) revealed similar results, replicating the five CpGs (eTable 4).

Associations of the statin-related CpGs with glycemic traits and type 2 diabetes

After a Bonferroni-correction ($2.5x10^{-3}$) and comprehensive assessment of potential confounders, increase in one standard deviation of methylation at cg06500161 (*ABCG1*) was associated with incident type 2 diabetes (odds ratio 1.34 (95% CI 1.22, 1.47)), fasting glucose ($1.03x10^{-3}$), insulin ($4.63x10^{-8}$) and HOMA-IR ($1.05x10^{-8}$). Increase in one standard deviation of methylation at cg27243685 (*ABCG1*) was associated with incident type 2 diabetes (1.24 (1.13, 1.38)) and approaching borderline statistical significance with augmented insulin ($4.62x10^{-3}$) , but not statistically significantly associated with glucose and HOMA-IR. Increase in one standard deviation of methylation at cg17901584 was approaching borderline statistical significance with incident type 2 diabetes ($6.2x10^{-3}$), but was not significantly associated with glucose, insulin and HOMA-IR. No significant findings were observed for cg05119988 (*SC4MOL*) and cg10177197 (*DHCR24*) (Table 2).

Causal mediation

Significant results were obtained for the models with outcome fasting insulin and HOMA-IR. For both cases the analyses showed that: i) current statin use has a significant overall effect of β=0.275 (0.190, 0.360) on insulin, and β=0.291 (0.199, 0.380) on HOMA-IR; ii) part of that effect goes directly or via mediator(s) other than cg06500161, called average direct effect (ADE). We found an ADE of β=0.233 (0.149, 0.320) on insulin, and β=0.242 (0.151, 0.330) on HOMA-IR; iii) Another part of such effect may go via an indirect path (indirect effect) possibly through cg06500161 methylation, with an average causal mediator effect (ACME) of β=0.043 (0.024, 0.060) on insulin and β=0.049 (0.028, 0.070) on HOMA-IR; iv) consequently, of the total effect of statins on insulin and HOMA-IR, 15.5% (0.086, 0.260) and 16.8% (0.095,

Table 1: DNA methylation sites associated with current statin use (compared to non-current use)

					Discovery panel [a]			Replication panel [b]			Both panels combined				
CpG	Chr	Position	Loc	Gene	Effect	SE	P-value	Effect	SE	P-value	Effect	SE	P-value	I^2 Het	P-value Het
cg17901584	**1**	55353706	TSS 1500	***DHCR24***	-0.0219	0.0021	**1.12E-25**	-0.0143	0.0033	**1.47E-5**	-0.0197	0.0018	**6.03E-29**	**74.8**	**0.0013**
cg06500161	**21**	43656587	gene body	***ABCG1***	0.0103	0.0010	**2.67E-23**	0.0115	0.0023	**3.76E-7**	0.0105	0.0009	**6.59E-29**	52.3	0.0627
cg03467813	6	3851047	gene body	*FAM50B*	-0.0218	0.0023	**2.11E-21**	0.0008	0.0069	0.9073	-0.0196	0.0022	**2.77E-19**	**99.3**	**<0.001**
cg05119988	**4**	166251189	5' UTR	***SC4MOL***	-0.0120	0.0017	**7.26E-12**	-0.0128	0.0032	**7.55E-5**	-0.0121	0.0015	**2.57E-15**	6.3	0.3761
cg27243685	**21**	43642366	gene body	***ABCG1***	0.0044	0.0008	**6.01E-9**	0.0069	0.0015	**6.41E-6**	-0.0115	0.0021	**2.52E-8**	13.8	0.3259
cg05575921	5	373378	gene body	*AHRR*	-0.0132	0.0022	**3.20E-9**	-0.0011	0.0056	0.8445	0.0049	0.0007	**5.15E-13**	**73.6**	**0.0020**
cg10177197	**1**	55316481	gene body	***DHCR24***	0.0039	0.0007	**3.94E-8**	0.0054	0.0016	**7.10E-4**	0.0041	0.0006	**1.65E-10**	0	0.5073
cg17475467	1	55316769	3 'UTR	*DHCR24*	0.0033	0.0008	1.68E-5	0.0049	0.0013	**2.11E-4**	0.0037	0.0007	**2.41E-8**	0	0.4684

Model adjusted for leukocyte proportions, batch effects, sex, age, smoking status, body mass index, systolic blood pressure, anti-hypertensive medication, presence of coronary heart diseases, prevalent type 2 diabetes.
Bonferroni correction for significance for discovery P < 1.0E-07 and P < 7.14E-03 for replication.
[a] Discovery panel (n=6820): Cohorts KORA-F4, SHIP-Trend, ESTHER, LOLIPOP.
[b] Replication panel (n=1450): the RS (sub-cohorts RSIII-1 and RS-Bios)
Abbreviations: Chr, chromosome; Loc, location; SE, standard error; Het, heterogeneity.
I^2 corresponds to the heterogeneity test.
Bold text indicates statistically significant associations.

Table 2: Association between DNA methylation of the replicated CpGs associated with statin use, and glycemic traits and type 2 diabetes

			log Glucose [a]			log Insulin [a]			log HOMA-IR [a]			Incident type 2 diabetes [b]			
CpG	Position	Gene	Effect	SE	P value	Effect	SE	P value	Effect	SE	P value	Effect	SE	P value	OR (95% CI)
cg17901584	55353706	*DHCR24*	-0.006	0.004	1.63E-1	-0.023	0.019	2.35E-1	-0.029	0.021	1.62E-1	-0.175	0.064	6.2E-3	0.84 (0.74, 0.95)
cg06500161	43656587	*ABCG1*	**0.010**	**0.003**	**1.03E-3**	**0.076**	**0.014**	**4.63E-8**	**0.086**	**0.015**	**1.05E-8**	**0.293**	**0.048**	**9.5E-10**	**1.34 (1.22, 1.47)**
cg05119988	166251189	*SC4MOL*	-0.003	0.003	3.59E-1	-0.004	0.015	7.84E-1	-0.007	0.016	6.65E-1	-0.058	0.055	2.9E-1	0.94 (0.84, 1.05)
cg27243685	43642366	*ABCG1*	-0.002	0.003	5.04E-1	0.041	0.014	4.62E-3	0.039	0.016	1.39E-2	**0.218**	**0.051**	**1.99E-5**	**1.24 (1.13, 1.38)**
cg10177197	55316481	*DHCR24*	-0.001	0.003	8.01E-1	-0.015	0.015	3.08E-1	-0.016	0.016	3.18E-1	0.080	0.052	1.2E-1	1.08 (0.98, 1.20)

Model adjusted for leukocyte proportions, batch effects, sex, age, smoking status, systolic blood pressure, anti-hypertensive medication, presence of coronary heart disease, statin use and body mass index.

Bold text indicates statistically significant associations after Bonferroni correction of P < 2.5E-03.

[a] Sample n=1,165 (RS) complete cases, of which 119 were current statin users. Non-fasting samples (n=25), prevalent type 2 diabetes cases (n=181) and former statin users (n=74) were excluded from the analysis.

[b] Sample size n=4,422 (RS n=626; KORA n=1,134; LOLIPOP n=2,659), of which 1,173 are cases of incident type 2 diabetes.

Abbreviations: Chr, chromosome; SE, standard error; HOMA-IR, homeostatic model assessment insulin resistance

0.270) are suggested to act via cg06500161 methylation, respectively. According to these results, cg06500161 methylation does not explain 100% of the effect of statins on insulin/HOMA-IR, thus we may call cg06500161 methylation a partial mediator, instead of a total mediator.

No significant findings were observed for the mediation model with outcome glucose (Table 3). We also investigated the potential mediation of cg27243685 on incident type 2 diabetes, with no significant findings. The sensitivity analysis showed that residual correlation due to unmeasured confounding of r≥0.2 would be needed to violate the sequential ignorability assumption.

Table 3: Causal mediation analysis on the significant associations between the statin-related CpGs (cg06500161 and cg27243685) as mediators and glycemic traits and incident type 2 diabetes as outcomes

	ACME estimate of mediator CpG (95% CI)	ADE estimate (95% CI)	Total Effect (95% CI)	Proportion mediated by cg06500161 (95% CI)
cg06500161				
log Glucose [a]	0.006 (0.003, 0.010)	0.009 (-0.007, 0.020)	0.015 (-6.79E-03, 0.030)	0.404 (-1.250, 2.600)
log Insulin [a]	0.043 (0.024, 0.060)	0.233 (0.149, 0.320)	0.275 (0.190, 0.360)	0.155 (0.086, 0.260)
log HOMA-IR [a]	0.049 (0.028, 0.070)	0.242 (0.151, 0.330)	0.291 (0.199, 0.380)	0.168 (0.095, 0.270)
Incident type 2 diabetes[b]	0.017 (-0.002, 0.040)	0.094 (0.024, 0.190)	0.111 (0.033, 0.220)	0.153 (-0.029, 0.390)
cg27243685				
Incident type 2 diabetes[b]	0.0005 (-0.006, 0.010)	0.127 (0.033, 0.23)	0.128 (0.035, 0.23)	0.004 (-0.053, 0.10)

Models adjusted for sex, age and cohort.
Bold text indicates statistically significant results.
[a] Sample n=1,180 (RS) complete cases, of which 178 were current statin users. Non-fasting samples (n=25), prevalent type 2 diabetes cases (n=181) and former statin users (n=74) were excluded from the analysis.
[b] Sample n=642 (RS) complete cases, of which 23 were cases of incident type 2 diabetes type 2 diabetes.
Abbreviations: CI, confidence interval; ACME, average causal mediator effect; ADE, average direct effect; HOMA-IR, homeostatic model assessment insulin resistance.

Further mediation models were performed to account for potential confounders. Additional to age, sex and cohort, we further adjusted for 1) BMI; 2) BMI and HDL; 3) smoking status, systolic blood pressure, anti-hypertensive medication and presence of coronary heart disease. Results did not materially change with these further adjustments, although the effect sizes attenuated (eTables 5-7).

Gene expression

To explore potential epigenetic regulations of gene expression by DNAm, the identified CpGs (predictors) were tested in association with transcriptome-wide gene expression in the RS. Cg06500161 was inversely associated with two expression probes at *ABCG1* (ILMN_1794782 and ILMN_2329927), after Bonferroni correction (4.71×10^{-7}). An inverse association was observed for cg27243685 and one probe (ILMN_1794782). These results indicate a lower expression of *ABCG1* with increasing cg06500161 and cg27243685 methylation. Furthermore, cg17901584 methylation was directly associated with *ABCG1* expression (ILMN_1794782) and suggestively associated with another probe at the same gene (ILMN_2329927). There were no significant findings for the other two CpGs after Bonferroni correction (eTable 8). Moreover, exposure to statin treatment was associated with lower *ABCG1* expression (eTable 9). When *ABCG1* expression was studied as predictor of insulin traits, we found an inverse association between ILMN_1794782 and fasting insulin and HOMA-IR (eTable 10).

Dose effect

Statin dose exposure was significantly associated with the five identified CpGs after Bonferroni-correction ($P<0.01$) (eTable 11). Exposure to increasing statin dose was nominally associated with lower levels of *ABCG1* expression at ILMN_1794782 and suggestively with ILMN_2329927 (eTable 12).

Sensitivity analyses

Residual confounding and confounding by indication: i) Additional adjustment of the EWAS for serum total-cholesterol or blood lipids HDL, LDL and triglycerides instead (eTable 13); ii) exclusion of prevalent type 2 diabetes and pre-type 2 diabetes cases (eTable 14); or iii) restriction of the EWAS to participants with LDL ≥70 mg/dL or ≥100 mg/dL (eTable 15) did not change the associations of the five CpGs. Further, exclusion of type 2 diabetes cases from the other analyses did not affect the results (data not shown).

Trans-ethnic replication: the new EWAS taking Caucasians only for discovery and South-Asians for replication showed that all five CpGs passed Bonferroni-correction threshold (eTable 16).

DISCUSSION

The current study sheds light on potential mechanisms linking statin use and type 2 risk diabetes, by firstly, identifying and replicating associations between statin therapy and methylation at five CpGs (cg06500161, cg27243685, cg17901584, cg10177197

and cg05119988) independent of blood lipids; and secondly, providing evidence on the partial mediaton via *ABCG1* methylation in the effect of statins on increased levels of insulin and insulin resistance under the sequential ignorability assumption.

Our results on the association between statin use and DNAm agree with a report from the Framingham Study who found epigenome-wide significance at only two sites (cg17901584 and cg06500161).[5] They examined this association in a smaller sample (*n*=1,545 versus 6,820 of our discovery panel), and contrary to our investigation, they did not proceed with replication nor examined associations with gene expression, metabolic markers and type 2 diabetes. Moreover, our work reports novel associations and adds to current knowledge that statin dosage might be implicated on the degree of methylation and *ABCG1* expression. The latter finding goes in line with an experimental study where macrophages treated with various types of statins showed lower *ABCG1* expression as dose increased.[22]

Sites cg06500161 and cg27243685 are annotated to *ABCG1*, cg17901584 and cg10177197 to *DHCR24*, and cg05119988 to of *SC4MOL*. *ABCG1* (ATP-Binding Cassette Member-1 Subfamily-G) encodes a protein that mediates the transport of different molecules, such as cholesterol efflux to the high-density lipoprotein, oxysterols and phospholipid transport in macrophages.[23,24] It is also involved in insulin secretion and sensitivity.[25] *ABCG1* expression was reduced in statin users compared to non-users,[26] and in type 2 diabetes patients.[27,28] Genes *DHCR24* (24-Dehydrocholesterol Reductase) and *SC4MOL* (Sterol-C4-methyl oxidase-like) code enzymes catalyzing different steps during cholesterol biosynthesis. *DHCR24* mutations are related to desmosterolosis[29,30] and Alzheimer´s disease.[31] *SC4MOL* deficiency protein produces congenital cataracts, microcephaly, growth delay, skin conditions and immune dysfunction.[32] An observational study found that *DHCR24* and *SC4MOL* were up-regulated among statin users,[26] while *SC4MOL* up-regulation increased type 2 diabetes risk.[27,28]

Methylation at four of the identified CpGs (cg06500161, cg27243685, cg17901584 and cg05119988) were previously associated with glycemic traits, type 2 diabetes and blood lipids while cg06500161 (*ABCG1*) was associated with increased glucose, insulin, HbA1c, HDL and triglycerides levels.[33-35] Our causal mediation analyses provided evidence on cg06500161 methylation to partially mediate the association between statin use and higher fasting insulin and HOMA-IR. Although, given the crossectional nature of these analyses our results must be interpreted with caution. Since the mediator was measured at the same time as the outcome, reverse causation can not be completely ruled out. However, in an effort to overcome this issue, we excluded type 2 diabetes cases. The mediation effect observed may be the result

of more complex interactions of combined effects across the epigenome. Moreover, the lack of significant findings in mediation by the other CpG at *ABCG1* (cg27243685) could reflect a power issue. Hence, future studies shall explore how the CpGs within the same gene interact and mediate the effects.

Based on our findings and the available evidence, we hypothesize that cg06500161 hypermethylation may be a consequence of statins use, possibly inducing a decrease on *ABCG1* transcription in blood. Further research is needed to investigate to what extent this down-regulation could in turn compromise downstream signals, resulting in impaired insulin metabolism. In this line, imapaired insulin sensitivity and secretion as a consequence of statin treatment was recently observed in a longitudinal study.[36] Furthermore, a functional study suggested an epigenetic regulation of *ABCG1* mediated by methylation-dependent transcription factor binding.[37] Our hypothesized model is displayed in Figure 2. Nevertheless, further investigations should test this hypothesis and assess the effect of statins in target tissues, and to what extent these findings are generalizable to populations with ethnic backgrounds other than Caucasians and South-Asians.

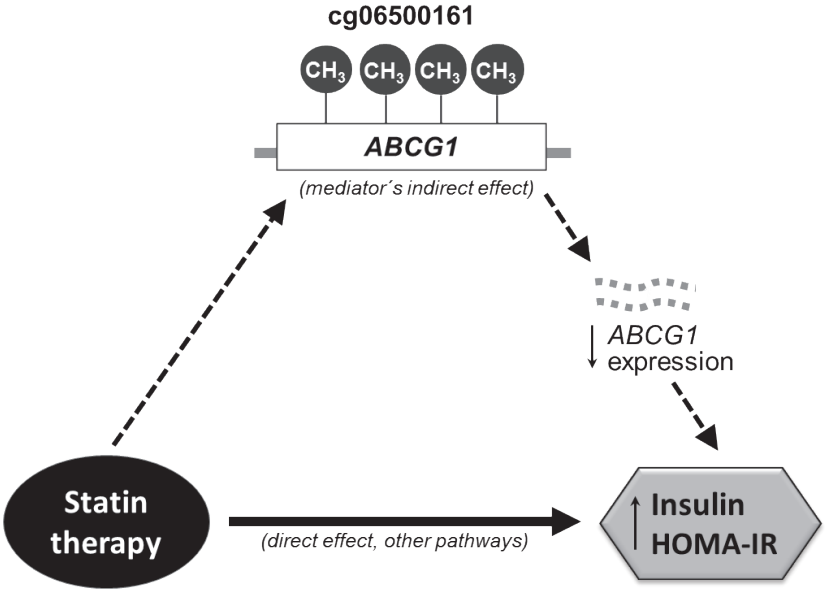


Figure 2. Scheme of the hypothesized mechanism linking statin therapy and risk of type 2 diabetes The solid black arrow represents the effect of statins on plasmatic insulin and HOMA-IR levels that goes directly or through a pathway different from the mediator analyzed in the current study (methylation at cg06500161). The dotted black arrows represent the suggested alternative pathway, where an indirect effect of statins on insulin and HOMA-IR is mediated by cg06500161 methylation. The black thin arrows indicate decreased or increased level of the parameter; in this figure they are representing decrease in gene expression of *ABCG1* and increase in serum levels of fasting insulin and HOMA-IR.

We highlight the large sample size with DNAm data in our study, the use of a replication panel and the comprehensive assessment of confounding factors. Addition of complementary data like gene expression, trans-ethnic replication and use of a causal inference method as mediation analysis, also add value to our research.

One limitation is the lack of data on DNAm and gene expression from specific tissues of interest for type 2 diabetes and drug metabolism (pancreas, liver, adipose tissue) as both DNAm and gene expression may be tissue specific.[38] Instead, we used whole peripheral blood, allowing for cell mixture. Consequently, we included adjustments for estimated blood leukocytes proportions.

Most cohorts had limited information on former statin users, thus it is likely that they were included in the reference group. However, the replication panel showed the same results are valid for the group of current statin users only, although we had limited numbers to investigate such associations. Furthermore, it was not possible to explore potential effects of cessation time or treatment duration on DNAm. A Mendelian randomization (MR) approach based on genetic instrumental variables to explore a potential causal effect of statins on DNAm was not possible due to the lack of adequate genetic variants that may mimic the randomization of the hypothesized adverse effect of statins on type 2 diabetes risk. The use of variants at the gene encoding 3hydroxy-3-methylglutaryl coenzyme-A reductase,[39] statins' target protein, would mimic the main effect of statins, rather than an alternative pathway leading to the hypothesized effect. Moreover, these variants were associated with LDL,[39] reflecting cholesterol metabolism. This violates the MR assumption that the instrumental variable is associated with the outcome only through the exposure, since high cholesterol is a risk factor for type 2 diabetes.

All cohorts used Illumina 450K array to measure DNAm, except one using Illumina EPIC array (n=247). Despite differences, EPIC array covers approximately 90% of the same sites on the 450K chip, both use the same technology, and showed overall per-sample correlations of r>0.99,[40] thus encouraging the application of combined data. Moreover, cohorts used different methods to diagnose type 2 diabetes, which might introduce error measurement.

The heterogeneity observed in the meta-analysis for some of the CpGs might be consequence of differences in statin use prevalence (ranged from 9.2% to 18.1%), types and dose of statin used across cohorts. Additionally, differences in ethnicity could have influenced heterogeneity (the association statin use and cg06500161 showed I^2=52.30 in the overall meta-analysis, while I^2=5.20 in Caucasians); as well as

different methods applied to collect data on medication and disease (pharmacy and general practitioner records or self-report).

Finally, some analyses were performed using data from the replication panel only, which may introduce type-II error.

CONCLUSION

We identified and replicated five DNAm sites associated with statin therapy, of which cg06500161 may be partially mediating the effect of statins on increasing insulin levels and HOMA-IR. This could be one potential mechanism linking statin therapy and type 2 diabetes onset observed in clinical trials and observational studies. Nevertheless, other biological pathways should not be discarded. From our study, it is not clear what might be the mechanism by which statins may induce DNAm changes; this shall be a challenge for future research.

Meanwhile, close monitoring of risk factors for type 2 diabetes in these patients is imperative to prevent further complications.

SUPPLEMENTARY ONLINE CONTENT

Epigenetic Link Between Statin Therapy and Type 2 Diabetes

eTable 1 DNA methylation assessment methods of each cohort

	RS-III-1	RS-BIOS	SHIP-Trend	ESTHER	KORA	LOLIPOP
Methylation assay	Illumina 450K array	Illumina 450K array	Illumina EPIC array	Illumina 450K array	Illumina 450K array	Illumina 450K array
IDAT extraction	Custom script	Beadstudio	Custom script	GenomeStudio	Custom script	minfi
Background correction	separate colors	separate colors	separate colors	separate colors	separate colors	minfi
Detection P value cutoff	0.01	NA	1.00E-16	0.01	0.01	0.01
Sample call rate threshold	99%	99%	95%	95%	95%	NA
Nbeads filter	3	NA	NA	NA	3	3
Normalization	SWAN	DASEN	CPACOR	Illumina	CPACOR	CPACOR

eTable 2-a. Baseline characteristics of the cohorts included in the discovery panel

	SHIP-Trend			ESTHER			KORA F4			LOLIPOP		
Country	Germany			Germany			Germany			United Kingdom		
Ethnicity	Caucasian			Caucasian			Caucasian			South Asian		
Study Design	Population based			Population-based			Population-based			Prospective type 2 diabetes Case/Control		
	Non-current users	Current users	Total	Non-current users	Current users	Total	Non-current users	Current users	Total	Non-current users	Current users	Total
Sample size	207	40	247	906	92	998	1455	271	1726	3390	459	3849
Age, years, mean (S.D.)	48.8 (13.1)	65.2 (8.3)	51.5 (13.8)	62.0 (6.5)	63.6 (6.6)	62.1 (6.5)	60.02 (8.84)	66.06 (7.21)	60.97 (8.88)	51.3 (10.1)	57.8 (9.4)	52.0 (10.3)
Sex female, n (%)	116 (56.0)	14 (35.0)	130 (52.6)	462 (51.0)	36 (39.1)	498 (49.9)	767 (52.7)	115 (42.4)	882 (51.1)	1325 (39.1)	134 (29.2)	1459 (37.9)
Body mass index, kg/m2,mean (SD)	27 (4.2)	28.7 (2.9)	27.2 (4.0)	27.8 (4.3)	27.2 (3.6)	27.8 (4.3)	27.88 (4.74)	29.35 (4.78)	28.11 (4.78)	27.4 (4.4)	28.0 (4.1)	27.5 (4.4)
Systolic blood pressure, mmHg, mean (SD)	123.6 (16.7)	128.4 (19.3)	124.3 (17.2)	139.5 (19.9)	141.5 (16.8)	139.7 (19.7)	124.79 (18.87)	124.82 (17.98)	124.8 (18.73)	131.7 (19.0)	138.6 (20.0)	132.5 (19.3)
Antihypertensive medication use, n (%) [missing, n (%)]	56 (27.1)	26 (65.0)	82 (33.2)	406 (44.8)	71 (77.2)	477 (47.8)	445 (30.6)	204 (75.3)	649 (37.6)	524 (21.7) [978, (28.8)]	307 (66.9) [0, (0.0)]	831 (28.9) [978, (25.4)]
Coronary heart disease, n (%) [a]	36 (17.4)	29 (72.5)	65 (26.3)	107 (11.8)	37 (40.2)	144 (14.4)	100 (6.9)	83 (30.5)	183 (10.6)	58 (1.7)	166 (36.2)	223 (5.8)
Prevalent type 2 diabetes, n (%) [b] [missing, n (%)]	0	0	0	136 (15.0)	20 (20.7)	156 (15.6)	99 (6.8)	62 (22.9)	161 (9.3)	206 (9.1) [1123, (33.1)]	56 (14.3) [67, (14.6)]	262 (6.8) [1190, (30.9)]
Incident type 2 diabetes, n (%) [b] [missing, n (%)]	NA	NA	NA	161 (17.8)	13 (13.7)	174 (17.4)	83 (5.7)	41 (15.1)	117 (6.8)	802 (35.4) [1123, (33.1)]	270 (68.9) [67, (14.6)]	1072 (40.3) [1190, (30.9)]
Fasting glucose, mmol/L, median (I.Q.R.)	NA	NA	NA	5.62 (1.28)	5.73 (0.83)	5.63 (1.22)	5.33 (0.84)	5.61 (1.19)	5.33 (0.89)	5.1 (0.8)	5.5 (1.0)	5.2 (0.9)

eTable 2-a. Baseline characteristics of the cohorts included in the discovery panel *(continued)*

	SHIP-Trend			ESTHER			KORA F4			LOLIPOP		
Country	Germany			Germany			Germany			United Kingdom		
Ethnicity	Caucasian			Caucasian			Caucasian			South Asian		
Study Design	Population based			Population-based			Population-based			Prospective type 2 diabetes Case/Control		
	Non-current users	Current users	Total	Non-current users	Current users	Total	Non-current users	Current users	Total	Non-current users	Current users	Total
Fasting insulin, uU/mL, median (I.Q.R.)	NA	NA	NA	NA	NA	NA	8.9 (6.5)	11 (8.9)	9.3 (7.3)	10.4 (9.0)	11.8 (9.9)	10.6 (9.3)
LDL cholesterol, mmol/L, mean (S.D.)	3.50 (0.90)	2.80 (0.70)	3.40 (0.90)	3.82 (0.97)	3.50 (1.14)	3.78 (1.00)	3.70 (1.17)	3.05 (0.93)	3.57 (1.24)	3.32 (0.88)	2.76 (0.91)	3.24 (0.90)
HDL cholesterol, mmol/L, median (I.Q.R.)	1.36 (0.41)	1.37 (0.48)	1.36 (0.42)	1.32 (0.60)	1.39 (0.57)	1.33 (0.57)	4.42 (0.52)	1.32 (0.43)	1.42 (0.49)	1.31 (0.39)	1.25 (0.30)	1.26 (0.39)
Total cholesterol, mmol/L, mean (S.D.)	5.50 (1.10)	4.80 (0.90)	5.40 (1.10)	4.26 (1.41)	4.33 (1.58)	4.27 (1.42)	5.76 (1.32)	5.12 (1.16)	5.68 (1.34)	5.36 (1.00)	4.83 (1.10)	5.29 (1.03)
Triglycerides, mmol/L, median (I.Q.R.)	1.30 (0.95)	1.52 (0.80)	1.35 (0.93)	1.24 (0.91)	1.41 (0.91)	1.26 (0.91)	1.23 (0.91)	1.49 (1.09)	1.26 (0.93)	1.40 (0.97)	1.62 (0.98)	1.40 (0.99)
Never smoker %	85 (41.0)	14 (42.5)	102 (41.2)	440 (48.6)	39 (42.4)	479 (48.0)	604 (41.5)	118 (43.5)	722 (41.8)	2875 (84.8)	358 (78.0)	3233 (84.0)
Former smoker %	78 (37.7)	15 (37.5)	93 (37.7)	295 (32.5)	40 (43.5)	335 (33.5)	630 (43.3)	124 (45.8)	754 (43.7)	251 (7.4)	61 (13.3)	312 (8.1)
Current smoker %	44 (21.3)	8 (20.0)	52 (21.1)	171 (18.9)	13 (14.1)	184 (18.5)	221 (15.2)	29 (10.7)	250 (14.5)	264 (7.8)	40 (8.7)	304 (7.9)

[a] Coronary heart disease: Revascularization by PCI or CABG; Angiographically severe coronary disease; Documented acute coronary syndrome (ACS; symptoms, ECG change and/or biochemistry).

[b] Type 2 Diabetes: physician diagnosis, fasting glucose ≥ 7.0 mmol/L or HbA1c>6.5%, or self-report

eTable 2-b: Baseline characteristics of the cohorts participating in the replication panel and in post-EWAS analyses

	Replication panel: The Rotterdam Study (RS)					
	RS-III-1			**RS-BIOS (RS-II-3 and RS-III-2)**		
Country	the Netherlands			the Netherlands		
Ethnicity	Caucasian			Caucasian		
Study Design	Population based			Population based		
	Non-current users	Current users	Total	Non-current users	Current users	Total
Sample size	641	90	731	546	173	719
Age, years, mean (S.D.)	59.1 (7.9)	65.1 (8.5)	59.9 (8.2)	67.14 (6.2)	69.1 (0.4)	67.6 (6.0)
Sex female, n (%)	355 (55.4)	40 (44.4)	395 (54.0)	326 (59.7)	92 (53.2)	414 (57.6)
Body mass index, kg/m2,mean (SD)	27.3 (4.8)	29.0 (4.5)	27.52 (4.8)	27.57 (4.0)	28.1 (4.2)	27.7 (4.1)
Systolic blood pressure, mmHg, mean (SD)	134.0 (19.9)	137.4 (18.4)	134.4 (4.5)	144.7 (22.1)	144.5 (21.3)	144.7 (21.9)
Antihypertensive medication, n (%)	147 (22.9)	70 (77.8)	217 (29.7)	187 (34.2)	123 (71.1)	310 (43.1)
Coronary heart disease, n (%) [a]	19 (3.0)	26 (29.0)	45 (6.2)	15 (2.7)	44 (25.4)	59 (8.2)
Prevalent type 2 diabetes, n (%) [b]	45 (7.0)	29 (32.2)	74 (10.1)	55 (10.1)	52 (30.1)	107 (14.9)
Incident type 2 diabetes, n (%) [b]	16 (2.5)	10 (11.1)	26 (3.6)	NA	NA	NA
Fasting glucose, mmol/L, median (I.Q.R.)	5.3 (0.8)	5.6 (1.7)	5.4 (0.7)	5.30 (0.80)	5.65 (1.22)	5.40 (0.90)
Fasting insulin, uU/mL, median (I.Q.R.)	10.8 (7.8)	15.6 (14.3)	11.2 (7.8)	9.50 (7.24)	12.31 (8.86)	10.08 (7.78)
LDL cholesterol, mmol/L, mean (S.D.)	3.61 (0.93)	2.51 (0.69)	3.48 (0.98)	3.58 (0.85)	2.57 (0.70)	3.33 (0.92)
HDL cholesterol, mmol/L, median (I.Q.R.)	1.37 (0.52)	1.17 (0.48)	1.37 (0.52)	1.49 (0.62)	1.37 (0.47)	1.46 (0.59)
Total cholesterol, mmol/L, mean (S.D.)	5.70 (1.04)	4.56 (0.76)	5.56 (1.04)	5.77 (0.87)	4.68 (0.92)	5.51 (1.02)
Triglycerides, mmol/L, median (I.Q.R.)	1.28 (0.83)	1.59 (0.97)	1.28 (0.81)	1.26 (0.72)	1.41 (0.83)	1.30 (0.76)
Never smoker, n (%)	192 (30.0)	21 (23.3)	213 (29.1)	192 (35.2)	56 (32.4)	248 (34.5)
Former smoker, n (%)	275 (42.9)	46 (51.1)	321 (43.9)	297 (54.4)	97 (56.1)	394 (54.8)
Current smoker, n (%)	174 (27.1)	23 (25.6)	197 (26.9)	57 (10.4)	20 (11.6)	77 (10.7)

[a] Coronary heart disease: Revascularization by PCI or CABG; Angiographically severe coronary disease; Documented acute coronary syndrome (ACS; symptoms, ECG change and/or biochemistry).

[b] Type 2 Diabetes: physician diagnosis, fasting glucose ≥ 7.0 mmol/L or HbA1c>6.5%, or self-report

eTable 3. Pearson correlation among the replicated CpGs in two the sub-cohorts of the replication panel (RS)

RS-III-1	cg06500161	cg27243685	cg17901584	cg10177197	cg05119988
cg06500161	1				
cg27243685	0.244	1			
cg17901584	0.176	0.289	1		
cg10177197	0.065	0.353	0.174	1	
cg05119988	-0.173	-0.092	-0.187	-0.107	1
RS-BIOS	cg06500161	cg27243685	cg17901584	cg10177197	cg05119988
cg06500161	1				
cg27243685	0.477	1			
cg17901584	-0.011	0.270	1		
cg10177197	0.090	0.303	0.143	1	
cg05119988	-0.201	-0.420	-0.222	-0.371	1

eTable 4 Replication of DNA methylation sites associated with current statin use in a subset of the replication sample, excluding former statin users

CpG	Chr	Position	location	Gene	Effect	SE	P-value
cg17901584	1	55353706	TSS1500	*DHCR24*	**-0.0179**	**0.0037**	**1.34E-06**
cg06500161	21	43656587	gene body	*ABCG1*	**0.0123**	**0.0027**	**3.63E-06**
cg05119988	4	166251189	5'UTR	*SC4MOL*	**-0.0121**	**0.0039**	**1.94 E-03**
cg27243685	21	43642366	gene body	*ABCG1*	**0.0072**	**0.0018**	**5.06E-05**
cg10177197	1	55316481	gene body	*DHCR24*	**-0.0045**	**0.0010**	**1.35E-05**

Model adjusted for leukocyte proportions, batch effects, sex, age, smoking status, body mass index, systolic blood pressure, anti-hypertensive medication, presence of coronary heart diseases, prevalent type 2 diabetes.
Bonferroni correction for significance for replication P < 7.14E-03.
Sample size n=1205 (RS)
Abbreviations: Chr, chromosome; SE, standard error. Bold text indicates statistically significant associations.

eTable 5: Causal mediation analysis on the significant associations between the statin-related CpG cg06500161 and glycemic traits and incident type 2 diabetes, with a model adjusted for sex, age, cohort and body mass index

Outcomes	ACME estimate of mediator cg06500161 (95% CI)	ADE estimate (95% CI)	Total Effect (95% CI)	Proportion mediated by cg06500161 (95% CI)
log Glucose [a]	0.004 (0.001, 0.010)	0.006 (-0.009, 0.020)	0.010 (-0.005, 0.030)	0.404 (-3.201, 3.920)
log Insulin [a]	**0.025 (0.011, 0.04)**	**0.205 (0.131, 0.280)**	**0.229 (0 .1562, 0.30)**	**0.106 (0.047, 0.190)**
log HOMA-IR [a]	**0.029 (0.014, 0.050)**	**0.212 (0.132, 0.290)**	**0.240 (0.160, 0.32)**	**0.119 (0.057, 0.210)**
T2D [b]	0.007 (-0.002, 0.020)	0.116 (0.024, 0.22)	0.12 (0.030, 0.230)	0.060 (-0.024, 0.230)

Models adjusted for sex, age, cohort and body mass index. Bold text indicates statistically significant results.
[a] Sample n=1,180 (RS) complete cases, of which 178 were current statin users. Non-fasting samples (n=25), prevalent type 2 diabetes cases (n=181) and former statin users (n=74) were excluded from the analysis.
[b] Sample n=640 (RS) complete cases, of which 23 were cases of incident type 2 diabetes.
Abbreviations: CI, confidence interval; ACME, average causal mediator effect; ADE, average direct effect; HOMA-IR, homeostatic model assessment insulin resistance; T2D, type 2 diabetes.

eTable 6: Causal mediation analysis on the significant associations between the statin-related CpG cg06500161 and glycemic traits and incident type 2 diabetes, with a model adjusted for sex, age, cohort, body mass index and high density lipoprotein cholesterol

Outcomes	ACME estimate of mediator cg06500161 (95% CI)	ADE estimate (95% CI)	Total Effect (95% CI)	Proportion mediated by cg06500161 (95% CI)
log Glucose [a]	0.003 (0.001, 0.010)	0.005 (-0.010, 0.020)	0.008 (-0.007, 0.020)	0.396 (-3.231, 3.63)
log Insulin [a]	**0.014 (0.004, 0.030)**	**0.184 (0.111, 0.260)**	**0.198 (0.125, 0.27)**	**0.071 (0.017, 0.15)**
log HOMA-IR [a]	**0.018 (0.006, 0.030)**	**0.189 (0.114, 0.260)**	**0.20 (0.131, 0.280)**	**0.085 (0.028, 0.17)**
T2D [b]	0.006 (-0.003, 0.020)	0.110 (0.019, 0.210)	0.116 (0.024, 0.220)	0.051 (-0.039, 0.230)

Models adjusted for sex, age, cohort, body mass index and high density lipoprotein cholesterol. Bold text indicates statistically significant results.
[a] Sample n=1,180 (RS) complete cases, of which 178 were current statin users. Non-fasting samples (n=25), prevalent type 2 diabetes cases (n=181) and former statin users (n=74) were excluded from the analysis.
[b] Sample n=640 (RS) complete cases, of which 23 were cases of incident type 2 diabetes.
Abbreviations: CI, confidence interval; ACME, average causal mediator effect; ADE, average direct effect; HOMA-IR, homeostatic model assessment insulin resistance; T2D, type 2 diabetes.

eTable 7: Causal mediation analysis on the significant associations between the statin-related CpG (cg06500161) and glycemic traits and incident type 2 diabetes, with a comprehensive assessment of confounder factors

Outcomes	ACME estimate of mediator cg06500161 (95% CI)	ADE estimate (95% CI)	Total Effect (95% CI)	Proportion mediated by cg06500161 (95% CI)
log Glucose [a]	0.004 (0.001, 0.010)	-0.003 (-0.019, 0.010)	0.002 (-0.015, 0.020)	2.382 (-6.352, 6.180)
log Insulin [a]	**0.024 (0.011, 0.040)**	**0.151 (0.069, 0.230)**	**0.174 (0.093, 0.260)**	**0.136 (0.060, 0.290)**
log HOMA-IR [a]	**0.028 (0.013, 0.050)**	**0.149 (0.061, 0.240)**	**0.176 (0.088, 0.270)**	**0.156 (0.069, 0.340)**
T2D [b]	0.004 (-0.004, 0.010)	0.090 (-0.004, 0.190)	0.094 (-0.001, 0.200)	0.047 (-0.123, 0.320)

Models adjusted for sex, age, cohort, body mass index, smoking status, systolic blood pressure, antihypertensive medication and presence of coronary heart disease. Bold text indicates statistically significant results.
a Sample n=1,160 (RS) complete cases, of which 178 were current statin users. Non-fasting samples (n=25), prevalent type 2 diabetes cases (n=181) and former statin users (n=74) were excluded from the analysis.
[b] Sample n=622 (RS) complete cases, of which 23 were cases of incident type 2 diabetes.
Abbreviations: CI, confidence interval; ACME, average causal mediator effect; ADE, average direct effect; HOMA-IR, homeostatic model assessment insulin resistance; T2D, type 2 diabetes.

eTable 8: Significant results of the gene expression-DNA methylation association analyses between the statin-related CpGs and expression probes

CpG	Chr	Position	Location	Gene at CpG	Illumina Probe ID	Gene at expression probe	Effect	P-value
cg06500161	21	43656587	gene body	***ABCG1***	**ILMN_1794782**	*ABCG1*	-13.000	**8.85E-13**
					ILMN_2329927	*ABCG1*	-8.504	**4.85E-12**
cg27243685	1	43642366	gene body	***ABCG1***	**ILMN_1794782**	*ABCG1*	-15.974	**1.31E-07**
cg17901584	1	55353706	TSS1500	***DHCR24***	**ILMN_1794782**	*ABCG1*	9.319	**2.21E-10**
					ILMN_2329927	*ABCG1*	4.975	6.15E-07

Sample size n=731, from cohort RS (RSIII-1). Bonferroni correction p < 4.7E-07. Bold text indicates significant associations.

eTable 9: Association between the CpGs-associated expression probes for *ABCG1* and current statin use as exposure

ILMN_1794782 (*ABCG1*)			ILMN_2329927 (*ABCG1*)		
Effect	SE	P-value	Effect	SE	P-value
-0.7999	**0.1648**	**1.54E-06**	**-0.3750**	**0.1186**	**0.0016**

Model: gene-expression ~ statin use, adjusted for age, sex, body mass index, coronary heart disease, systolic blood pressure, high-density lipoprotein cholesterol; n=631 (RS).
Cases of prevalent type 2 diabetes and former users of statins were excluded. Bold text indicates significant associations.

eTable 10: Association between significant outcomes found in the causal mediation analyses (fasting insulin and HOMA-IR) and the CpGs-associated expression probes for gene *ABCG1*

		log Insulin [b]			log HOMA-IR [b]		
Illumina Probe	Gene at probe	Effect	SE	P-value	Effect	SE	P-value
ILMN_1794782	***ABCG1***	**-0.0290**	**0.0137**	**0.0355**	**-3.130E-02**	**1.5E-2**	**0.0354**
ILMN_2329927	*ABCG1*	-0.0252	0.019774	0.2038	-2.921E-02	2.1E-2	0.1718

Model: glycemic trait (insulin or HOMA-IR) ~ statin use, adjusted for age, sex, body mass index, coronary heart disease, systolic blood pressure, high-density lipoprotein cholesterol; n=616 (RS).
Cases of prevalent type 2 diabetes and former users of statins were excluded.
Bold text indicates significant associations.

eTable 11: Association between DNA methylation at the identified statin-related CpGs and statin dose among current users of statins

CpG	Chr	Position	location	Gene	Effect	SE	P-value
cg17901584	1	55353706	TSS1500	*DHCR24*	-0.0058	0.0019	**2.104E-03**
cg06500161	21	43656587	gene body	*ABCG1*	0.0037	0.0014	**9.579E-03**
cg05119988	4	166251189	5'UTR	*SC4MOL*	-0.0068	0.0021	**1.317E-03**
cg27243685	21	43642366	gene body	*ABCG1*	0.0025	0.0009	**5.012E-03**
cg10177197	1	55316481	gene body	*DHCR24*	0.0025	0.0013	**9.510E-03**

Model adjusted for leukocyte proportions, batch effects, sex, age, smoking status, body mass index, systolic blood pressure, anti-hypertensive medication, presence of coronary heart diseases and prevalent type 2 diabetes.
Bonferroni correction for significance of $P < 0.05/5$ replicated CpGs = 0.01.
Sample size n=303 current statin users with dose data, from cohorts SHIP-Trend and RS.
Abbreviations: Chr, chromosome; SE, standard error.

eTable 12: Association between gene expression and statin use

Statin use (exposure)	Illumina Probe	Gene at probe	Effect	SE	P-value
Current use	**ILMN_1794782**	***ABCG1***	**-0.0490**	**0.0177**	**5.92 E-03**
	ILMN_2329927	*ABCG1*	-0.0496	0.0256	0.0529

Models adjusted for age, sex, body mass index, coronary heart disease, systolic blood pressure, high-density lipoprotein cholesterol.
Cases of prevalent type 2 diabetes and former users of statins were excluded.
Sample size n=631 (RS)
Bold text indicates significant associations.

eTable 13: DNA methylation sites associated with current statin use (compared to non-current use) in the discovery panel and replication panel (EWAS) in models further adjusted for blood lipids

Model additionally adjusted for serum total cholesterol [a]					Discovery panel [b]			Replication panel [c]			Both panels combined		
CpG	Chr	Position	location	Gene	Effect	SE	P-value	Effect	SE	P-value	Effect	SE	P-value
cg17901584	**1**	55353706	TSS1500	***DHCR24***	-0.0204	0.0021	3.69E-22	-0.0108	0.0034	1.66E-03	**-0.0178**	**0.0018**	**4.18E-23**
cg06500161	**21**	43656587	gene body	***ABCG1***	0.01	0.001	1.05E-21	0.0102	0.0024	1.73E-05	**0.01**	**0.001**	**9.44E-26**
cg05119988	**4**	166251189	5'UTR	***SC4MOL***	-0.0111	0.0018	2.63E-10	-0.0107	0.0034	1.61E-03	**-0.011**	**0.0016**	**1.65E-12**
cg05575921	**5**	373378	gene body	*AHRR*	-0.0136	0.0022	1.06E-09	-0.0009	0.0059	8.79E-01	**-0.012**	**0.0021**	**8.49E-09**
cg27243685	**21**	43642366	gene body	***ABCG1***	0.0045	0.0008	7.65E-09	0.0065	0.0016	6.32E-05	**0.0048**	**0.0007**	**4.01E-12**
cg10177197	**1**	55316481	gene body	***DHCR24***	0.0036	0.0007	3.90E-07	0.0042	0.0017	1.41E-02	**0.0037**	**0.0007**	**1.82E-08**
Model additionally adjusted for HDL-C, LDL-C and TG [d]					Discovery panel [b]			Replication panel [c]			Both panels combined		
CpG	Chr	Position	location	Gene	Effect	SE	P-value	Effect	SE	P-value	Effect	SE	P-value
cg17901584	**1**	55353706	TSS1500	***DHCR24***	**-0.0188**	**0.0022**	**2.27E-18**	**-0.0108**	**0.0035**	**1.79E-03**	**-0.0166**	**0.0018**	**1.15E-19**
cg06500161	**21**	43656587	gene body	***ABCG1***	**0.0074**	**0.001**	**1.21E-12**	**0.0093**	**0.0023**	**7.32E-05**	**0.0077**	**0.001**	**5.37E-16**
cg15348274	12	123451191	TSS200	*ABCB9*	**-0.0148**	**0.0026**	**1.08E-08**	0.0028	0.0055	6.05E-01	**-0.0116**	**0.0023**	**7.62E-07**
cg05575921	5	373378	gene body	*AHRR*	**-0.013**	**0.0023**	**2.05E-08**	-0.0002	0.0059	9.79E-01	**-0.0113**	**0.0022**	**1.68E-07**
cg05119988	**4**	166251189	5'UTR	***SC4MOL***	**-0.0098**	**0.0018**	**7.73E-08**	**-0.0092**	**0.0034**	**7.03E-03**	**-0.0097**	**0.0016**	**1.86E-09**
cg27243685	**21**	43642366	gene body	***ABCG1***	0.0035	0.0008	1.08E-05	**0.0063**	**0.0016**	**7.70E-05**	**0.004**	**0.0007**	**1.24E-08**
cg10177197	**1**	55316481	gene body	***DHCR24***	0.0037	0.0007	3.66E-07	0.0041	0.0017	1.71E-02	**0.0038**	**0.0007**	**1.98E-08**

[a] Model adjusted for leukocyte proportions, batch effects, sex, age, smoking status, body mass index, systolic blood pressure, anti-hypertensive medication, presence of coronary heart diseases, prevalent type 2 diabetes, serum total cholesterol.
Bonferroni correction for significance for discovery P < 1.0E-07 and P < 7.14E-03 for replication.
[b] Discovery panel (n=6,820): Cohorts KORA-F4, SHIP-Trend, ESTHER, LOLIPOP.
[c] Replication panel (n=1,450): Sub-cohorts from The Rotterdam Study (RSIII-1 and RS-Bios)
[d] Model adjusted for leukocyte proportions, batch effects, sex, age, smoking status, body mass index, systolic blood pressure, anti-hypertensive medication, presence of coronary heart diseases, prevalent type 2 diabetes, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol and triglycerides.
Abbreviations: Chr, chromosome; SE, standard error. Bold text indicates statistically significant associations.

eTable 14: Replication of the identified statin-related CpGs in an independent sample excluding type 2 diabetes and pre-type 2 diabetes cases in the replication panel

CpG	Chr	Position	location	Gene	Effect	SE	P-value
cg17901584	1	55353706	TSS1500	*DHCR24*	-0.0165	0.0041	**5.21E-05**
cg06500161	21	43656587	gene body	*ABCG1*	0.0086	0.0028	**2.08E-03**
cg05119988	4	166251189	5'UTR	*SC4MOL*	-0.0136	0.0039	**5.36E-04**
cg27243685	21	43642366	gene body	*ABCG1*	0.0068	0.0019	**3.53E-04**
cg10177197	1	55316481	gene body	*DHCR24*	0.0081	0.0020	**4.24E-05**

Model adjusted for leukocyte proportions, batch effects, sex, age, smoking status, body mass index, systolic blood pressure, anti-hypertensive medication, presence of coronary heart diseases
Sample from RS (n=1065). Bonferroni correction P < 7.14E-0
Abbreviations: Chr, chromosome; SE, standard error. Bold text indicates statistically significant associations.

eTable 15: Replication of the association between current statin use and DNA methylation at the five identified CpGs in a set of participants from the replication panel with high LDL-C levels (≥ 70 mg/dL and ≥ 100 mg/dL)

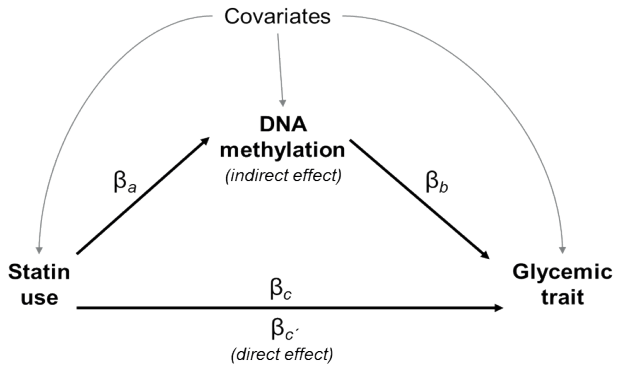
			All participants from the replication panel [a]			Participants from the replication panel with LDL-C levels ≥ 70 mg/dL only [b]			Participants from the replication panel with LDL-C levels ≥ 100 mg/dL only [c]		
CpG site	Chr	Gene	Effect	SE	P-value	Effect	SE	P-value	Effect	SE	P-value
cg17901584	1	*DHCR24*	-0.0145	0.0035	3.06E-05	-0.0156	0.0036	1.38E-05	-0.0202	0.0046	9.12E-06
cg06500161	21	*ABCG1*	0.0139	0.0024	7.84E-09	0.0129	0.0025	2.72E-07	0.0143	0.0031	4.44E-06
cg05119988	4	*SC4MOL*	-0.0118	0.0035	6.81E-04	-0.0085	0.0036	0.0188	-0.0133	0.0045	3.33E-03
cg27243685	21	*ABCG1*	0.0074	0.0016	6.08E-06	0.0077	0.0017	7.74E-06	0.0067	0.0021	1.83E-03
cg10177197	1	*DHCR24*	0.0060	0.0017	5.10E-04	0.0051	0.0018	4.85E-03	0.0061	0.0024	9.79E-03

All analyses adjusted for leukocyte proportions, batch effects, sex, age, smoking status, body mass index, systolic blood pressure, anti-hypertensive medication, presence of coronary heart diseases, prevalent type 2 diabetes.
Bonferroni correction for replication P < 7.14E-03.
[a] Sample size n=1320 (RS) (RS), from which n=247 participants were current statin users and n=1320 were never users.
[b] Sample size participants LDL ≥ 70 mg/dL only, n=1267 (RS), from which n=216 participants were current statin users and n=1051 were never users.
[c] Sample size participants LDL ≥ 100 mg/dL only, n=1065 (RS), from which n=110 participants were current statin users and n=955 were never users.
Abbreviations: Chr, chromosome; SE, standard error; mg/dL, milligrams per deciliter.

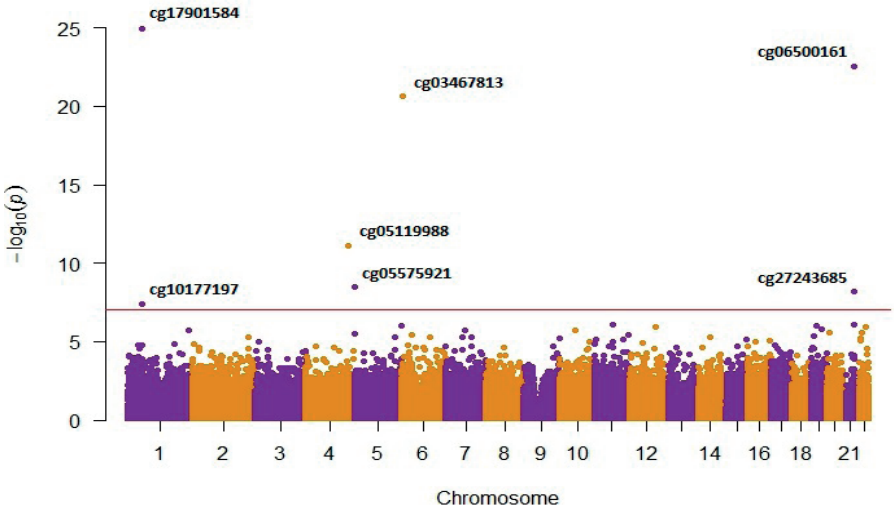
eTable 16: Trans-ethnic replication of the association of current statin use (compared to non-current users)

					Caucasian group as discovery panel [a]			British-Asian group as replication panel [b]		
CpG	Chr	Position	location	Gene	Effect	SE	P-value	Effect	SE	P-value
cg06500161	21	43656587	TSS1500	*ABCG1*	0.0129	0.0013	4.60E-22	**0.008**	**0.001**	**1.01E-09**
cg17901584	1	553153706	gene body	*DHCR24*	-0.0199	0.0022	2.42E-19	**-0.019**	**0.003**	**4.09E-11**
cg27243685	21	43642366	5'UTR	*ABCG1*	0.0063	0.0009	3.69E-12	**0.003**	**0.001**	**2.76E-03**
cg10177197	1	55316481	gene body	*DHCR24*	0.005	0.0008	1.68E-09	**0.003**	**0.001**	**6.99E-03**
cg05119988	4	166251189	gene body	*SC4MOL*	-0.0118	0.0021	1.59E-08	**-0.012**	**0.002**	**3.26E-08**
cg17475467	1	55316769	location	*DHCR24*	0.0048	0.0009	3.96E-08	0.002	0.001	0.027

Model adjusted for leukocyte proportions, batch effects, sex, age, smoking status, body mass index, systolic blood pressure, anti-hypertensive medication, presence of coronary heart diseases, prevalent type 2 diabetes.
Bonferroni correction for significance for discovery P < 1.0E-07 and P < 8.33E-03 for replication.
[a] Sensitivity analysis-Discovery panel (n=4,349): Cohorts KORA, SHIP-Trend, ESTHER and RS.
[b] Sensitivity analysis-Replication panel (n=3,849): Cohort LOLIPOP.
Abbreviations: Chr, chromosome; SE, standard error.
Bold text indicates the CpGs that replicated the independent panel.



eFigure 1. Causal Mediation Analysis
Schematic representation of the causal mediation analysis. βc' represents the causal direct effect of statin use (exposure) on the glycemic traits and type 2 diabetes (outcome). βc represents a non-causal pathway. βa represents the effect of statin use on DNA methylation. βb represents the effect of DNA methylation (mediator) on the outcome, which would be mediating an indirect effect of statins on the outcome. Covariates that may confound these associations are included in the model.



eFigure 2. Manhattan Plot of the epigenome-wide associations between current statin use and DNA methylation (compared to non-current use)
Using a model adjusted for batch effects, leukocyte proportions and cardiometabolic risk factors, we found genome-wide associations of seven methylation sites annotated to *DHCR24* (cg17901584 and cg10177197, chromosome 1), *SC4MOL* (cg05119988, chromosome 4), *ABCG1* (cg06500161 and cg27243685, chromosome 21), *AHRR* (cg05575921, chromosome 5) and *FAM50B* (cg03467813, chromosome 6). Of these, five CpGs, annotated to *DHCR24, ABCG* and *SC4MOL*. From them, five could be replicated in an independent cohort.

REFERENCES

1. Cholesterol Treatment Trialists C, Baigent C, Blackwell L, et al. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. Lancet 2010;376:1670-81.
2. Casula M, Mozzanica F, Scotti L, et al. Statin use and risk of new-onset diabetes: A meta-analysis of observational studies. Nutr Metab Cardiovasc Dis 2017;27:396-406.
3. Thakker D, Nair S, Pagada A, Jamdade V, Malik A. Statin use and the risk of developing diabetes: a network meta-analysis. Pharmacoepidemiol Drug Saf 2016;25:1131-49.
4. Allen SC, Mamotte CDS. Pleiotropic and Adverse Effects of Statins-Do Epigenetics Play a Role? J Pharmacol Exp Ther 2017;362:319-26.
5. Dogan MV, Grumbach IM, Michaelson JJ, Philibert RA. Integrated genetic and epigenetic prediction of coronary heart disease in the Framingham Heart Study. PLoS One 2018;13:e0190549.
6. Muka T, Nano J, Voortman T, et al. The role of global and regional DNA methylation and histone modifications in glycemic traits and type 2 diabetes: A systematic review. Nutr Metab Cardiovasc Dis 2016;26:553-66.
7. Raum E, Rothenbacher D, Low M, Stegmaier C, Ziegler H, Brenner H. Changes of cardiovascular risk factors and their implications in subsequent birth cohorts of older adults in Germany: a life course approach. Eur J Cardiovasc Prev Rehabil 2007;14:809-14.
8. Rathmann W, Haastert B, Icks A, et al. High prevalence of undiagnosed diabetes mellitus in Southern Germany: target populations for efficient screening. The KORA survey 2000. Diabetologia 2003;46:182-9.
9. Volzke H, Alte D, Schmidt CO, et al. Cohort profile: the study of health in Pomerania. Int J Epidemiol 2011;40:294-307.
10. Chahal NS, Lim TK, Jain P, Chambers JC, Kooner JS, Senior R. Does subclinical atherosclerosis burden identify the increased risk of cardiovascular disease mortality among United Kingdom Indian Asians? A population study. Am Heart J 2011;162:460-6.
11. Ikram MA, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol 2017;32:807-50.
12. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 2013;45:1238-43.
13. Schurmann C, Heim K, Schillert A, et al. Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. PLoS One 2012;7:e50938.
14. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 2012;13:86.
15. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014;30:1363-9.
16. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R Package for Causal Mediation Analysis. J Stat Softw 2014;59.

17. Steenaard RV, Ligthart S, Stolk L, et al. Tobacco smoking is associated with methylation of genes related to coronary artery disease. Clin Epigenetics 2015;7:54.
18. Imai K, Keele L, Yamamoto T. Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. Stat Sci 2010;25:51-71.
19. Piepoli MF, Hoes AW, Agewall S, et al. [2016 European Guidelines on cardiovascular disease prevention in clinical practice] Wytyczne ESC dotyczace prewencji chorob ukladu sercowonaczyniowego w praktyce klinicznej w 2016 roku. Kardiol Pol 2016;74:821-936.
20. Perk J, De Backer G, Gohlke H, et al. European guidelines on cardiovascular disease prevention in clinical practice (version 2012) : the fifth joint task force of the European society of cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of nine societies and by invited experts). Int J Behav Med 2012;19:403-88.
21. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009;41:1149-60.
22. Wong J, Quinn CM, Gelissen IC, Jessup W, Brown AJ. The effect of statins on ABCA1 and ABCG1 expression in human macrophages is influenced by cellular cholesterol levels and extent of differentiation. Atherosclerosis 2008;196:180-9.
23. Vaughan AM, Oram JF. ABCA1 and ABCG1 or ABCG4 act sequentially to remove cellular cholesterol and generate cholesterol-rich HDL. J Lipid Res 2006;47:2433-43.
24. Engel T, Fobker M, Buchmann J, et al. 3beta,5alpha,6beta-Cholestanetriol and 25-hydroxycholesterol accumulate in ATP-binding cassette transporter G1 (ABCG1)-deficiency. Atherosclerosis 2014;235:122-9.
25. Olivier M, Tanck MW, Out R, et al. Human ATP-binding cassette G1 controls macrophage lipoprotein lipase bioavailability and promotes foam cell formation. Arterioscler Thromb Vasc Biol 2012;32:2223-31.
26. Obeidat M, Fishbane N, Nie Y, et al. The Effect of Statins on Blood Gene Expression in COPD. PLoS One 2015;10:e0140022.
27. Ding J, Reynolds LM, Zeller T, et al. Alterations of a Cellular Cholesterol Metabolism Network Are a Molecular Feature of Obesity-Related Type 2 Diabetes and Cardiovascular Disease. Diabetes 2015;64:3464-74.
28. Mauldin JP, Nagelin MH, Wojcik AJ, et al. Reduced expression of ATP-binding cassette transporter G1 increases cholesterol accumulation in macrophages of patients with type 2 diabetes mellitus. Circulation 2008;117:2785-92.
29. Waterham HR, Koster J, Romeijn GJ, et al. Mutations in the 3beta-hydroxysterol Delta24-reductase gene cause desmosterolosis, an autosomal recessive disorder of cholesterol biosynthesis. Am J Hum Genet 2001;69:685-94.
30. Andersson HC, Kratz L, Kelley R. Desmosterolosis presenting with multiple congenital anomalies and profound developmental delay. Am J Med Genet 2002;113:315-9.
31. Lamsa R, Helisalmi S, Hiltunen M, et al. The association study between DHCR24 polymorphisms and Alzheimer's disease. Am J Med Genet B Neuropsychiatr Genet 2007;144B:906-10.
32. He M, Kratz LE, Michel JJ, et al. Mutations in the human SC4MOL gene encoding a methyl sterol oxidase cause

psoriasiform dermatitis, microcephaly, and developmental delay. J Clin Invest 2011;121:976-84.
33. Dekkers KF, van Iterson M, Slieker RC, et al. Blood lipids influence DNA methylation in circulating cells. Genome Biol 2016;17:138.
34. Hidalgo B, Irvin MR, Sha J, et al. Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network study. Diabetes 2014;63:801-7.
35. Dayeh T, Tuomi T, Almgren P, et al. DNA methylation of loci within ABCG1 and PHOSPHO1 in blood DNA is associated with future type 2 diabetes risk. Epigenetics 2016;11:482-8.
36. Cederberg H, Stancakova A, Yaluri N, Modi S, Kuusisto J, Laakso M. Increased risk of diabetes with statin treatment is associated with impaired insulin sensitivity and insulin secretion: a 6 year follow-up study of the METSIM cohort. Diabetologia 2015;58:1109-17.
37. Pfeiffer L, Wahl S, Pilling LC, et al. DNA methylation of lipid-related genes affects blood lipid levels. Circ Cardiovasc Genet 2015;8:334-42.
38. Lokk K, Modhukur V, Rajashekar B, et al. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. Genome Biol 2014;15:r54.
39. Burkhardt R, Kenny EE, Lowe JK, et al. Common SNPs in HMGCR in micronesians and whites associated with LDLholesterol levels affect alternative splicing of exon13. Arterioscler Thromb Vasc Biol 2008;28:2078-84.
40. Solomon O, MacIsaac J, Quach H, et al. Comparison of DNA methylation measured by Illumina 450K and EPIC BeadChips in blood of newborns and 14-year-old children. Epigenetics 2018;13:655-64.

Chapter 4.

Link between dietary exposures and prevention of type 2 diabetes

Chapter 4.1.

C-reactive protein partially mediates the inverse association between coffee consumption and risk of type 2 diabetes: the UK Biobank and the Rotterdam Study cohorts

Carolina Ochoa-Rosales, Niels van der Schaft, Kim Braun, Frederick K. Ho, Fanny Petermann-Rocha, Jill P. Pell, M. Arfan Ikram, Carlos A. Celis-Morales*, Trudy Voortman*.

Submitted

ABSTRACT

Background: Coffee consumption has been linked to lower risk of type 2 diabetes (T2D). We hypothesized that this may be mediated by the effect of coffee compounds on inflammation and adipokines.

Objective: To study the association between coffee consumption and risk of T2D, and the potential mediating role of biomarkers of inflammation.

Research design and methods: Using two large population-based cohorts, the UK-Biobank (UKB; n=145,370) and the Rotterdam Study (RS; n=7,172), we investigated associations of coffee consumption with: incident T2D using proportional hazard regressions; longitudinal changes of insulin resistance (HOMA-IR) through linear mixed models; and serum concentrations of inflammation biomarkers and adipokines using linear regressions. We investigated inflammatory response as potential mediator of the observed associations between coffee and T2D, using mediation analyses. Models were adjusted for sociodemographic, lifestyle and health-related factors.

Results: During a median follow-up of 9.4 (RS) and 7.4 (UKB) years, 685 and 2,290 incident T2D cases occurred, respectively. In both cohorts, increase in coffee consumption of 1 cup/day was associated with 4 to 6% lower lower risk of T2D (RS, HR=0.94 (95%CI 0.90, 0.98); UKB, HR=0.96 (0.94, 0.98)); with lower log-transformed HOMA-IR (RS, β=-0.017 (-0.024, -0.010)); and with lower log-transformed CRP (RS, β=-0.014 (-0.022, -0.005); UKB, β=-0.011 [-0.012; -0.009]). Coffee-related changes in CRP mediated 3.4% (0.6%; 14.8%, RS) and 9.6% (5.6%; 24.0%, UKB) of the total effect of coffee on T2D onset. Evidence for mediation was also found for adiponectin. Associations were generally stronger among women, never smokers, those with overweight and consumers of ground coffee as compared to instant coffee.

Conclusions: We observed an inverse association between coffee consumption and T2D risk. Subclinical inflammation partly mediates this association.

Key words: Coffee, Type 2 Diabetes Mellitus, Inflammation, C-Reactive Protein, Adiponectin, Adipokines, Epidemiology, Cohort Studies, Mediation Analysis

INTRODUCTION

Coffee is one of the most frequently consumed beverages over the world.[1-3] Coffee contains several bioactive compounds such as chlorogenic acids, caffeine and polyphenols, although the exact composition depends on the genus, brewing and roasting process.[4] Given coffee´s popularity and enriched composition, it has been of great interest to study its potential effects on health. Coffee consumption has previously been linked to lower risk of cardiovascular and metabolic diseases.[5,6] Recent meta-analyses and reviews have consistently concluded that coffee consumption is associated with a decreased risk of developing type 2 diabetes (T2D).[7-9] However, potential mechanisms underlying these associations are still being investigated.[10-12] Coffee consumption has been hypothesized to affect gut microbiome content and diversity and adenosine receptor signaling, and stimulate thermogenesis [7] In addition, higher coffee intake has been associated with lower concentrations of inflammatory markers which may, in turn, favourably affect processes related to cardiometabolic disease.[13]

T2D is partly considered an inflammatory disease, with a large body of evidence from epidemiological studies supporting a positive association with markers of inflammation and oxidative stress such reactive oxygen species (ROS) and C-reactive protein (CRP) with T2D risk.[14-17]. Moreover, we previously demonstrated that many other inflammatory markers, including interleukin (IL) 13, IL-17 and extracellular newly identified receptor for advanced glycation end-products binding protein (EN-RAGE), were associated with insulin resistance and incident T2D. [18] Adipokines, which are molecules secreted by adipose tissue, are also known to play a role in the inflammatory response, regulation of energy homeostasis and insulin resistance. Given that adiposity is an important risk factor for T2D, it is important to also consider the role of adipokines in relation to the association between inflammation and T2D.[19]

Studies on coffee consumption thus far have shown contradicting results with regard to its effects on inflammation.[20] This could be due to discrepancies between studies in the amount and type of coffee consumed, duration of the exposure to coffee and demographic characteristics of the population under study. In addition, evidence regarding a potential mediating role of inflammatory status in the effect of coffee intake on T2D risk is limited.[21] Thus, the aim of this study was to firstly assess the association between coffee consumption and T2D risk; and secondly, to determine the potential role of the inflammatory markers as mediators in such association through causal mediation analysis, among participants from two population-based prospective studies: UK Biobank (UKB) and the Rotterdam Study (RS).

METHODS

Study design

This study involves both prospective and cross-sectional analyses, and is embedded in two large ongoing population-based cohorts: UKB in the United Kingdom and the RS in the Netherlands. UKB, is a prospective cohort study that recruited 502,536 women and men aged 37 to 73 years. Participants were recruited at 22 research centres across England, Scotland and Wales, between April 2006 and December 2010.[22-24] Baseline data were collected at the time of recruitment, and follow-up data for the longitudinal analyses were available, at the time of conducting the study, up to September 27th, 2017.

The RS is a population-based cohort study conducted in the district of Ommoord in Rotterdam, the Netherlands. Its design has been described elsewhere.[25] Briefly, in 1990 the first cohort (RS-I) of the RS recruited 7,983 participants aged 55 years or older. The second cohort (RS-II) started in 2000 with 3,011 new participants who had become 55 years of aged or moved into the district. The third cohort (RS-III) was initiated with the inclusion of 3,932 new individuals aged 45 years or older. Follow-up examinations were performed every 3-5 years at the research centre. The current study used baseline data from three sub-cohorts (RS-I-3, RS-II-1, RS-III-1). For the prospective analyses, the follow-up data of these participants were used, which was complete up to January 1st, 2012.

Assessment of coffee consumption

In UKB, data on coffee consumption were collected through food-frequency questionnaires (FFQs) self-administrated at the baseline visit at each of the assessment centres using a touch-screen questionnaire. Questions on coffee intake included coffee type, wherein coffee drinkers had to report whether it was mostly decaffeinated, instant, or ground coffee. Details on the questions[26] and on the development and validation of the web-based questionnaire[27] can be found elsewhere. In RS, data on coffee consumption were obtained through self-report during home interviews among participants of sub-cohorts RS-I and RS-II, and through a FFQ for sub-cohort RS-III-1. Additional data on consumption of specific food groups among RS-I and RS-II participants were collected through a 170-item FFQ in which subjects were asked about the frequency and amount of foods consumed in the past. Participants of sub-cohort RS-III completed an updated and more extensive 389-item FFQ.

Assessment of inflammatory markers and adipokines

In UKB, C-reactive protein (CPR, ug/mL) was measured in plasma samples collected at the baseline assessment centres and analysed using an immune-turbidimetric assay (Beckman Coulter AU5800). In RS, high-sensitivity CRP was measured in serum collected at the research centre and stored at -80°C. A rate near-infrared particle immunoassay (IMMAGE Immunochemistry System; Beckman Coulter, San Diego, CA) was performed. A random subset of 856 samples of citrate plasma (200 µL) from RS was sent in July 2008 to Rules-Based Medicine, Austin, Texas to measure specific biomarkers of inflammation and adipokines: EN-RAGE (ng/ml), IL-13 (pg/mL), IL-17 (IL17, pg/mL), IL-18 (pg/mL), IL-1 receptor antagonist (IL1ra, pg/mL), complement factor H (CFH, ug/mL), complement 3 (C3, mg/mL), tumor necrosis factor receptor 2 (TNFRII, ng/mL), adiponectin (ug/mL) and leptin (ng/mL). They were quantified using multiplex immunoassay on a custom-designed human multianalyte profile. The intra-assay variability was less than 4% and the inter-assay variability was less than 13%.

Insulin resistance and T2D ascertainment

In UKB, incident T2D was derived from linkage to primary care data. Records were extracted for 45% of the cohort (228,495 participants). The end of coverage (extract date) was May 2017 for Scotland, September 2017 for Wales and August 2017 for England. Detailed linkage procedures are available at http://biobank.ndph.ox.ac.uk/showcase/showcase/docs/primary_care_data.pdf. We defined incident T2D as primary care diagnosed with ICD-10 (international classification of diseases, 10th revision) code E11. ICD codes were converted to read codes using UKB's look-up table. All participants with T2D from primary care data and who were diagnosed before their UKB baseline assessment visit were excluded from the analyses. In RS cases of T2D, plasma fasting levels of glucose and insulin were ascertained through active follow-up using general practitioners' records, hospital discharge letters and glucose measurements from follow-up visits, which take place every 3-5 years.[28] In both cohorts, T2D was defined according to the current WHO guidelines: fasting blood glucose ≥7.0 mmol/L, non-fasting blood glucose ≥11.1 mmol/L (when fasting samples were unavailable), or use of blood glucose-lowering medication (RS).[28] Information on blood glucose-lowering medication use was extracted from structured home interviews and linkage to pharmacy dispensing records. At baseline, more than 95% of RS population was covered by the pharmacies in the study area. All potential incident cases of T2D were independently adjudicated by two study physicians. In case of disagreement, consensus was reached through the input of an endocrinologist. In RS, glucose concentrations were measured in blood samples at baseline and at follow-up using the glucose hexokinase method within 1 week of sampling.[29]

Insulin levels were determined by means of electrochemiluminescence immunoassay technology using a Modular Analytics E170 analyzer (Roche Diagnostics GmbH, Mannheim, Germany). Homeostatic model assessment–insulin resistance (HOMA-IR) was calculated as: HOMA-IR = glucose (nmol/L) * insulin (uU/L)/22.5.

Covariates assessment

Variables considered as potential confounders included demographic factors (age, sex, ethnicity (UKB only) and measures of socioeconomic status (education level in RS and deprivation index[30] in UKB)); lifestyle factors (smoking, diet, physical activity, and alcohol and tea intake); and cardiovascular risk factors (hypertension, blood lipids, and body mass index (BMI)). Further details of the measurement procedures can be found elsewhere for RS[25] and UKB.[31] Briefly, ethnicity in UKB was assessed at recruitment through self-report and classified into 5 categories (white, mixed, south Asian, black, Chinese). As a measure of socioeconomic status in UKB, Townsend deprivation index scores were derived from national census data about car ownership, household overcrowding, owner occupation, and unemployment aggregated for postcodes of residence,[30] while self-reporteded highest education level was used in RS.

Data on lifestyle factors were obtained through self-reported questionnaires. In both cohorts, smoking status was categorized into never, former or current smoking. Physical activity was self-reported using the validated International Physical Activity Questionnaire in UKB,[32] and using Zutphen[33] (RS-I and RS-II) and LASA[34] (RS-III) questionnaires in RS. Activity was expressed for both cohorts in MET-hours/week.

In UKB, self-reported data on tea, alcohol and food groups consumption were collected through the aforementioned twenty-nine questions-FFQ, which additionally asked about daily tea intake (number of cups/day) and alcohol consumption frequency (daily or almost daily; 3-4 times a week; once or twice per week; 1-3 times a month; special occasions only; and never) at baseline. Further, the FFQ also assessed the intake of 10 food items. These data were used to build a diet quality score for the adherence to the recommendations of the UK Eatwell Guide,[35] the Food-Based Dietary Guidelines from the European Food Safety Authority),[36] or the or median intake for those food items without a consumption guideline.[37] In RS, food consumption was assessed using 170-item (RS-I and RS-II) and 389-item (RS-III) FFQs. A diet quality score (0–14) reflecting adherence to Dutch dietary guidelines was calculated by adding the scores for 14 components of the Dutch dietary guidelines, as described elsewhere.[38]

In both cohorts, physical measures and collection of blood samples were assessed at the research centres. BMI was calculated as weight (kg) divided by height (metres)

squared. Blood pressure was the mean of 2 consecutive measures. Total cholesterol (total-C) and high-density lipoprotein cholesterol (HDL-C) were measured in fasting blood samples, using the cholesterol oxidase-peroxidase (CHO-POD) enzymatic reaction and enzyme immune-inhibition methods respectively (AU5800 chemistry analyzer, Beckman Coulter). Disease information was collected by self-administrated questionnaires and by general practitioner records. Definitions of hypertension and cardiovascular disease (CVD) cases can be found elsewhere for RS[25] and UKB.[31]

Study population

After exclusion of participants with missing data on coffee intake, T2D incidence and CRP levels; and exclusion of baseline cases of T2D and CVD, final samples of 145,370 participants in UKB and 7,172 in RS, were analysed. A subset of n=846 in RS had available data on specific inflammatory markers and adipokines. A subset of n=111,159 in UKB had data on coffee type. Figure 1 shows details on study sample selection.

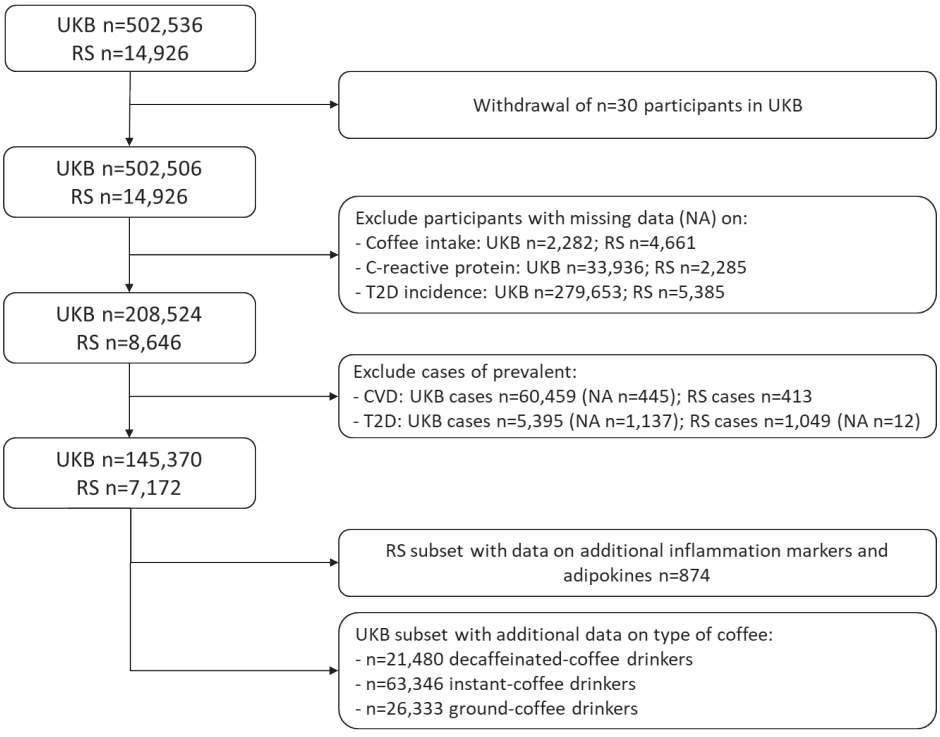


Figure 1: Flux diagram of the study samples. Abbreviations: Rotterdam Study, RS; UK Biobank, UKB; type 2 diabetes, T2D; cardiovascular disease, CVD.

Statistical analyses

Analyses were performed separately in RS and UKB. For all analyses, coffee consumption was analysed per one cup/day increase and in categories. For consistency participants in both cohorts were classified into four categories according to their daily consumption of coffee: (i) non-consumers (0 cups); (ii) 0.5-2 cups/day; (iii) 3-4 cups/day; and (iv) 5 or more cups/day. Non-coffee consumers were used as the reference group. Cox proportional hazards regression was used to assess the association between coffee intake and incident T2D. Results were reported as hazard ratios (HR) with corresponding 95% confidence intervals (CIs). The timescale in these regression models was follow-up time in years from baseline examination until incident T2D, death, or withdrawal from the study, whichever came first. Schoenfeld residuals were inspected to detect deviance from the proportional hazards assumption. Linear mixed-effect models were used to study associations between coffee intake and HOMA-IR changes over time in participants of RS for whom at least two measurements of HOMA-IR were available. HOMA-IR was transformed to the natural logarithmic scale to approximate a normal distribution. Random intercept (participants) and random slopes (time) were included in the models. Results were expressed as coefficients (β) with corresponding 95% CI.

For the aforementioned regressions, two statistical models were designed. Model 1 was adjusted for age, sex, RS sub-cohort (RS only), highest attained level of education (RS only), deprivation index (UKB only), smoking, physical activity, diet quality score and daily alcohol and tea intake. Model 2 was additionally adjusted for BMI, hypertension, and serum total/HDL cholesterol ratio. In RS, alcohol intake in grams/day was transformed to the natural logarithmic scale.

To investigate inflammatory response as a potential underlying mechanism linking coffee consumption and T2D, we conducted linear regression analyses to examine the cross-sectional associations between coffee intake and CRP in both cohorts, and leptin, adiponectin, EN-RAGE, TNFR-II, IL-17, IL-1RA, CHF, IL-18, IL-13, C3 in RS. Biomarkers were natural log-transformed. Results were expressed as coefficients (β) with corresponding 95% CI.

The biomarkers significantly associated with one-cup increase in coffee consumption were subsequently tested in a mediation analysis to dissect the total effect of coffee consumption into direct and indirect effect, the latter being the effect that goes via a mediator (biomarkers). We first estimated the associations between coffee consumption and biomarker (*a,* in Figure 2), between biomarker and T2D adjusted for coffee intake (*b*) and between coffee intake and T2D adjusted for biomarker (*c´*).

Next, we used the *mediation* R package[39] to estimate the proportion of the total effect (*c*) of coffee on T2D that is mediated by coffee-related changes in the biomarkers concentrations, by comparing the effect estimates of each association of coffee consumption with T2D, under the sequential ignorability assumption, which assumes no unmeasured confounding. Quasi-Bayesian confidence intervals were constructed for the estimated effects with 10,000 simulations. Results were expressed as proportion mediated in percentage and its 95% CIs. Ten-fold multiple imputation with chained equations was performed to account for missing data on covariates, ranging from 0% to 19.7% (RS) or from 0 to 18.9% (UKB). A two-sided p-value <0.05 was considered statistically significant. nonlinearity was evaluated with penalized splines[40] using *pspline* R package. No significant evidence for nonlinearity was found. The results shown in this work correspond to the pooled results of the ten imputed datasets. All analyses were performed using R version 3.6.1 (R foundation for Statistical Computing, Vienna, Austria).

RESULTS

Baseline characteristics

The mean age was 65.1 (SD=9.4) years among RS participants (n=7,172), of which 59.7% (n=4,283) were female. In UKB (n=145,370) the mean age was 55.2 (SD=8.10) years and 58.0% (n=84,343) were female. During a median follow-up of 9.41 (IQR=4.46; 11.76) years in RS and 7.44 (IQR=6.81; 8.34) in UKB; 685 and 2,290 incident T2D cases occurred, respectively. Table 1 and Supplementary Tables S-1 and S-2 show more details on participant's characteristics.

Table 1: Summary of baseline characteristics of the Rotterdam Study and UK Biobank participants, stratified by coffee consumption categories

	Non-coffee drinkers	0.5-2 cups/ day	3-4 cups/ day	5 or more cups/day	Total
The Rotterdam Study					
n, (%)	542 (7.6)	1,594 (22.2)	2,883 (40.2)	2,053 (28.6)	7,172
Age, mean (SD), years	62.1 (9.8)	68.4 (10.1)	65.8 (9.2)	62.3 (7.8)	65.1 (9.4)
Sex, (%) women	373 (68.8)	1064 (66.8)	1844 (61.8)	1002 (48.8)	4283 (59.7)
Smoking status					
Never smokers	252 (46.5)	652 (40.9)	994 (33.3)	453 (22.1)	2351 (32.8)
Former smoker	209 (38.6)	730 (45.8)	1939 (46.8)	839 (40.9)	3174 (44.3)
Current smoker	81 (14.9)	212 (13.3)	593 (19.9)	761 (37.0)	1647 (22.9)

Table 1: Summary of baseline characteristics of the Rotterdam Study and UK Biobank participants, stratified by coffee consumption categories (*continued*)

	Non-coffee drinkers	0.5-2 cups/ day	3-4 cups/ day	5 or more cups/day	Total
Physical activity, median (IQR), MET-hours/week	61.9 (68.4)	71.3 (60.9)	60.8 (60.8)	70.5 (67. 6)	72.0 (63.4)
BMI, mean (SD), kg/m2	26.8 (4.7)	26.7 (4.0)	26.9 (3.8)	27.1 (4.0)	26.9 (4.0)
CRP median (IQR), ug/mL	1.40 (2.80)	1.27 (2.9)	1.50 (2.70)	1.36 (2.47)	1.5 (2.7)
Incident type 2 diabetes cases, n (%)	52 (9.6)	171 (10.7)	276 (9.6)	187 (9.1)	685 (9.6)
The UK Biobank					
n, (%)	31,773 (21.9)	66,153 (45.5)	30,440 (20.9)	17,004 (11.7)	145,370
Age, mean (SD), years	53.9 (8.09)	55.8 (8.10)	55.5 (8.03)	54.5 (7.94)	55.2 (8.10)
Sex, (%) women	19,818 (62.4)	39,088 (59.1)	16,715 (54.9)	8,722 (51.3)	84,343 (58.0)
Smoking status, n (%)					
Never smokers	19,659 (61.9)	39,535 (59.8)	16,624 (54.6)	7,559 (44.4)	83,377 (57.4)
Former smoker	9,077 (28.6)	21,335 (32.3)	10,325 (33.9)	5,711 (33.6)	46,448 (32.0)
Current smoker	3,037 (9.6)	5,283 (8.0)	3,491 (11.5)	3,734 (22.0)	15,545 (10.7)
Physical activity, median (IQR), MET-hours/week	48.1 (46.8)	45.8 (43.7)	45.0 (43.3)	47.3 (45.9)	46.3 (44.4)
BMI, mean (SD), kg/m^2	26.2 (4.4)	25.9 (4.0)	26.6 (4.2)	27.1 (4.3)	26.3 (4.20)
CRP median (IQR), ug/mL	1.15 (1.84)	1.03 (1.57)	1.08 (1.65)	1.76 (1.19)	1.08 (1.67)
Incident type 2 diabetes cases, n (%)	205 (1.6)	298 (1.2)	147 (1.1)	95 (1.4)	745 (1.3)

Coffee consumption, insulin resistance and incident T2D

After comprehensive adjustment for covariates (model 2), higher coffee intake was associated with lower T2D risk in both RS (HR=0.94 per cup/day increase (95%CI=0.90; 0.98)) and UKB (HR=0.96 (0.94; 0.99)). Compared to non-consumers, those drinking ≥5 coffee cups/day had a HR of 0.62 (0.45; 0.86) and 0.80 (0.69; 0.93) times lower risk of developing T2D, in RS and UKB respectively (Table 2). Among RS participants, we further observed an association between higher coffee consumption and a longitudinal decrease in HOMA-IR levels (β=-0.017 log-HOMA-IR per cup/day increase (-0.024; -0.010)).

Coffee consumption and markers of inflammation

After comprehensive adjustment for covariates (model 2), higher coffee intake in dose-response analyses was associated with lower circulating levels of CRP in both RS (β=-0.014 ug/mL log-CRP per cup/day (95%CI=-0.022; -0.005)) and UKB (β=-0.004 (-0.006; -0.002)), (Table 3). Among the subset of participants from RS with additional data on other inflammatory markers, we observed that higher coffee intake was as-

Table 2: Associations of coffee consumption with longitudinal measures of HOMA-IR and incident T2D in The Rotterdam Study and with incident T2D in the UK Biobank

		The Rotterdam Study				UK Biobank	
		Longitudinal HOMA-IR[a] (n=4,138)		Incident T2D (n = 7,172)		Incident T2D (n = 145,370)	
	Coffee (cups per day)	β (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Model 1	Per cup increase	-0.011 (-0.020; -0.002)	**0.013**	0.95 (0.91; 0.99)	**0.033**	(0.95; 0.99)	**0.043**
	0	ref		ref		ref	
	0.5-2	0.044 (-0.021; 0.109)	0.187	0.83 (0.60; 1.14)	0.253	0.90 (0.81; 0.99)	**0.004**
	3-4	-0.005 (-0.065; 0.054)	0.856	0.67 (0.50; 0.91)	**0.011**	0.84 (0.74; 0.96)	**0.012**
	≥ 5	-0.028 (-0.089; 0.033)	0.376	0.67 (0.49; 0.92)	**0.014**	0.88 (0.75; 1.01)	0.085
	p-trend		**0.028**		**0.005**		**0.015**
Model 2	Per cup increase	-0.017 (-0.024; -0.010)	**<0.001**	0.94 (0.90; 0.98)	**0.008**	0.96 (0.94; 0.99)	**0.002**
	0	ref		ref		ref	
	0.5-2	0.020 (-0.033; 0.073)	0.465	0.79 (0.57; 1.09)	0.147	0.92 (0.83; 1.02)	0.124
	3-4	-0.022 (-0.071; 0.026)	0.366	0.64 (0.47; 0.87)	**0.004**	0.82 (0.72; 0.94)	**0.004**
	≥ 5	-0.064 (-0.114; -0.014)	**0.012**	0.62 (0.45; 0.86)	**0.004**	0.80 (0.69; 0.93)	**0.005**
	P-trend		**<0.001**		**0.001**		**0.001**

Model 1: adjusted for age, sex, Rotterdam Study cohort (RS only), level of education (RS only), deprivation index (UKB only), ethnicity (UKBB only), tea consumption, alcohol consumption, smoking status, physical activity, and diet quality score; Model 2: additionally adjusted for body mass index, hypertension, ratio serum total cholesterol/HDL.
Cases of incident T2D in the Rotterdam Study n=685 and in the UK Biobank n=2,290
[a]natural logarithm-transformed
Bold text indicates statistically significant associations (p<0.05).

sociated with higher adiponectin concentrations (β=0.025 per cup/day (0.007; 0.042)). Higher coffee intake was also associated with lower complement-3 and higher IL-18, but only in certain categories of coffee consumption as compared to non-consumers (e.g. 3-4 cups/day β=-0.078 (-0.137; -0.020) for complement-3 and β=0.175 (0.029; 0.322) for IL-18). No significant findings were observed for the other inflammatory markers. The results did not change after adjustment for CRP (Table 4).

Mediation analyses

A schematic representation of the causal mediation analyses and results are shown in Figure 2 and Table 5. In RS, the association between higher coffee intake and lower CRP levels was statistically significant ($a_{CRP\text{-}RS}$, β=-0.014, (-0.022; -0.005)); as well as the association between CRP and incident T2D, adjusting for coffee intake ($b_{CRP\text{-}RS}$, HR=1.17 (1.04; 1.31)). Similarly, coffee intake was associated with incident T2D independent of CRP ($c'_{independent\ of\ CRP\text{-}RS}$, HR=0.94 (0.90; 0.99)). The mediation analysis in RS demonstrated that, of the total effect of coffee on T2D onset, there was a statisti-

Table 3: Associations of coffee consumption with CRP levels per categories of consumption

	Coffee (cups/day)	The Rotterdam Study (n = 7,172) β (95% CI)	P	UK Biobank (n = 145,370) β (95% CI)	P
Model 1	Per cup increase	-0.009 (-0.017; 0.000)	0.058	-0.002 (-0.003; 0.001)	0.101
	0	ref		**ref**	
	0.5-2	-0.030 (-0.096; 0.036)	0.371	-0.048 (-0.056; -0.039)	**<0.001**
	3-4	-0.048 (-0.110; 0.014)	0.128	-0.039 (-0.049; -0.028)	**<0.001**
	≥ 5	-0.067 (-0.131; -0.002)	**0.043**	-0.020 (-0.032; -0.008)	**<0.001**
	Trend		**0.026**		**<0.001**
Model 2	Per cup increase	-0.014 (-0.022; -0.005)	**0.002**	-0.011 (-0.012; -0.009)	**-0.011 (-0.012; -0.009)**
	0	ref		ref	
	0.5-2	-0.039 (-0.102; 0.023)	0.217	-0.041 (-0.048; -0.033)	**-0.041 (-0.048; -0.033)**
	3-4	-0.064 (-0.123; -0.005)	**0.034**	-0.059 (-0.069; -0.050)	**-0.059 (-0.069; -0.050)**
	≥ 5	-0.084 (-0.153; -0.030)	**0.004**	-0.066 (-0.077; -0.054)	**-0.066 (-0.077; -0.054)**
	Trend		**0.001**		**<0.001**

Model 1: adjusted for age, sex, Rotterdam Study cohort (RS only), level of education (RS only), deprivation index (UKBB only), ethnicity (UKBB only), tea consumption, alcohol consumption, smoking status, physical activity, and diet quality score; Model 2: further adjusted for body mass index, hypertension, ratio serum total cholesterol/HDL.
Natural logarithm transformation was used for CRP (ug/mL).
Bold text indicates statistically significant associations (p<0.05).

cally significant proportion mediated by coffee-related changes in CRP, of 3.4% (0.6%; 14.8%). Further, in UKB there was a significant inverse association between coffee intake and CRP ($a_{CRP\text{-}UKB}$, β=-0.011 (-0.012; -0.009)); as well as the effect of CRP on T2D incidence, adjusting by coffee consumption ($b_{CRP\text{-}UKB}$, HR=1.45 (1.36; 1.53)). The association between coffee consumption and T2D incidence adjusting for CRP was also significant ($c'_{independent\ of\ CRP\text{-}UKBB}$, HR=0.97 (0.94; 0.99)). The mediation analysis in UKB suggested that changes in CRP mediate 9.6% (5,7%; 24.0%) of the total effect of coffee on T2D.

Additionally, adiponectin, complement 3 and IL-18 were also tested in RS, adjusting for CRP in each case. Evidence for mediation was found only for adiponectin, whose effect on T2D incidence was independent of coffee consumption and CRP levels ($b_{adiponectin\text{-}RS}$, HR=0.58 (0.32; 0.83)). In this case, the direct effect of coffee on T2D did not remain significant after adjustment for adiponectin ($c'_{independent\ of\ CRP+adiponectin\text{-}RS}$, HR=0.90 (0.80; 1.01)). Estimates for complement-3 and IL-18 did not suggest they were part of the causal pathway.

Table 4: Associations between coffee intake and T2D-related inflammatory markers and adipokines in RS (n=846), with and without additional adjustment for CRP

	Coffee (cups/day)	Multivariable adjusted (model 2)		Additionally adjusted for CRP	
		β (95% CI)	P	β (95% CI)	P
Adiponectin (ug/mL)	Per cup increase	0.022 (0.003; 0.041)	**0.023**	0.023 (0.005; 0.041)	**0.010**
	0	ref		ref	
	0.5-2	0.093 (-0.065; 0.251)	0.247	0.068 (-0.078; 0.213)	0.362
	3-4	0.143 (-0.010; 0.296)	0.068	0.137 (-0.003; 0.278)	0.056
	≥ 5	0.162 (0.001; 0.323)	**0.049**	0.137 (-0.011; 0.285)	0.070
	Trend		**0.029**		**0.020**
Leptin (ng/mL)	Per cup increase	0.009 (-0.013; 0.032)	0.417	0.011 (-0.012; 0.033)	0.358
	0	ref		ref	
	0.5-2	-0.006 (-0.193; 0.181)	0.948	0.008 (-0.177; 0.194)	0.930
	3-4	0.027 (-0.154; 0.208)	0.770	0.035 (-0.144; 0.215)	0.699
	≥ 5	0.009 (-0.182; 0.199)	0.930	0.027 (-0.162; 0.215)	0.783
	Trend		0.729		0.644
ENRAGE (ng/ml)	Per cup increase	0.009 (-0.013; 0.032)	0.417	0.011 (-0.012; 0.033)	0.358
	0	ref		ref	
	0.5-2	-0.006 (-0.193; 0.181)	0.948	0.008 (-0.177; 0.194)	0.930
	3-4	0.027 (-0.154; 0.208)	0.770	0.035 (-0.144; 0.215)	0.699
	≥ 5	0.009 (-0.182; 0.199)	0.930	0.027 (-0.162; 0.215)	0.783
	Trend		0.729		0.644
C3 (mg/mL)	Per cup increase	-0.004 (-0.011; 0.003)	0.282	-0.004 (-0.011; 0.003)	0.261
	0	ref		ref	
	0.5-2	-0.081 (-0.141; -0.021)	**0.009**	-0.068 (-0.126; -0.011)	**0.021**
	3-4	-0.078 (-0.137; -0.020)	**0.008**	-0.073 (-0.129; -0.017)	**0.010**
	≥ 5	-0.083 (-0.145; -0.022)	**0.008**	-0.071 (-0.130; -0.013)	**0.017**
	Trend		0.137		0.141
TNFR II (ng/mL)	Per cup increase	0.001 (-0.011; 0.014)	0.828	0.001 (-0.011; 0.013)	0.819
	0	ref		ref	
	0.5-2	0.042 (-0.059; 0.143)	0.411	0.054 (-0.046; 0.153)	0.292
	3-4	0.030 (-0.068; 0.128)	0.552	0.034 (-0.062; 0.131)	0.487
	≥ 5	0.028 (-0.075; 0.131)	0.590	0.039 (-0.062; 0.141)	0.448
	Trend		0.926		0.969

Model adjusted for age, sex, cohort, highest education level, tea consumption, alcohol consumption, smoking, physical activity, and diet score, body mass index, hypertension, ratio total cholesterol/HDL and CRP.

Abbreviations: EN-RAGE, Extracellular Newly identified Receptor for Advanced Glycation End-products binding protein; C3, complement 3; TNFR II, tumor Necrosis Factor Receptor 2.

Bold text indicates statistically significant associations ($p<0.05$).

Table 4 - Continued: Associations between coffee intake and T2D-related inflammatory markers and adipokines in RS (n=846), with and without additional adjustment for CRP

	Coffee (cups per day)	Multivariable adjusted (model 2)		Additionally adjusted for CRP	
		β (95% CI)	P	β (95% CI)	P
IL-17 (pg/mL)	Per cup increase	-0.005 (-0.022; 0.012)	0.591	-0.003 (-0.020; 0.014)	0.698
	0	ref		ref	
	0.5-2	0.034 (-0.104; 0.172)	0.627	0.032 (-0.105; 0.170)	0.647
	3-4	0.059 (-0.075; 0.192)	0.389	0.061 (-0.071; 0.194)	0.365
	≥ 5	0.016 (-0.125; 0.156)	0.828	0.018 (-0.122; 0.158)	0.804
	Trend		0.996		0.918
IL-1RA (pg/mL)	Per cup increase	0.000 (-0.026; 0.025)	0.973	-0.005 (-0.030; 0.021)	0.717
	0	ref		ref	
	0.5-2	0.147 (-0.065; 0.358)	0.175	0.151 (-0.057; 0.359)	0.155
	3-4	0.143 (-0.062; 0.348)	0.171	0.132 (-0.069; 0.333)	0.198
	≥ 5	0.072 (-0.144; 0.287)	0.515	0.061 (-0.150; 0.273)	0.569
	Trend		0.718		0.528
CFH (ug/mL)	Per cup increase	0.012 (-0.006; 0.030)	0.179	0.011 (-0.007; 0.029)	0.231
	0	ref		ref	
	0.5-2	0.110 (-0.041; 0.262)	0.154	0.108 (-0.043; 0.259)	0.162
	3-4	0.094 (-0.053; 0.240)	0.212	0.087 (-0.059; 0.234)	0.242
	≥ 5	0.154 (-0.001; 0.308)	0.051	0.147 (-0.007; 0.301)	0.061
	Trend		0.121		0.149
IL-18 (pg/mL)	Per cup increase	0.004 (-0.015; 0.022)	0.691	0.001 (-0.017; 0.019)	0.904
	0	ref		ref	
	0.5-2	0.160 (0.009; 0.311)	**0.038**	0.161 (0.012; 0.310)	**0.035**
	3-4	0.175 (0.029; 0.322)	**0.019**	0.168 (0.023; 0.313)	**0.023**
	≥ 5	0.120 (-0.034; 0.274)	0.127	0.113 (-0.039; 0.265)	0.146
	Trend		0.716		0.874
IL-13 (pg/mL)	Per cup increase	0.010 (-0.001; 0.021)	0.074	0.010 (-0.001; 0.021)	0.081
	0	ref		ref	
	0.5-2	-0.009 (-0.099; 0.082)	0.850	-0.013 (-0.103; 0.077)	0.779
	3-4	0.018 (-0.070; 0.105)	0.690	0.013 (-0.074; 0.101)	0.767
	≥ 5	0.023 (-0.069; 0.115)	0.621	0.019 (-0.073; 0.110)	0.692
	Trend		0.224		0.238

Model adjusted for age, sex, cohort, highest education level, tea consumption, alcohol consumption, smoking, physical activity, and diet score, body mass index, hypertension, ratio total cholesterol/HDL.
Abbreviations: EN-RAGE, Extracellular Newly identified Receptor for Advanced Glycation End-products binding protein; TNFR II, tumor Necrosis Factor Receptor 2.
Bold text indicates statistically significant associations ($p<0.05$).

Table 5: Causal mediation analyses for the association of coffee intake (cups/day) with T2D by inflammatory markers in the UK biobank and Rotterdam Study

	a biomarker ~ coffee[a]		*b* T2D ~ biomarker + coffee[a]		*c´* T2D ~ coffee + biomarker[a]		Proportion mediated	
	β (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	Percentage (%) (95% CI)	P
	UK Biobank							
CRP[b]	-0.011 (-0.012; -0.009)	**<0.001**	1.45 (1.36; 1.53)	**<0.001**	0.96 (0.94; 0.99)	**0.001**	9.6 (5.7; 24.4)	**0.002**
	The Rotterdam Study							
CRP[b]	-0.014 (-0.022; -0.005)	**0.002**	1.17 (1.04; 1.31)	**0.008**	0.94 (0.90; 0.99)	**0.010**	3.4 (0.6; 4.8)	**0.015**
Adiponectin[b,c]	0.025 (0.007; 0.042)	**0.006**	0.58 (0.32; 0.83)	**0.006**	0.90 (0.80; 1.01)	0.079	12.2 (-12.9; 81.2)	0.066
C3[b,c]	-0.006 (-0.012; 0.001)	0.094	2.35 (0.78; 7.1)	0.129	0.89 (0.79; 0.99)	**0.048**	3.4 (-4.7; 26.0)	0.207
IL-18[b,c]	0.001 (-0.017; 0.019)	0.090	1.36 (0.88; 2.12)	0.170	0.89 (0.79; 0.99)	**0.047**	-0.1 (-15.9; 12.8)	0.894

[a] Regressions adjusted for age, sex, Rotterdam Study cohort (RS only), level of education (RS only), deprivation index (UKBB only), ethnicity (UKBB only), tea consumption, alcohol consumption, smoking status, physical activity, diet quality score, hypertension, body mass index and ratio serum total cholesterol/HDL.
[b] Naturally logarithm transformed variables
[c] Additionally adjusted for CRP.
Bold text indicates statistically significant associations (p<0.05).

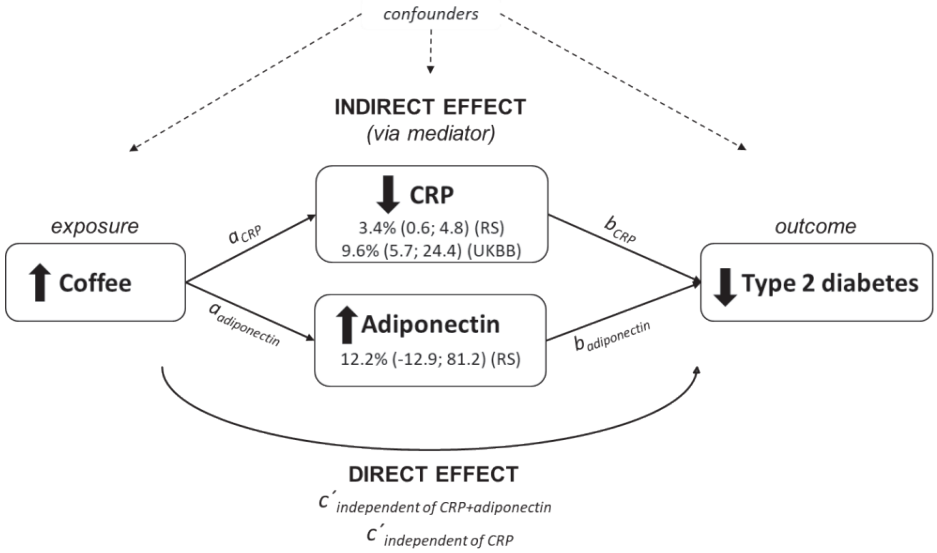


Figure 2. Schematic representation of the causal mediation analyses. These analyses estimated the proportion of the total effect of higher coffee intake (increase in 1 cup/day) on type 2 diabetes incidence that are potentially mediated by coffee-related changes in serum C-reactive protein levels (CRP) and adiponectin. In the figure, a_{CRP} and $a_{adiponectin}$ represent the potential effect of coffee consumption on log-CRP and log-adiponectin levels, respectively. $a_{CRP\text{-}RS}$: β=-0.014 (-0.022; -0.005); $a_{CRP\text{-}UKB}$: β=-0.010 (-0.012; -0.007) and $a_{adiponectin\text{-}RS}$: β=0.025 (0.007; 0.042). Further, b_{CRP} and $b_{adiponectin}$ correspond to the effect of coffee-related changes in log-CRP and log-adiponectin levels on incident T2D, respectively, controlling for coffee consumption. $b_{CRP\text{-}RS}$: HR=1.17 (1.04; 1.31); $b_{CRP\text{-}UKB}$: HR=1.45 (1.36; 1.53); and $b_{adiponectin}$: HR=0.58 (0.32; 0.83). c' relates to the effect of coffee consumption on T2D risk that goes directly or via mediators other than CRP and/or adiponectin. $c'_{independent\ of\ CRP\text{-}UKB}$: HR=0.97 (0.94; 0.99); $c'_{independent\ of\ CRP\text{-}RS}$: HR=0.94 (0.90; 0.99); and $c'_{independent\ of\ CRP+adiponectin\text{-}RS}$: HR=0.90 (0.80; 1.01). The analyses estimated that coffee-related changes in CRP levels may constitute a partial mediator mechanism in the effect of coffee intake on T2D, with a proportion mediated of 3.4% (0.6; 4.8) and 9.6% (5.7; 24.4), among RS and UKB participants respectively. Adiponectin may also constitute a potential mediator. Dotted lines represent confounding among the exposure, mediators and outcome variables. Confounders considered were age, sex, education (RS only), deprivation index (UKB only), ethnicity (UKB only), tea consumption, alcohol consumption, smoking, physical activity, diet score, body mass index, hypertension, ratio total cholesterol/HDL model 2).

Additional analyses

The analyses for the association between coffee consumption and T2D were repeated excluding cases of incident T2D which occurred during the first 2 years of follow-up (RS n=589; UKB n=547). No meaningful differences were observed (Supplementary Table S-3). We identified significant interactions between sex, age, smoking and BMI and coffee intake in cups/day for some of the outcomes (Supplementary Table S-4) and, therefore, conducted stratified analyses (Supplementary Table S-5 to S-8). The associations between coffee consumption and T2D, HOMA-IR and CRP were stronger among women, but were significant in both men and women, except for T2D

and HOMA-IR among RS participants (Supplementary Table S-5). In terms of age, we observed that the association between coffee intake and CRP became slightly weaker with older age (interaction coefficient β=0.00090 ug/mL log-CRP per cup/day, (0.00001; 0.00179), p-interaction 0.046). Compared with never smokers, we observed stronger associations among never smokers in both cohorts and among former smokers in RS. Results among former versus current smokers between RS and UKB were inconsistent (Supplementary Tables S-6 and S-7). Those with a higher BMI (≥25 kg/m2) showed stronger associations with lower CRP and HOMA-IR (Supplementary Table S-8). Associations between coffee and type 2 diabetes and CRP were stronger among consumers of regular ground coffee, than among drinkers of instant or decaffeinated coffee (Supplementary Table S-9)

DISCUSSION

In these large population-based cohort studies, we confirmed an association between higher coffee consumption and lower T2D risk. We additionally report that higher coffee consumption is associated with lower longitudinal HOMA-IR and with lower plasma CRP, and higher adiponectin and IL-18 concentrations. Furthermore, the association between coffee consumption and risk of T2D could be mediated by CRP levels.

Although overall findings were generally consistent between the RS and UKB, a few differences between the two cohorts were observed in the strength of associations or for certain categories of coffee intake. These differences might be due to dissimilar coffee consumption habits and in the characteristics of both study populations. Average coffee consumption was higher among individuals in the RS cohort as compared to those in UKB. UKB participants reported modal intake of 0.5-2 coffee cups/day (43.2%) followed by 3-4 cups (23.5%), and 21.7% did not drink coffee. In RS, modal consumption was 3-4 cups/day (41.6%), followed by ≥ 5 cups/day (28.6%), and only 7.6% reported not drinking any coffee. The observed distribution of coffee consumption frequencies in this study agrees with recent reports on countries' trends in coffee intake.[1,2] The Netherlands is one of the countries with the highest per capita coffee consumption worldwide. Contrarily, the United Kingdom lies in 45th place regarding coffee consumption, tea being the most popular hot beverage among the British. To account for this difference, tea consumption was included as potential confounder in our models. In addition to these differences, UKB participants were on average around ten years younger than RS participants.

Our findings confirm the evidence from several previous studies of the protective effect of coffee consumption on T2D risk.[7,8,41,42] Several previous studies suggested potential mechanisms explaining these observations. One of them is the action of chlorogenic acid, a phenolic coffee compound, reducing intestinal glucose uptake and reducing blood glucose levels. Furthermore, it has been hypothesized that the antioxidant capacity of some coffee compounds may reduce oxidative stress. Coffee consumption may also modulate adenosine receptor signaling and, in turn, insulin and glucagon signaling. Modulation of the microbiome content and diversity by coffee has also been suggested.[12,43]

We observed beneficial associations between higher coffee intake and lower plasma CRP, lower complement-3 and higher adiponectin concentrations; association with blood IL-18 concentrations; but no significant association with the rest of the studied biomarkers. Ours is the first study to suggest that CRP and adiponectin might be in the causal pathway mediating the association of higher coffee consumption and lower risk of T2D, although with opposite direction effects, by providing evidence from causal mediation analyses. In the same line, a recent study using an instrumental variable analysis (Mendelian randomization) suggested a causal effect of CRP on increasing the risk of T2D.[44]

The small proportion (3.34% in RS) of the total coffee effect that appeared to be mediated by coffee-related lower CRP, may suggest the existence of a more complex network of interplaying metabolic and inflammatory biomarkers and alternative indirect effects of coffee that were not investigated or for which our analyses were underpowered in the current study. For example, CRP is produced in response to increase in the concentrations of pro-inflammatory cytokines IL-6, IL-1 and IL-17[45,46] and recent evidence showed that coffee intake may be associated with a broader range of biomarkers of inflammation and metabolism beyond CRP alone,[47] like adiponectin.

We found evidence supporting a mediator role of adiponectin in the effect of coffee intake on T2D risk. Further, when adiponectin was included as predictor of T2D risk, the direct effect of coffee consumption (*c'*) was significantly attenuated. This result might suggest that adiponectin could potentially constitute a complete mediator. However, as aforementioned, coffee is hypothesized to be involved in a more complex interplay of biomarkers,[47] which is in agreement with our mediation findings on CRP. Thus, a potential alternative explanation for the observed significant attenuation of coffee's direct effect, may be the influence of a different pathway or the presence of another mediator of similar strength but in the opposite direction.[48]

Moreover, since the analysis on adiponectin was performed in a smaller sample of participants with available data on specific biomarkers, power issues explaining the attenuation of coffee's direct effect cannot be ruled out. A previous Mendelian randomization study found no consistent evidence for a causal association between genetically lower levels of serum adiponectin and higher risk of T2D or higher fasting insulin.[49] Therefore, our mediation analysis result on adiponectin must be interpreted cautiously. Whether adiponectin might constitute a biomarker reflecting further underlying causal mechanism needs additional research. Adiponectin has been shown to have anti-inflammatory and insulin-sensitizing effects.[50] Lower concentrations of serum adiponectin have been associated with obesity and insulin sensitivity. Thus, adiponectin has been suggested as a therapeutic target for diabetes and obesity.[50] Available evidence on coffee intake and adiponectin is contradictory, with reports from clinical trials and population studies showing both non-significant results and significant positive trends.[21,47,51,52] The latter is supported by our findings, and we further suggest a dose-response significant association of coffee intake and higher adiponectin concentrations.

Furthermore, we showed a significant positive association between higher coffee consumption and plasma IL-18 in some categories of coffee consumption. Our results on plasma TNFRII and leptin disagree with a recent observational study reporting a significant trend between higher coffee consumption and lower plasma TNFRII and leptin.[47] Our findings on IL1ra were in agreement with clinical trials that found no changes in plasma concentrations after coffee intervention.[53,54]

Notably, we observed some disparities in the strength of the association of higher coffee consumption and lower T2D risk, and lower CRP concentrations among the different smoking statuses, as well as across cohorts. Never smokers had a stronger association between coffee consumption and T2D and CRP in both cohorts, as well as former smokers in RS. However, in UKB, the results among former and current smokers were inconsistent. These discrepancies might be explained by the differences in smoking habits between the two cohorts, as well as coffee consumption patterns. Smoking may play an important role increasing oxidative stress and consequent inflammatory response.[55,56] A study reported that current smokers with long smoking duration and greater daily smoking quantity had elevated oxidative stress markers and lower antioxidant levels than never and former smokers; but that this is reversible after a long period of smoking cessation.[56] Future studies on coffee and health outcomes should explore to what extent current and former smoking may interfere with coffee´s antioxidant effects.

Detrimental effects of coffee, like raising serum lipids, are attributed to coffee oils, which could be removed from the coffee extract by using a filter. A meta-analysis of clinical trials reported that the increment in total cholesterol was greater among patients in trials using unfiltered coffee as treatment.In our study, ground coffee, which is generally filtered, showed the strongest beneficial association with lower type 2 diabetes risk and CRP concentration.

Strengths of this study are the large sample size from two population-based cohort studies, UKB and the RS, and the longitudinal analyses of repeated measures of HOMA-IR in the latter. Moreover, we made use of a causal inference method, causal mediation analysis, including time-to-event data on incident T2D. An additional strength of our study is the comprehensive consideration of confounding factors, including sociodemographic, lifestyle, health-related and metabolic factors. Among several other confounding factors, our models were additionally adjusted for smoking status, alcohol consumption and diet quality. Furthermore, we carried out stratified analyses providing further insight into how the association between coffee consumption and T2D differed between sub-groups. There are also some limitations to our study that need consideration. The analyses on the subset of specific inflammatory markers and adipokines were limited in sample size, therefore conclusions from their analyses must be carefully considered. Additionally, our study did not assess differences between caffeinated and decaffeinated coffee. There is controversy regarding a potential effect of caffeine on inflammatory markers, as concluded by a recent systematic review of clinical trials.[20] Another limitation of our study is that we did not include adjustments for sugar intake with coffee. However, we did adjust for overall diet quality. Further, in the mediation analyses, we used as the mediator CRP concentrations which were measured at the same time point as the exposure coffee intake. However, use of a cross-sectional analysis will not have introduced reverse causation since circulating concentrations of CRP do not change coffee consumption behavior.

CONCLUSION

Our findings from two large population-based cohort studies highlight the association between coffee consumption and lower T2D risk as well as lower insulin resistance over time. Furthermore, we observe that higher coffee intake was associated with lower CRP concentrations. Finally, ours is the first study to demonstrate that changes in CRP related to coffee consumption may mediate part of the inverse association between coffee consumption and T2D.

SUPPLEMENTARY MATERIAL

Table S-1: Detailed baseline characteristics of the Rotterdam Study participants by coffee intake categories

	Non-coffee drinkers	0.5-2 cups/ day	3-4 cups/ day	5 or more cups/day	Total
n, (%)	542 (7.6)	1,594 (22.2)	2,983 (40.2)	2,053 (28.6)	7,172
Age, mean (SD), years	62.1 (9.8)	68.4 (10.1)	65.8 (9.2)	62.3 (7.8)	65.1 (9.4)
Sex, % women	373 (68.8)	1064 (66.8)	1844 (61.8)	1002 (48.8)	4283 (59.7)
Education level					
Primary	63 (11.6)	221 (13.9)	338 (11.3)	207 (10.1)	832 (11.6)
Intermediate	211 (38.9)	644 (40.4)	1271 (42.6)	816 (39.7)	2939 (41.0)
Higher general	160 (29.5)	455 (28.5)	879 (29.5)	595 (29.0)	2086 (29.1)
University	108 (19.9)	274 (17.2)	495 (16.6)	435 (21.2)	1315 (18.3)
Tea consumption, n (%)consumers	486 (89.7)	1465 (91.9)	2704 (90.6)	1589 (77.4)	6241 (87.0)
Alcohol consumption, median (IQR), grams/day	1.20 (0.00; 8.57)	3.57 (0.54; 10.25)	6.40 (0.71; 14.29)	6.43 (0.71; 15.0)	5.71 (0.54; 14.29)
Diet quality score, mean (SD)	7.1 (2.0)	6.9 (1.9)	6.8 (1.9)	6.3 (1.8)	6.7 (1.9)
Smoking status					
Never smokers	252 (46.50)	652 (40.9)	994 (33.3)	453 (22.1)	2351 (32.8)
Former smoker	209 (38.6)	730 (45.8)	1939 (46.8)	839 (40.9)	3174 (44.3)
Current smoker	81 (14.9)	212 (13.3)	593 (19.9)	761 (37.0)	1647 (22.9)
Physical activity, median (IQR), MET-hours/week	61.4 (31.2; 97.9)	70.8 (43.1; 103.8)	74.1 (46.6; 107.5)	70.5 (37.2; 104.6)	72.0 (42.0; 105.1)
BMI, mean (SD), kg/m2	26.8 (4.7)	26.7 (4.0)	26.9 (3.8)	27.1 (4.0)	26.9 (4.0)
Hypertension, n (%)	306 (56.5)	1046 (65.3)	1788 (60.0)	1099 (53.5)	4237 (59.1)
Total cholesterol, mean (SD), mmol/L	5.69 (1.04)	5.84 (1.03)	5.82 (0.97)	5.8 (0.97)	5.8 (0.99)
HDL cholesterol, median (IQR), mmol/L	1.37 (1.16; 1.68)	1.40 (1.16; 1.71)	1.40 (1.16; 1.67)	1.36 (1.13; 1.63)	1.38 (1.16; 1.67)
CRP median (IQR), ug/mL	1.40 (0.50; 3.30)	1.66 (0.70; 3.61)	1.50 (0.60; 3.30)	1.36 (0.57; 3.04)	1.5 (0.60; 3.30)
Incident type 2 diabetes cases, n (%)	52 (9.6)	171 (10.7)	276 (9.6)	187 (9.1)	685 (9.6)

Table S-2: Detailed baseline characteristics of the UK Biobank participants by coffee intake categories

	Non-coffee drinkers	0.5-2 cups/ day	3-4 cups/ day	5 or more cups/day	Total
n, (%)	31,773 (21.9)	66,153 (45.5)	30,440 (20.9)	17,004 (11.7)	145,370
Age, mean (SD), years	53.92 (8.09)	55.81 (8.10)	55.54 (8.03)	54.45 (7.94)	55.18 (8.10)
Sex, (%) women	19,818 (62.4)	39,088 (59.1)	16,715 (54.9)	8,722 (51.3)	84,343 (58.0)
Deprivation index	-1.93 [-3.52; 0.83]	-2.36 [-3.76; -0.04]	-2.40 [-3.79; -0.10]	-2.14 [-3.62; 0.48]	-2.26 [-3.70; 0.19]
Ethnicity, n (%)					
White	29,367 (92.4)	63,504 (96.0)	29,810 (97.9)	16,750 (98.5)	139,431 (95.9)
Mixed	508 (1.6)	855 (1.3)	261 (0.9)	138 (0.8)	1,762 (1.2)
South Asian	1,155 (3.6)	1,045 (1.6)	201 (0.7)	65 (0.4)	2,463 (1.7)
Black	561 (1.8)	557 (0.8)	142 (0.5)	46 (0.3)	1,306 (0.9)
Chinese	182 (0.6)	195 (0.3)	26 (0.1)	5 (0.0)	408 (0.3)
Tea consumption, mean (SD), cups	4.54 (3.29)	3.82 (2.54)	2.51 (2.22)	1.91 (2.89)	3.48 (2.84)
Alcohol consumption, n (%), frequency	3.37 (1.58)	2.80 (1.42)	2.67 (1.40)	2.90 (1.49)	2.91 (1.48)
Daily or almost daily	4,299 (13.5)	13,512 (20.4)	7,131 (23.4)	3,493 (20.5)	28,435 (19.6)
3-4 times a week	5,801 (18.3)	17,032 (25.7)	8,149 (26.8)	3,980 (23.4)	34,962 (24.1)
Once or twice per week	8,513 (26.8)	18,355 (27.7)	8,133 (26.7)	4,458 (26.2)	39,459 (27.1)
1-3 times a month	4,226 (13.3)	7,429 (11.2)	3,101 (10.2)	2,032 (12.0)	16,788 (11.5)
Special occasions only	4,736 (14.9)	6,212 (9.4)	2,467 (8.1)	1,848 (10.9)	15,263 (10.5)
Never	4,198 (13.2)	3,613 (5.5)	1,459 (4.8)	1,193 (7.0)	10,463 (7.2)
Diet quality score (SD)	5.11 (1.81)	5.25 (1.77)	5.06 (1.77)	4.76 (1.81)	5.12 (1.79)
Smoking status, n (%)					
Never smokers	19,659 (61.9)	39,535 (59.8)	16,624 (54.6)	7,559 (44.4)	83,377 (57.4)
Former smoker	9,077 (28.6)	21,335 (32.3)	10,325 (33.9)	5,711 (33.6)	46,448 (32.0)
Current smoker	3,037 (9.6)	5,283 (8.0)	3,491 (11.5)	3,734 (22.0)	15,545 (10.7)
Physical activity, median (IQR), MET-hours/week	32.0 (14.4; 65.1)	31.6 (14.9; 62.4)	31.1 (14.4; 62.2)	32.4 (14.9; 67.6)	31.6 (14.8; 63.1)
BMI, mean (SD), kg/m2	26.65 (4.52)	26.33 (4.17)	26.85 (4.25)	27.38 (4.47)	26.63 (4.31)
Hypertension, n (%)	130 (0.4)	218 (0.3)	92 (0.3)	56 (0.3)	495 (0.3)
Total cholesterol, mean (SD), mmol/L	5.75 (1.07)	5.89 (1.08)	5.93 (1.07)	5.93 (1.08)	5.87 (1.08)
HDL cholesterol, median (IQR), mmol/L	1.42 (1.19; 1.68)	1.47 (1.23; 1.75)	1.45 (1.22; 1.72)	1.40 (1.18; 1.65)	1.44 (1.21; 1.72)
CRP median (IQR), ug/mL	1.29 (0.63; 2.71)	1.15 (0.58; 2.34)	1.18 (0.59; 2.41)	1.30 (0.65; 2.64)	1.20 (0.60; 2.47)
Incident type 2 diabetes cases, n (%)	599 (1.9)	979 (1.5)	431 (1.4)	281 (1.7)	2290 (1.6)

Table S-3: Associations between coffee intake and incident type 2 diabetes excluding cases of incident type 2 diabetes during the first 2 years of follow- up.

	Coffee (cups/day)	**Rotterdam Study (n=7,076)**		**UK Biobank (n = 144,823)**	
		HR (95% CI)	P	HR (95% CI)	P
Model 1	Per cup increase	0.96 (0.91; 1.01)	0.089	0.96 (0.93; 0.984)	**0.002**
	0	ref		ref	
	0.5-2	0.96 (0.66; 1.39)	0.824	0.86 (0.76; 0.97)	**0.014**
	3-4	0.78 (0.55; 1.11)	0.168	0.81 (0.70; 0.94)	**0.005**
	≥ 5	0.76 (0.52; 1.03)	0.148	0.78 (0.78; 0.93)	**0.005**
	Trend		**0.028**		**0.002**
Model 2	Per cup increase	0.95 (0.90; 0.99)	**0.026**	0.94 (0.91; 0.96)	**<0.001**
	0	ref		ref	
	0.5-2	0.91 (0.63; 1.31)	0.608	0.88 (0.78; 0.99)	**0.032**
	3-4	0.73 (0.51; 1.04)	0.085	0.77 (0.66; 0.89)	**<0.001**
	≥ 5	0.70 (0.48; 1.02)	0.062	0.70 (0.59; 0.84)	**<0.001**
	Trend		**0.010**		**<0.001**

Model 1: adjusted for age, sex, Rotterdam Study cohort (RS only), level of education (RS only), deprivation index (UKBB only), ethnicity (UKBB only), tea consumption, alcohol consumption, smoking status, physical activity, and diet quality score; Model 2: additionally adjusted for BMI, hypertension, ratio serum total cholesterol/HDL.
Natural logarithm transformation was used for CRP (ug/mL). Cases of incident type 2 diabetes in the RS n=589 and in the UKBB n=547.
Bold text indicates statistically significant associations (p<0.05).

Table S-4: Interaction terms of coffee intake with covariates for the main analyses

		Rotterdam Study			**UK Biobank**	
		T2D	CRP[a]	HOMA-IR[a]	T2D	CRP[a]
Sex		0.999	0.222	**0.041**	0.650	**<0.001**
Age		0.981	**0.046**	0.774	0.991	0.161
Smoking						
	Former	0.982	0.604	0.958	0.076	0.853
	Current	0.305	**0.002**	0.602	0.914	**<0.001**
BMI		0.267	**0.015**	0.298	0.085	**<0.001**

Model adjusted for age, sex, Rotterdam Study cohort (RS only), level of education (RS only), deprivation index (UKBB only), ethnicity (UKBB only), tea consumption, alcohol consumption, smoking status, physical activity, and diet quality score, body mass index, hypertension, ratio serum total cholesterol/HDL; corresponding to model 2 in the main analyses.
[a]Natural logarithm transformed. Bold text indicates statistically significant associations (p<0.05).
Cases of incident type 2 diabetes among RS n=685 and in the UKBB n=2,290.
Abbreviations: BMI, body mass index; T2D, type 2 diabetes

Table S-5: Sex-stratified associations of coffee intake with type 2 diabetes incidence, and log transformed CRP and HOMA-IR in RS and UKBB

	Rotterdam Study				**The UK Biobank**			
	Women (n = 4,283)		**Men (n = 2,889)**		**Women (n = 84,343)**		**Men (n = 61,027)**	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
T2D	0.93 (0.88; 0.99)	**0.021**	0.95 (0.88; 1.02)	0.135	0.96 (0.93; 1.00)	**0.047**	0.96 (0.93; 0.99)	**0.005**
	β (95% CI)	P	β (95% CI)	P	β (95% CI)	P	β (95% CI)	P
CRP[a]	-0.016 (-0.027; -0.005)	**0.004**	-0.006 (-0.020; 0.007)	0.340	*-0.016 (-0.018; -0.013)*	***<0.001***	*-0.004 (-0.007; -0.002)*	***0.002***
HOMA-IR[a,b]	*-0.023 (-0.032; -0.014)*	***<0.001***	*-0.012 (-0.023; -0.001)*	***0.040***				

Model: adjusted for age, Rotterdam Study cohort (RS only), level of education (RS only), deprivation index (UKBB only), ethnicity (UKBB only), tea consumption, alcohol consumption, smoking status, physical activity, and diet quality score, body mass index, hypertension, ratio serum total cholesterol/HDL. Cases of incident type 2 diabetes among RS women n=405; RS men n=280; UKBB women n=958; UKBB men n=1,332.
[a]Natural logarithm transformed variables
[b]Sample size n=2,463 and n=1,675 men in RS.
Bold text indicates statistically significant associations (p<0.05).
Italic text indicates associations for which interaction terms between coffee consumption and sex were statistically significant (p<0.05).

Table S-6: Smoking-stratified associations of coffee intake with outcomes type 2 diabetes incidence, CRP and HOMA-IR in the Rotterdam Study

	Never smoker (n = 2,336)		**Former smoker (n = 3,161)**		**Current smoker (n = 1,637)**	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
T2D	0.91 (0.84; 0.99)	**0.024**	0.93 (0.87; 0.10)	**0.035**	1.00 (0.91; 1.08)	0.901
	β (95% CI)	P	β (95% CI)	P	β (95% CI)	P
CRP[a]	*-0.019 (-0.033; -0.005)*	***0.010***	-0.026 (-0.039; -0.012)	**<0.001**	*0.010 (-0.007; 0.027)*	*0.239*
HOMA-IR[a,b]	-0.013 (-0.026; -0.001)	**0.035**	-0.018 (-0.028; -0.007)	**0.001**	-0.021 (-0.037; -0.006)	**0.007**

Model: adjusted for age, sex, Rotterdam Study (RS) cohort level of education, tea consumption, alcohol consumption, physical activity, diet quality score, body mass index, hypertension, ratio serum total cholesterol/HDL. Cases of incident type 2 diabetes among RS never smokers n=206; RS former smokers n=295; RS current smokers n=178.
[a]Naturally logarithm transformed variables.
[b]Sample never smokers n=1,352; former smokers n=1,902; current smokers n=872.
Bold text indicates statistically significant associations (p<0.05).
Italic text indicates associations for which interaction terms between coffee consumption and smoking category were statistically significant (p<0.05).

Table S-7: Smoking-stratified associations of coffee intake with outcomes type 2 diabetes incidence and CRP in the UK Biobank

	Never smoker (n = 83,169)		**Former smoker (n = 46,294)**		**Current smoker (n = 15,495)**	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
T2D	0.90 (0.84; 0.96)	**0.001**	1.04 (0.97; 1.11)	0.284	0.90 (0.81; 0.10)	**0.046**
	β (95% CI)	P	β (95% CI)	P	β (95% CI)	P
CRP[a]	*-0.012 (-0.016; -0.009)*	***<0.001***	-0.011 (-0.016; -0.006)	**<0.001**	*0.001 (-0.008; 0.010)*	*0.821*

Model: adjusted for age, sex, deprivation index, ethnicity, tea consumption, alcohol consumption, physical activity, diet quality score, body mass index, hypertension, ratio serum total cholesterol/HDL.
[a]Natural logarithm transformation was used for CRP (ug/mL).
Cases of incident type 2 diabetes among UKBB never smokers n=1,113; UKBB former smokers n=792; UKBB current smokers n=375.
Bold text indicates statistically significant associations (p<0.05).
Italic text indicates associations for which interaction terms between coffee consumption and smoking category were statistically significant (p<0.05).

Table S-8: Body composition-stratified associations of coffee intake with outcomes T2D incidence, CRP and HOMA-IR in RS and UKB

	Rotterdam Study				**The UK Biobank**			
	Normal weight (BMI <25) (n = 2,425)		**Overweight/obese (BMI ≥25) (n = 4,699)**		**Normal weight (BMI <25) (n = 88,903)**		**Overweight/obese (BMI ≥25) (n = 56,072)**	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
T2D	0.84 (0.76; 0.93)	**0.001**	0.98 (0.93; 1.03)	0.369	0.86 (0.74; 1.00)	0.057	0.96 (0.92; 1.00)	0.084
	β (95% CI)	P	β (95% CI)	P	β (95% CI)	P	β (95% CI)	P
CRP[a]	*-0.003 (-0.018; 0.012)*	*0.706*	*-0.013 (-0.024; -0.003)*	***0.013***	*-0.005 (-0.009; -0.001)*	***0.021***	*-0.007 (-0.010; -0.003)*	***<0.001***
HOMA-IR[a,b]	-0.010 (-0.022; 0.003)	0.128	-0.014 (-0.024; -0.005)	**0.004**				

Model: adjusted for age, sex, Rotterdam Study cohort (RS only), level of education (RS only), deprivation index (UKBB only), ethnicity (UKB only), tea consumption, alcohol consumption, smoking status, physical activity, and diet quality score, hypertension, ratio serum total cholesterol/HDL.
Cases of incident type 2 diabetes among RS normal weight n=128; RS overweight/obese n=552; UKBB normal weight n=196; UKB
[a]Naturally logarithm transformed variables
[b]For HOMA-IR analysis the sample sizes were n=1,389 normal weight, and n=2,735.
overweight/obese n=2,083.
Bold text indicates statistically significant associations (p<0.05).
Italic text indicates associations for which interaction terms between coffee consumption and smoking category were statistically significant (p<0.05).

Table S-9: Associations of type 2 diabetes incidence and CRP, with specific types of coffee consumption in UKB

	UK Biobank					
	Decaffeinated coffee[a]		**Instant coffee**[b]		**Ground coffee**[c]	
	HR (95% CI)	P	HR (95% CI)	P	HR (95% CI)	P
Type 2 diabetes	0.97 (0.93; 1.01)	0.147	0.96 (0.93; 0.99)	**0.003**	0.88 (0.83; 0.93)	**<0.001**
	β (95% CI)	P	β (95% CI)	P	β (95% CI)	P
CRP[d]	-0.015 (-0.018; -0.012)	**<0.001**	-0.012 (-0.012; -0.008)	**<0.001**	-0.026 (-0.030; -0.023)	**<0.001**

Model: adjusted for age, sex, deprivation index, ethnicity, tea consumption, alcohol consumption, smoking status, physical activity, diet quality score, hypertension and ratio serum total cholesterol/HDL.
Sample of noncoffee drinkers n=31,773
[a] Total sample n=53,253 of whom n=21,480 were decaffeinated-coffee drinkers.
[b] Total sample n=95,119 of whom n=63,346 were instant-coffee drinkers.
[c] Total sample n=58,106 of whom n=26,333 were ground-coffee drinkers.
[c] Naturally logarithm transformed variable

REFERENCES

1 Bernard, K. Top 10 Coffee Consuming Nations, <https://www.worldatlas.com/articles/top-10-coffee-consuming-nations.html> (2018).
2 Smith, O. Mapped: The countries that drink the most coffee, <https://www.telegraph.co.uk/travel/maps-and-graphics/countries-that-drink-the-most-coffee/#comments> (2017).
3 Sung, K. C. et al. Inflammation in the Prediction of Type 2 Diabetes and Hypertension in Healthy Adults. Arch Med Res 48, 535-545, doi:10.1016/j.arcmed.2017.11.010 (2017).
4 Ludwig, I. A., Clifford, M. N., Lean, M. E., Ashihara, H. & Crozier, A. Coffee: biochemistry and potential impact on health. Food Funct 5, 1695-1717, doi:10.1039/c4fo00042k (2014).
5 van Dam, R. M. & Feskens, E. J. Coffee consumption and risk of type 2 diabetes mellitus. Lancet 360, 1477-1478, doi:10.1016/S0140-6736(02)11436-X (2002).
6 Ding, M., Bhupathiraju, S. N., Satija, A., van Dam, R. M. & Hu, F. B. Long-term coffee consumption and risk of cardiovascular disease: a systematic review and a dose-response meta-analysis of prospective cohort studies. Circulation 129, 643-659, doi:10.1161/CIRCULATIONAHA.113.005925 (2014).
7 Carlstrom, M. & Larsson, S. C. Coffee consumption and reduced risk of developing type 2 diabetes: a systematic review with meta-analysis. Nutr Rev 76, 395-417, doi:10.1093/nutrit/nuy014 (2018).
8 Poole, R. et al. Coffee consumption and health: umbrella review of meta-analyses of multiple health outcomes. BMJ 359, j5024, doi:10.1136/bmj.j5024 (2017).
9 Grosso, G., Godos, J., Galvano, F. & Giovannucci, E. L. Coffee, Caffeine, and Health Outcomes: An Umbrella Review. Annu Rev Nutr 37, 131-156, doi:10.1146/annurev-nutr-071816-064941 (2017).
10 Donath, M. Y., Dalmas, E., Sauter, N. S. & Boni-Schnetzler, M. Inflammation in obesity and diabetes: islet dysfunction and therapeutic opportunity. Cell Metab 17, 860-872, doi:10.1016/j.cmet.2013.05.001 (2013).
11 Donath, M. Y. & Shoelson, S. E. Type 2 diabetes as an inflammatory disease. Nat Rev Immunol 11, 98-107, doi:10.1038/nri2925 (2011).
12 Tsalamandris, S. et al. The Role of Inflammation in Diabetes: Current Concepts and Future Perspectives. Eur Cardiol 14, 50-59, doi:10.15420/ecr.2018.33.1 (2019).
13 Lopez-Garcia, E., van Dam, R. M., Qi, L. & Hu, F. B. Coffee consumption and markers of inflammation and endothelial dysfunction in healthy and diabetic women. Am J Clin Nutr 84, 888-893, doi:10.1093/ajcn/84.4.888 (2006).
14 Kanmani, S., Kwon, M., Shin, M. K. & Kim, M. K. Association of C-Reactive Protein with Risk of Developing Type 2 Diabetes Mellitus, and Role of Obesity and Hypertension: A Large Population-Based Korean Cohort Study. Sci Rep 9, 4573, doi:10.1038/s41598-019-40987-8 (2019).
15 Lainampetch, J. et al. Association of Tumor Necrosis Factor Alpha, Interleukin 6, and C-Reactive Protein with the Risk of Developing Type 2 Diabetes: A Retrospective Cohort Study of Rural Thais. J Diabetes Res 2019, 9051929, doi:10.1155/2019/9051929 (2019).

16 Liu, C. et al. Adiponectin, TNF-alpha and inflammatory cytokines and risk of type 2 diabetes: A systematic review and meta-analysis. Cytokine 86, 100-109, doi:10.1016/j.cyto.2016.06.028 (2016).
17 Odegaard, A. O. et al. Oxidative stress, inflammation, endothelial dysfunction and incidence of type 2 diabetes. Cardiovasc Diabetol 15, 51, doi:10.1186/s12933-016-0369-6 (2016).
18 Brahimaj, A. et al. Novel inflammatory markers for incident pre-diabetes and type 2 diabetes: the Rotterdam Study. Eur J Epidemiol 32, 217-226, doi:10.1007/s10654-017-0236-0 (2017).
19 Hjerkind, K. V., Stenehjem, J. S. & Nilsen, T. I. Adiposity, physical activity and risk of diabetes mellitus: prospective data from the population-based HUNT study, Norway. BMJ Open 7, e013142, doi:10.1136/bmjopen-2016-013142 (2017).
20 Paiva, C. et al. Consumption of coffee or caffeine and serum concentration of inflammatory markers: A systematic review. Crit Rev Food Sci Nutr 59, 652-663, doi:10.1080/10408398.2017.1386159 (2019).
21 Jacobs, S. et al. Evaluation of various biomarkers as potential mediators of the association between coffee consumption and incident type 2 diabetes in the EPIC-Potsdam Study. Am J Clin Nutr 100, 891-900, doi:10.3945/ajcn.113.080317 (2014).
22 Collins, R. What makes UK Biobank special? Lancet 379, 1173-1174, doi:10.1016/S0140-6736(12)60404-8 (2012).
23 Palmer, L. J. UK Biobank: bank on it. Lancet 369, 1980-1982, doi:10.1016/S0140-6736(07)60924-6 (2007).
24 Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 12, e1001779, doi:10.1371/journal.pmed.1001779 PMEDICINE-D-12-02351 (2015).
25 Ikram, M. A. et al. Objectives, design and main findings until 2020 from the Rotterdam Study. Eur J Epidemiol 35, 483-517, doi:10.1007/s10654-020-00640-5 (2020).
26 Bradbury, K. E., Young, H. J., Guo, W. & Key, T. J. Dietary assessment in UK Biobank: an evaluation of the performance of the touchscreen dietary questionnaire. J Nutr Sci 7, e6, doi:10.1017/jns.2017.6600066 (2018).
27 Liu, B. 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 14, 1998-2005, doi:10.1017/S1368980011000942 (2011).
28 Ligthart, S. et al. Lifetime risk of developing impaired glucose metabolism and eventual progression from prediabetes to type 2 diabetes: a prospective cohort study. Lancet Diabetes Endocrinol 4, 44-51, doi:10.1016/S2213-8587(15)00362-9 (2016).
29 Neeley, W. E. Simple automated determination of serum or plasma glucose by a hexokinase-glucose-6 -phosphate dehydrogenase method. Clin Chem 18, 509-515 (1972).
30 Mackenbach, J. P. Health and Deprivation - Inequality and the North - Townsend,P, Phillimore,P, Beattie,A. Health Policy 10, 207-207, doi 10.1016/0168-8510(88)90006-1 (1988).
31 Centre, U. B. C. UK Biobank: Protocol for a large-scale prospective epidemiological resource <https://www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf>.

32 Craig, C. L. et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 35, 1381-1395, doi:10.1249/01.MSS.0000078924.61453.FB (2003).
33 Caspersen, C. J., Bloemberg, B. P., Saris, W. H., Merritt, R. K. & Kromhout, D. The prevalence of selected physical activities and their relation with coronary heart disease risk factors in elderly men: the Zutphen Study, 1985. Am J Epidemiol 133, 1078-1092, doi:10.1093/oxfordjournals.aje.a115821 (1991).
34 Stel, V. S. et al. Comparison of the LASA Physical Activity Questionnaire with a 7-day diary and pedometer. J Clin Epidemiol 57, 252-258, doi:10.1016/j.jclinepi.2003.07.008 (2004).
35 Cobiac, L. J., Scarborough, P., Kaur, A. & Rayner, M. The Eatwell Guide: Modelling the Health Implications of Incorporating New Sugar and Fibre Guidelines. PLoS One 11, e0167859, doi:10.1371/journal.pone.0167859 PONE-D-16-35010 (2016).
36 EFSA Panel on Dietetic Products N, A. Scientific Opinion on establishing Food-Based Dietary Guidelines, <https://efsa.onlinelibrary.wiley.com/doi/pdf/10.2903/j.efsa.2010.1460> (
37 Waijers, P. M., Feskens, E. J. & Ocke, M. C. A critical review of predefined diet quality scores. Br J Nutr 97, 219-231, doi:10.1017/S0007114507250421 (2007).
38 Voortman, T. et al. Adherence to the 2015 Dutch dietary guidelines and risk of non-communicable diseases and mortality in the Rotterdam Study. Eur J Epidemiol 32, 993-1005, doi:10.1007/s10654-017-0295-2 (2017).
39 Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. mediation: R Package for Causal Mediation Analysis. J Stat Softw 59 (2014).
40 Govindarajulu, U. S., Malloy, E. J., Ganguli, B., Spiegelman, D. & Eisen, E. A. The comparison of alternative smoothing methods for fitting non-linear exposure-response relationships with Cox models in a simulation study. Int J Biostat 5, Article 2, doi:10.2202/1557-4679.1104 (2009).
41 Tunnicliffe, J. M. & Shearer, J. Coffee, glucose homeostasis, and insulin resistance: physiological mechanisms and mediators. Appl Physiol Nutr Metab 33, 1290-1300, doi:10.1139/H08-123 (2008).
42 Ding, M., Bhupathiraju, S. N., Chen, M., van Dam, R. M. & Hu, F. B. Caffeinated and decaffeinated coffee consumption and risk of type 2 diabetes: a systematic review and a dose-response meta-analysis. Diabetes Care 37, 569-586, doi:10.2337/dc13-1203 (2014).
43 Pollack, R. M., Donath, M. Y., LeRoith, D. & Leibowitz, G. Anti-inflammatory Agents in the Treatment of Diabetes and Its Vascular Complications. Diabetes Care 39 Suppl 2, S244-252, doi:10.2337/dcS15-3015 (2016).
44 Cheng, L. et al. Exposing the Causal Effect of C-Reactive Protein on the Risk of Type 2 Diabetes Mellitus: A Mendelian Randomization Study. Front Genet 9, 657, doi:10.3389/fgene.2018.00657 (2018).
45 Agrawal, A., Cha-Molstad, H., Samols, D. & Kushner, I. Transactivation of C-reactive protein by IL-6 requires synergistic interaction of CCAAT/enhancer binding protein beta (C/EBP beta) and Rel p50. J Immunol 166, 2378-2384, doi:10.4049/jimmunol.166.4.2378 (2001).
46 Agrawal, A., Cha-Molstad, H., Samols, D. & Kushner, I. Overexpressed nuclear factor-kappaB can participate in endogenous C-reactive protein induction, and enhances the effects of C/

EBPbeta and signal transducer and activator of transcription-3. Immunology 108, 539-547, doi:.1046/j.1365-2567.2003.01608.x (2003).

47 Hang, D. et al. Coffee consumption and plasma biomarkers of metabolic and inflammatory pathways in US health professionals. Am J Clin Nutr 109, 635-647, doi:10.1093/ajcn/nqy295 (2019).

48 Zhang, Z. et al. Causal mediation analysis in the context of clinical research. Ann Transl Med 4, 425, doi:10.21037/atm.2016.11.11 (2016).

49 Yaghootkar, H. et al. Mendelian randomization studies do not support a causal role for reduced circulating adiponectin levels in insulin resistance and type 2 diabetes. Diabetes 62, 3589-3598, doi:10.2337/db13-0128 (2013).

50 Achari, A. E. & Jain, S. K. Adiponectin, a Therapeutic Target for Obesity, Diabetes, and Endothelial Dysfunction. Int J Mol Sci 18, doi:10.3390/ijms18061321 (2017).

51 Gavrieli, A. et al. Caffeinated coffee does not acutely affect energy intake, appetite, or inflammation but prevents serum cortisol concentrations from falling in healthy men. J Nutr 141, 703-707, doi:10.3945/jn.110.137323 (2011).

52 Wedick, N. M. et al. Effects of caffeinated and decaffeinated coffee on biological risk factors for type 2 diabetes: a randomized controlled trial. Nutr J 10, 93, doi:10.1186/1475-2891-10-93 (2011).

53 Kempf, K. et al. Cardiometabolic effects of two coffee blends differing in content for major constituents in overweight adults: a randomized controlled trial. Eur J Nutr 54, 845-854, doi:10.1007/s00394-014-0763-3 (2015).

54 Ramakers, B. P. et al. Circulating adenosine increases during human experimental endotoxemia but blockade of its receptor does not influence the immune response and subsequent organ injury. Crit Care 15, R3, doi:10.1186/cc9400 (2011).

55 Foronjy, R. & D'Armiento, J. The Effect of Cigarette Smoke-derived Oxidants on the Inflammatory Response of the Lung. Clin Appl Immunol Rev 6, 53-72, doi:10.1016/j.cair.2006.04.002 (2006).

56 Zhou, J. F. et al. Effects of cigarette smoking and smoking cessation on plasma constituents and enzyme activities related to oxidative stress. Biomed Environ Sci 13, 44-55 (2000).

Chapter 5.

Type 2 diabetes, sex differences and risk of cardiovascular disease and mortality

Chapter 5.1.

Serum uric acid and risk of fatal and nonfatal cardiovascular outcomes and all cause-mortality: the role of sex and type 2 diabetes

Carolina Ochoa-Rosales, Niels van der Schaft, Frederick K. Ho, Jill P Pell, M. Arfan Ikram, Carlos A. Celis-Morales*, Trudy Voortman*, *Manuscript*

ABSTRACT

Background. Evidence indicates that higher serum uric acid (SUA), is a risk factor for cardiovascular disease (CVD) and premature mortality. SUA concentrations are known to differ by sex and also affect type 2 diabetes (T2D) risk. However, few studies have reported whether T2D status and sex modify the association of SUA with fatal and nonfatal CVD-related events and all-cause mortality.

Methods. We included 370,355 participants from the UK Biobank prospective cohort in the current study. Cox proportional hazard models were used to assess associations of SUA levels at baseline with fatal and non-fatal cardiovascular disease (CVD), heart failure (HF), myocardial infarction (MI) and stroke, as well as all-cause mortality. For all these outcomes, we tested interaction of SUA with sex and T2D status.

Results. During a median follow-up of 6.96 years, 11,356 all-cause and 3,274 CVD deaths, and 23,286 nonfatal CVD events occurred. Higher SUA concentrations were associated with higher risk of CVD-related fatal and nonfatal outcomes and all-cause mortality (HR from 1.06 per 1 mg/dL higher SUA [1.04; 1.08] for all-cause mortality, to HR=1.32 [1.25; 1.38] for non-fatal HF), except for MI, where no association was found. Most of the associations followed a J-shape. SUA-related higher risk of most fatal outcomes tended to be stronger among women than men and stronger among diabetics as compared to non-diabetics (p-interaction <0.05), with the strongest association among diabetic women. For nonfatal outcomes, SUA-related higher risk of CVD events also tended to be stronger among both diabetic women and men as compared to nondiabetics.

Conclusion. Using a large population-based study, we confirmed that higher SUA levels are directly associated with a higher risk of occurrence of the majority of CVD-related fatal and nonfatal events, and all-cause mortality. These associations generally followed a J-shaped curve. T2D status and sex may modify some of these associations, with more detrimental associations among women or diabetic participants.

INTRODUCTION

Many epidemiological studies, have implied that hyperuricemia, a state with high levels of serum uric acid (SUA), is a risk factor for all-cause mortality and cardiovascular disease (CVD) events,[1-6] the last being the leading cause of death worldwide for both sexes.[7] However, the epidemiology, pathophysiology and outcomes of CVD-related events may significantly differ between men and women.[8,9] SUA levels are modulated by endogenous purine metabolism, purine intake from diet, urinary urate excretion and intestinal uricolysis.[10] SUA levels vary according to age and sex, with the highest levels generally found among adult men.[10] In clinical practice, guidelines differ in regard to the recommended SUA concentration cut-off for hyperuricemia diagnosis. While some of them recommend a cut off of 7.0 mg/dL (or 420 μmol/L) in the general population, other guidance documents suggest sex-specific thresholds to diagnose hyperuricemia: 6.0 mg/dL (or 360 μmol/L) in women and 7.0 mg/dL in men,[11] based on the fact that SUA concentrations tend to be lower among women.[12]

In addition, a major risk factor for CVD is type 2 diabetes (T2D); it confers about two-fold higher risk for vascular diseases, independently from other conventional risk factors.[13] Studies have shown that higher SUA levels are positively associated with incident T2D or pre-T2D,[14-19] which may affect men and women differently. [15] For example, a recent study reported significantly higher SUA-related risk of T2D among pre-diabetic men but not women.[15] Nevertheless, other studies reported no differences according to sex.[18]

Few studies on SUA and risk of CVD and all-cause mortality have investigated potential sex disparities or differential risk among T2D patients.[20-22] The majority of them showed some limitations, such as the use of a specific population (e.g., only diabetic patients, or only among men, or among participants with specific comorbidities), instead of a population-based approach. Further, conclusions from those studies are inconsistent, showing contradictory findings for men and women across outcomes and across studies,[4,21-26] as well as high heterogeneity, according to a recent meta-analysis.[27] This conflicting scenario, prevents us to conclude whether sex and T2D status act as effect modifiers of the association between SUA and health outcomes. This gap may lead to underestimation of the effect of SUA on cardiovascular health in men or women or according to other existing comorbidities such as T2D, and therefore, to the accurate identification of risk groups. Research on how both sex and T2D combined modify the association of SUA with CVD-related outcomes and all-cause mortality is limited and inconclusive. The lack of large population

samples for within-study comparisons of diabetic and healthy men and women, is a significant drawback in the current evidence. A better understanding on how SUA levels represent a risk for CVD events in different subpopulations (according to sex and T2D status) from population-based studies may help in explaining inconsistent results from previous studies and in tailoring prevention strategies against mortality and morbidity from CVD related diseases, e.g. the extent to which the use of a SUA-lowering intervention could be of benefit to reduce future events of stroke or heart failure in diabetic or nondiabetic women or men.

Therefore, we aimed to firstly investigate the association of SUA with fatal and nonfatal CVD incidence and all-cause mortality; and secondly, to study to what extent sex and T2D status modify the association between SUA and these health outcomes, using data from the UK Biobank, a large prospective cohort study.

METHODS

Study design

The study was embedded in the UK Biobank, a large prospective population-based cohort study that recruited 502,536 women and men aged 37 to 73 years, between April 2007 and December 2010, at 22 research centres across England, Scotland and Wales.[28-30] Participants completed a self-administered, touch-screen questionnaire, a nurse-led face-to-face interview and provided biological samples at the baseline assessment visit.

Exposure ascertainment

Participants´ blood samples were collected at the research centre according to standard procedures described in the online protocol.[31] Serum uric acid (SUA) level was assessed using the uricase PAP method (AU5800 chemistry analyser, Beckman Coulter).

Outcomes assessment

For the fatal events, date of death was obtained from death certificates held within the National Health Service (NHS) Information Centre (England and Wales) and the NHS Central Register Scotland (Scotland). As for nonfatal events, date and specific cause of hospital admission were obtained from hospital records linked to Health Episode Statistics (England and Wales) and Scottish Morbidity Records 01 (Scotland). Detailed information on the linkage procedure can be found at http://content.digital.nhs.uk/services.

End of follow-up was recorded as the date of death or the date of end of follow-up for the assessment centre attended (31st January 2018 in Wales or England and 30th November 2016 in Scotland), or the first date of hospitalisation for the outcomes, whichever came first.

The outcomes in the current study included fatal and nonfatal events of cardiovascular disease (CVD), ischemic coronary heart disease (IHD), heart failure (HF), myocardial infarction (MI) and stroke, as well as deaths from any cause. The outcome events were defined as a hospital admission or death with code according to the International Classification of Diseases, 10th Revision (ICD-10),[32] as follows: MI: I21, I21.4, I21.9; HF: I50.0, I50.2, I50.9; stroke: ICD I60, I61, I63 and I64; and CVD: I05-I89.

Covariates assessment

The confounder variables considered were sociodemographic factors (age, sex, ethnicity, deprivation index), lifestyle factors at baseline (smoking behaviour, fruits and vegetables consumption, oily fish intake, red meat intake, processed meat intake, alcohol consumption, sedentary and sleeping time), cardiovascular risk factors at baseline (systolic blood pressure, blood total cholesterol (total-C), blood high-density lipoprotein cholesterol (HDL-C), body mass index (BMI) and type 2 diabetes status). Further details of the measurement procedures can be found elsewhere.[31] Briefly, ethnicity was self-reported at recruitment and classified into five categories (white, mixed, south Asian, black, Chinese). Townsend deprivation index scores were used as a measure of socioeconomic status, and derived from national census data about car ownership, household overcrowding, owner occupation, and unemployment aggregated for postcodes of residence.[33] Data on food consumption were collected through food-frequency questionnaires (FFQs) administrated at the research centre and via internet. The FFQ consisted of twenty-nine questions about diet and eighteen about alcohol consumption over the past year. Details on the questionnaires' questions[34] and validation of the web-based questionnaire[35] can be found elsewhere. Smoking behaviour was categorized into never, former, or current smoking. Data on sleeping time and sedentary time, such as time spent TV viewing (hours/day) were collected through a questionnaire, as described elsewhere.[36]

Physical measures and collection of blood samples were assessed at the research centre during the participant´s visit, using standardised procedures. Body mass index (BMI) was calculated as weight (kg) divided by height (metres) squared, and then categorised according to the WHO criteria.[37] Systolic blood pressure (SBP) in mmHg was the result of the mean of two measurements. Blood total cholesterol (total-C) and high-density lipoprotein cholesterol (HDL-C) were measured in mmol/L; blood measured in mmol/L

using the cholesterol oxidase-peroxidase (CHO-POD) enzymatic reaction and enzyme immunoinhibition methods respectively (AU5800 chemistry analyser, Beckman Coulter). Disease history information for prevalent type 2 diabetes (T2D), gout and kidney disease at baseline were collected by self-administrated questionnaires, and they referred to diseases diagnosed by a physician. Further details on the measurements protocols can be found in the UK Biobank online protocol.[31]

Study population

Participants with missing data on SUA (n=33,479) and prevalent T2D (n=4,294) were excluded, as well as participants with baseline diagnose of cardiovascular disease (n=16,857). Additionally, individuals under conditions that may alter SUA concentrations at baseline were also excluded. They correspond to subjects diagnosed with gout (n=6,546), who may be under treatment to lower SUA levels; and participants with kidney disease (n=1,169), whose SUA levels may be affected by impaired kidney function, when kidney does not eliminate uric acid efficiently. Further, subjects with missing data on relevant confounders were excluded (n=71,311). A remaining sample of 370,355 participants was available for analyses. Figure 1 shows details on study sample selection.

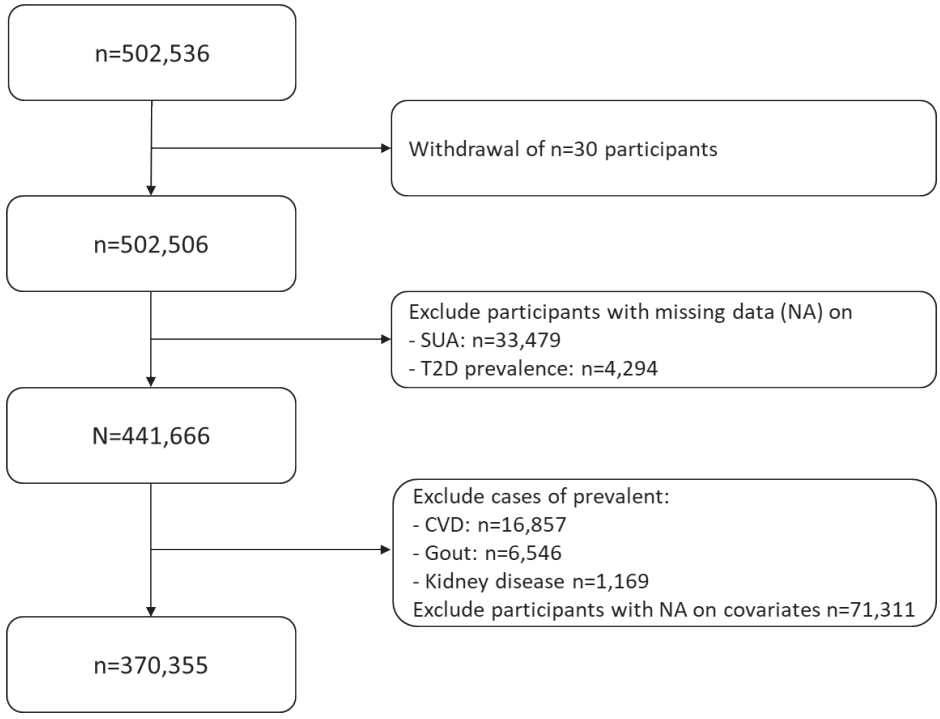


Figure 1: Flow chart of study sample selection

Statistical analyses

Cox proportional hazard models with fitted penalised cubic splines were used to assess nonlinear associations between SUA levels at baseline and fatal and nonfatal outcomes. The statistical model considered the following sociodemographic, lifestyle and health-related factors as confounders: age, sex, ethnicity, deprivation index, smoking, fruits and vegetables consumption, oily fish intake, red meat intake, processed meat intake, alcohol consumption, sedentary time, sleeping time, T2D diagnose, SBP, serum total cholesterol, serum HDL-C, BMI and type 2 diabetes status. Interaction effects between SUA levels and sex or T2D diabetes were investigated by fitting interaction terms. Evidence for interaction was found when $P<0.05$, after which analyses were stratified by sex and/or T2D. The results were reported as Hazard Ratio (HR) and the 95% confidence intervals (CIs). All analyses were performed using R Statistical Software V. 3.6.1.

Ethics

The UK Biobank study was approved by the North West Multi-Centre Research Ethics Committee; participants provided written informed consent for data collection, analysis, and record linkage. This study is part of UK Biobank project 7155 (NHS National Research Ethics Service 16/NW/0274).

RESULTS

Baseline participants' characteristics by sex and diabetes status are shown in Tables 1 and 2, and further details in Supplementary Tables S-1 and S-2. Participants included in this study were aged 56.4 (SD=8.1) years on average, and 55.6% were women, 55.5% were never smokers and 66.0% were overweight or obese at baseline. Mean SUA concentration was 5.16 (SD=1.32) µm/dL (range 1.50 – 14.7 mg/dL), which were higher among men than women (4.54 (1.09) vs 5.93 (1.16) mg/dL, $P<0.001$); and higher among diabetic participants as compared with nondiabetics (5.59 (1.39) vs 5.14 (1.31) µm/dL, $P<0.001$). After a median follow-up time of 6.96 years, a total of 11,356 deaths from any cause and 3,274 CVD deaths occurred, as well as 23,286 nonfatal CVD events. Details on number of events per group according to sex and T2D are shown in Tables 1 and 2.

In the overall sample and using a model adjusted for sociodemographic, lifestyle and health-related factors, we found significant associations between higher SUA concentration and higher risk of dying from any cause or from cause-specific CVD, except for MI (linear HR ranging from 1.06 [95% CI 1.04; 1.08] per 1 mg/dL of SUA

Table 1: Summary of participants´ baseline characteristics by sex or diabetes status

	Total sample	Women	Men	Nondiabetics	Diabetics
n	370,355	205,917	164,438	355,442	14,913
Age, years; mean (SD)	56.36 (8.09)	56.31 (7.99)	56.41 (8.22)	56.21 (8.10)	59.92 (6.87)
Sex, women; n (%)	205917 (55.6)	-	-	200168 (56.3)	5749 (38.6)
Deprivation index; median [IQR]	-2.21 [-3.67, 0.37]	-2.20 [-3.66, 0.33]	-2.22 [-3.69, 0.41]	-2.24 [-3.69, 0.31]	-1.42 [-3.25, 1.83]
Smoking, n (%)					
Never	205615 (55.5)	123186 (59.8)	82429 (50.1)	198707 (55.9)	6908 (46.3)
Former	126714 (34.2)	64833 (31.5)	61881 (37.6)	120308 (33.8)	6406 (43.0)
Current	38026 (10.3)	17898 (8.7)	20128 (12.2)	36427 (10.2)	1599 (10.7)
Body mass index category (BMI), n (%)					
Underweight (BMI < 18.5)	1949 (0.5)	1551 (0.8)	398 (0.2)	1931 (0.5)	18 (0.1)
Normal (BMI 18.5 to 24.9)	124247 (33.5)	81505 (39.6)	42742 (26.0)	122650 (34.5)	1597 (10.7)
Overweight (BMI 25.0 to 29.9)	157663 (42.6)	75656 (36.7)	82007 (49.9)	152460 (42.9)	5203 (34.9)
Obese (BMI ≥30.0)	86496 (23.4)	47205 (22.9)	39291 (23.9)	78401 (22.1)	8095 (54.3)
Systolic blood pressure, mmHg; mean (SD)	137.77 (18.62)	135.20 (19.20)	140.98 (17.33)	137.60 (18.66)	141.67 (17.03)
HDL-C, mmol/dL; mean (SD)	1.46 (0.38)	1.60 (0.38)	1.29 (0.31)	1.47 (0.38)	1.20 (0.32)
Total cholesterol, mmol/dL; mean (SD)	5.75 (1.12)	5.90 (1.11)	5.56 (1.10)	5.80 (1.10)	4.49 (1.01)
Uric acid, mg/dL; mean (SD)	5.16 (1.32)	4.54 (1.09)	5.93 (1.16)	5.14 (1.31)	5.59 (1.39)
Type 2 diabetes cases; n (%)	14913 (4.0)	5749 (2.8)	9164 (5.6)	-	-

Table 2: Summary of participants´ baseline characteristics by sex and diabetes status

	Nondiabetic women	Diabetic women	Nondiabetic men	Diabetic men
Sample n	200168	5749	155274	9164
Age, years; mean (SD)	56.21 (7.99)	59.93 (6.77)	56.20 (8.24)	59.92 (6.94)
Deprivation index; median [IQR]	-2.22 [-3.67, 0.28]	-1.24 [-3.13, 1.93]	-2.25 [-3.70, 0.33]	-1.53 [-3.34, 1.76]
Smoking, n (%)				
Never	119894 (59.9)	3292 (57.3)	78813 (50.8)	3616 (39.5)
Former	62868 (31.4)	1965 (34.2)	57440 (37.0)	4441 (48.5)
Current	17406 (8.7)	492 (8.6)	19021 (12.2)	1107 (12.1)
Never	17306 (8.6)	1247 (21.7)	8643 (5.6)	1066 (11.6)
Body mass index category, n (%)				
Underweight	1542 (0.8)	9 (0.2)	389 (0.3)	9 (0.1)
Normal	80891 (40.4)	614 (10.7)	41759 (26.9)	983 (10.7)
Overweight	74057 (37.0)	1599 (27.8)	78403 (50.5)	3604 (39.3)
Obese	43678 (21.8)	3527 (61.3)	34723 (22.4)	4568 (49.8)
Systolic blood pressure, mmHg; mean (SD)	135.06 (19.23)	140.17 (17.52)	140.88 (17.37)	142.61 (16.64)
HDL-C, mmol/dL; mean (SD)	1.61 (0.37)	1.32 (0.34)	1.30 (0.31)	1.12 (0.28)
Total cholesterol, mmol/dL; mean (SD)	5.93 (1.10)	4.74 (1.01)	5.63 (1.07)	4.34 (0.97)
Uric acid, mg/dL; mean (SD)	4.52 (1.08)	5.26 (1.39)	5.94 (1.15)	5.79 (1.35)

higher concentration for all-cause mortality, to HR=1.31 [1.23; 1.40] for fatal HF; Table 3). Subsequent nonlinear analysis (P for nonlinearity ≤0.001) in the total population (Figures 2-A and 2-B) showed J-shaped associations. SUA concentrations below approximately 6 mg/dL did not seem to affect the risk of death from any cause or CVD event; whereas values above 6 mg/dL were associated with increased risk of both cause-specific and all-cause mortality. Similar results were observed for risk of fatal HF and stroke events (Table 3 and Supplementary Figures S-1-A and S-1-B).

These associations were generally stronger among women than among men (e.g. HR=1.23 [1.17, 1.29] and HR=1.10 [1.07, 1.14] for risk of CVD mortality, respectively; P for sex interaction <0.0001); and stronger among diabetics as compared to non-diabetics (e.g. HR=1.23 [1.16, 1.31] and HR=1.12 [1.09, 1.16] for risk of CVD mortality, respectively; P for diabetes interaction=0.020), (Table 3). These differences were also observed in the nonlinear analyses, following similar J-shapes as observed in the total population (Figure 2, Supplementary Figures S-1 and S-2). For example, among women, and based on visual inspection of the graphs, SUA concentrations lower

Table 3. Associations of serum uric acid with mortality and incident health outcomes by sex or type 2 diabetes

	Total sample (n = 370,355)		Women (n = 205917)		Men (n = 164438)		P-sex inter-action	Nondiabetics (n = 355442)		Diabetics (n = 14913)		P-diabetes inter-action
	HR [95% CI]	Nr. of events	HR [95% CI]	Nr. of events	HR [95% CI]	Nr. of events		HR [95% CI]	Nr. of events	HR [95% CI]	Nr. of events	
Fatal outcomes												
All-cause	**1.06 [1.04, 1.08]**	11356	**1.11 [1.08, 1.14]**	4928	**1.03 [1.01, 1.05]**	6428	**7.22E-05**	**1.05 [1.03, 1.07]**	10287	**1.14 [1.09, 1.19]**	1069	**0.001**
CVD	**1.15 [1.11, 1.18]**	3274	**1.23 [1.17, 1.29]**	1102	**1.10 [1.07, 1.14]**	2172	**3.80E-04**	**1.12 [1.09, 1.16]**	2772	**1.23 [1.16, 1.31]**	502	**0.020**
HF	**1.31 [1.23, 1.40]**	567	**1.44 [1.29, 1.60]**	182	**1.23 [1.13, 1.34]**	385	**0.009**	**1.32 [1.23, 1.43]**	459	**1.25 [1.10, 1.42]**	108	0.257
MI	1.01 [0.93, 1.09]	482	0.93 [0.78, 1.10]	117	1.03 [0.94, 1.12]	365	0.236	1.03 [0.94, 1.13]	401	0.95 [0.80, 1.12]	81	0.408
Stroke	**1.12 [1.04, 1.20]**	532	**1.18 [1.06, 1.32]**	262	1.06 [0.96, 1.17]	270	0.204	1.07 [0.98, 1.16]	461	**1.32 [1.13, 1.54]**	71	**0.049**
Nonfatal outcomes												
CVD	**1.06 [1.05, 1.08]**	23286	**1.07 [1.05, 1.09]**	9626	**1.06 [1.04, 1.07]**	13660	0.073	**1.06 [1.05, 1.07]**	21450	**1.10 [1.07, 1.14]**	1836	**0.009**
HF	**1.32 [1.25, 1.38]**	884	**1.34 [1.24, 1.46]**	354	**1.30 [1.22, 1.39]**	530	0.665	**1.29 [1.20, 1.37]**	683	**1.38 [1.27, 1.51]**	201	0.141
MI	1.01 [0.98, 1.05]	3070	1.01 [0.95, 1.07]	874	1.01 [0.97, 1.05]	2196	0.952	1.01 [0.98, 1.04]	2819	1.02 [0.93, 1.12]	251	0.458
Stroke	**1.07 [1.03, 1.11]**	2667	**1.06 [1.01, 1.12]**	1166	**1.07 [1.03, 1.12]**	1501	0.859	**1.06 [1.02, 1.10]**	2415	1.02 [0.93, 1.12]	252	0.282

Model adjusted by age, sex, ethnicity, deprivation index, gout diagnosis, comorbidity, smoking, fruits and vegetables consumption, oily fish intake, red meat intake, processed meat intake, alcohol intake, sedentary time, sleeping time, body mass index category, systolic blood pressure, blood total cholesterol, blood high-density lipoprotein cholesterol.

Abbreviations: Cardiovascular disease, CVD; Heart failure, HF; Myocardial infarction, MI.

Bold text indicates statistical significance P >0.05

than approximately 4-5 mg/dL did not seem to be associated to all-cause mortality. However, concentrations above 4-5 mg/dL started to be detrimental for women, showing a higher risk of fatal CVD events. In men, the harmful effect of SUA on higher risk of fatal CVD events were observed in concentration ranges of approximately 4 mg/dl and lower, and then from approximately 6-7 mg/dL onwards, with a U-shaped curve (Figure 2). Associations with stroke were not significant among men or non-diabetics (Table 3).

In line with the findings for total or CVD mortality, significant associations were found between higher SUA concentrations and higher risk of all nonfatal events studied, except for MI (linear HR ranging from 1.06 [95% CI 1.05; 1.08] per 1 mg/dL of SUA higher concentration in a linear model for nonfatal CVD, to HR=1.32 [1.25; 1.38] for nonfatal HF; Table 3). However, for the non-fatal CVD-related events, most associations did not differ according to sex or T2D status (P for interaction >0.05), except for nonfatal CVD among diabetics and nondiabetics (P for diabetes interaction =0.009). Association with stroke was not significant among diabetics. Also, for nonfatal outcomes, most associations observed followed a J-shape. SUA concentrations above 4 mg/dL showed adverse effects among women, increasing the risk nonfatal CVD and HF events; whereas the detrimental effect in men was observed at SUA concentrations of 6-7 mg/dL and over. However, in levels around 8-9 mg/dL and above the curves of women and men overlapped and sex differences were not observed anymore in our study population (P for sex interaction >0.05; Supplementary Figure S-2).

Further stratification was carried out to compare groups of nondiabetic women, diabetic women, nondiabetic men and diabetic men (Table 4, Figure 2 and Supplementary Figures S-1 and S-2). In relation to higher SUA levels, we observed that diabetic women tended to have a strongest risk of CVD mortality (linear HR=1.35 [1.20; 1.51]) and death from any cause (HR=1.20 [1.11; 1.30]).

When comparing women and men, we found that T2D status modified SUA-related risk for some mortality outcomes. Men with diabetes showed higher risks of all-cause, CVD and stroke mortality of similar strength than those among nondiabetic women (Table 4). Most of these associations also followed a J-shaped curve, similar to those described above (Figure 2 and Supplementary Figure S-1); with exception of SUA-related risk of all-cause mortality among nondiabetic men, who presented a U-shaped curve, as previously described for overall male population. Diabetes status did not seem to modify the SUA-related higher risk of HF mortality among women.

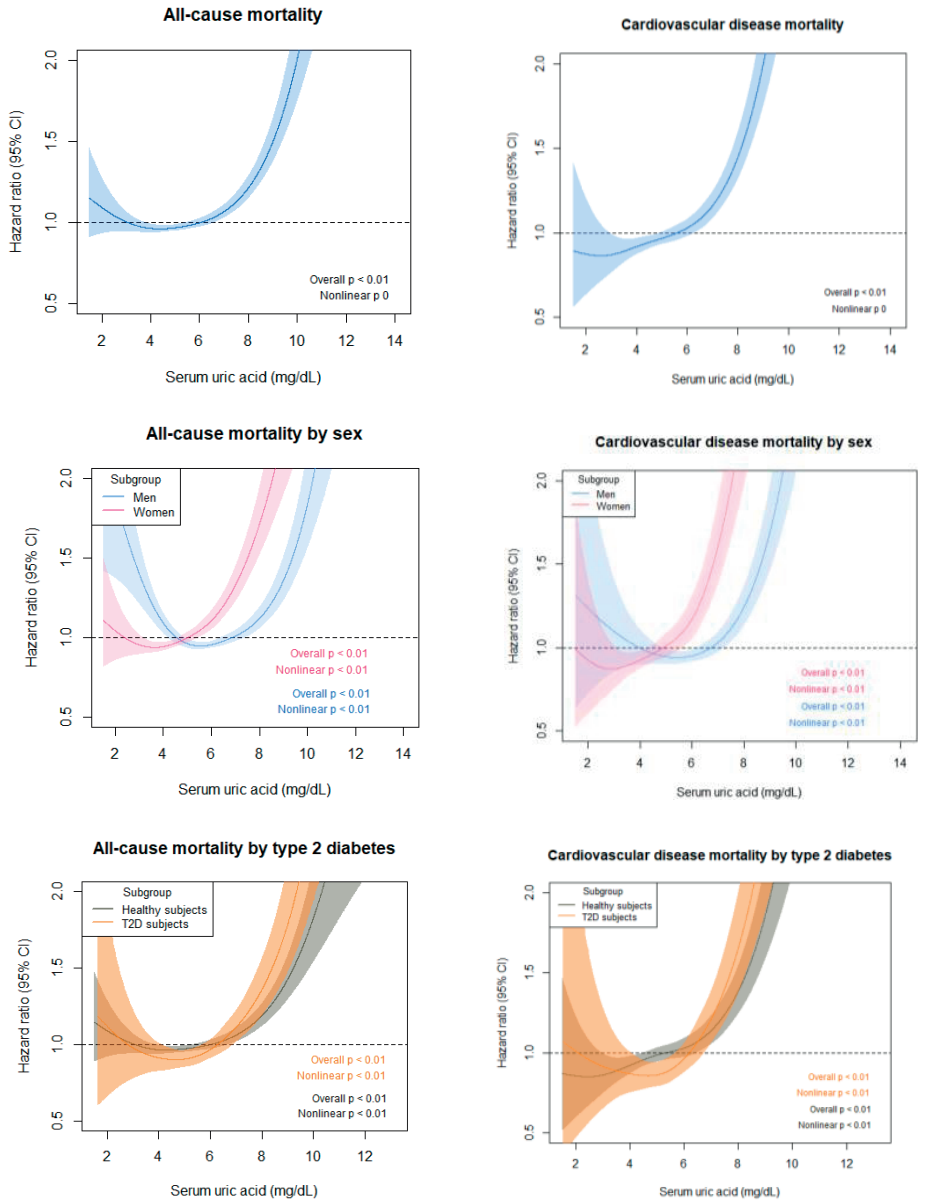


Figure 2: Associations between SUA concentrations and all-cause and cardiovascular disease mortality , according to sex and type 2 diabetes status.

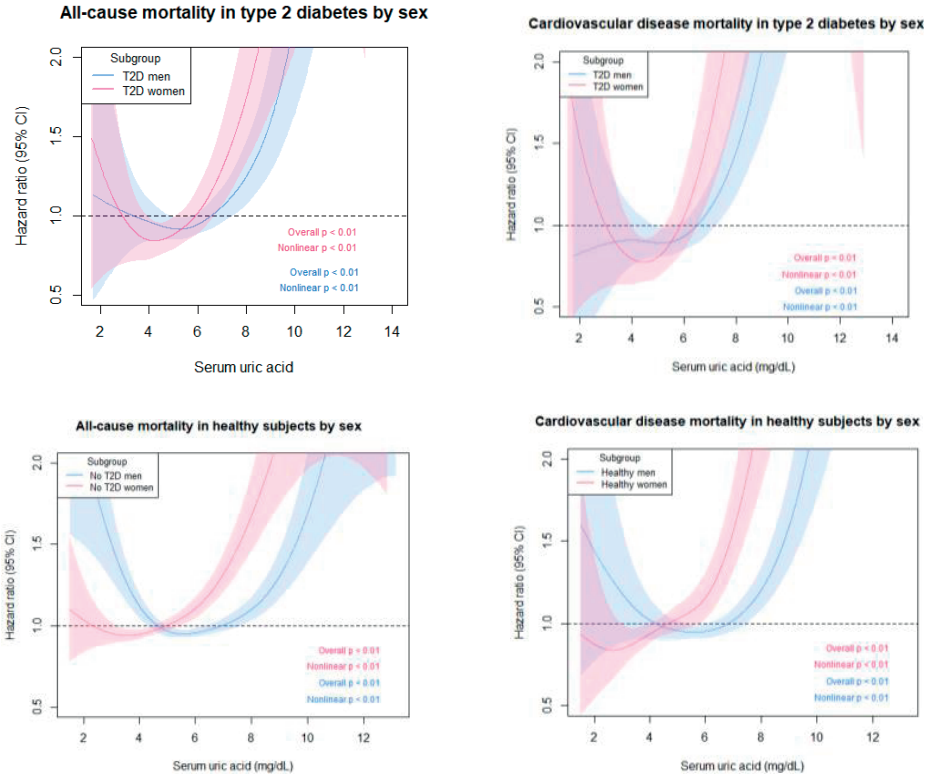


Figure 2: Associations between SUA concentrations and all-cause and cardiovascular disease mortality , according to sex and type 2 diabetes status. *(continued)*

Regarding nonfatal outcomes, T2D status seemed to modify SUA-associated risk of nonfatal CVD event and stroke, equally among men and women. Associations of SUA with fatal MI and stroke were not significant in any of the stratified groups (Table 4).

DISCUSSION

Overall, our findings are in line with those from previous reports showing that hyperuricemia is a risk factor for all-cause mortality and fatal CVD-related events. Our study found that that higher SUA concentrations are generally associated with higher risk of fatal and nonfatal events of CVD, HF, stroke and all-cause mortality, in participants free of CVD, gout or kidney disease at baseline and independent of several sociodemographic, lifestyle and health related factors, including sex and T2D status. However, SUA concentrations were not associated with a higher risk of fatal or nonfatal MI.

Table 4: Associations of serum uric acid with mortality and incident health outcomes by type 2 diabetes and sex

	Nondiabetic women (n= 200168)		Diabetic women (n = 5749)		Nondiabetic men (n = 155274)		Diabetic men (n = 9164)	
	HR [95% CI]	Nr. of events	HR [95% CI]	Nr. of events	HR [95% CI]	Nr. of events	HR [95% CI]	Nr. of events
Fatal outcomes								
All cause	**1.10 [1.07, 1.13]**	4622	**1.20 [1.11, 1.30]**	306	1.02 [0.99, 1.04]	5665	**1.11 [1.06, 1.17]**	763
CVD	**1.20 [1.14, 1.27]**	976	**1.35 [1.20, 1.51]**	126	**1.08 [1.04, 1.12]**	1796	**1.19 [1.11, 1.28]**	376
HF	**1.47 [1.28, 1.67]**	147	**1.38 [1.12, 1.70]**	35	**1.25 [1.14, 1.37]**	312	1.17 [0.99, 1.38]	73
MI	0.95 [0.79, 1.14]	106	0.77 [0.49, 1.23]	11	1.04 [0.94, 1.16]	295	0.97 [0.81, 1.17]	70
Stroke	**1.16 [1.03, 1.31]**	233	1.23 [0.96, 1.58]	29	0.98 [0.87, 1.10]	228	**1.37 [1.13, 1.67]**	42
Nonfatal outcomes								
CVD	**1.06 [1.04, 1.08]**	9095	**1.14 [1.07, 1.21]**	531	**1.05 [1.03, 1.07]**	12355	**1.10 [1.05, 1.14]**	1305
HF	**1.28 [1.16, 1.42]**	282	**1.47 [1.28, 1.70]**	72	**1.27 [1.17, 1.38]**	401	**1.35 [1.21, 1.52]**	129
MI	1.00 [0.94, 1.07]	821	1.03 [0.85, 1.26]	53	1.01 [0.97, 1.05]	1998	1.02 [0.92, 1.13]	198
Stroke	1.05 [0.99, 1.11]	1086	**1.18 [1.01, 1.38]**	80	**1.07 [1.02, 1.12]**	1329	1.09 [0.98, 1.22]	172

Model adjusted by age, sex, ethnicity, deprivation index, gout diagnosis, comorbidity, smoking, fruits and vegetables consumption, oily fish intake, red meat intake, processed meat intake, alcohol intake, sedentary time, sleeping time, body mass index category, systolic blood pressure, blood total cholesterol, blood high-density lipoprotein cholesterol.
Abbreviations: Cardiovascular disease, CVD; HF; Myocardial infarction, MI.
Bold text indicates statistical significance P >0.05

Moreover, we observed significant interactions of SUA with sex and T2D status, with generally stronger associations among diabetic women, and further provided evidence that the strength of some of the associations for specific CVD-related outcomes may differ among the groups of diabetic and nondiabetic men and women. The majority of the observed associations were J-shaped, with generally no associations across the lower range of SUA (below around 4 mg/dL in women and 6 mg/dL in men); except for the group of overall men and nondiabetic men specifically, for whom the association of SUA with higher risk of all-cause mortality followed a U-shaped curve, with significantly higher risk at SUA concentrations of below 4 mg/dL and above 6 mg/dL approximately.

In general, women have lower SUA levels than men,[38] which we also observed in our study population (mean SUA of 4.54 (SD=1.09) mg/dL in women and 5.93 (SD=1.16) mg/dL in men). It has been hypothesized that sex hormones may play a role in these different effects,[9,39] presumably due to estrogen effect. Postmenopausal women, a population at higher risk of CVD, have higher SUA concentrations than women before menopause onset.[38] In addition, differences in epigenetic mechanisms may also explain sex differences in cardiometabolic health.[40]

Our study supports the existing sex-specific cut-offs approach for hyperuricemia diagnosis (6.0 mg/dL (or 360 µmol/L) for women and 7.0 mg/dL (or 420 µmol/L) for men),[11] based on our findings that women started to show detrimental effects associated with higher SUA concentrations at lower concentrations than men, for most of the outcomes under study, such as fatal and nonfatal CVD, as well as for all-cause and stroke mortality. A similar tendency was observed among diabetic women as compared to diabetic men, for higher risk of dying from heart disease event or any cause. However, we observed that higher risk of most of the outcomes started being significant from 4.0 mg/dL in women and 6.0 mg/dL for men. In consequence, hyperuricemia diagnosis cut-offs may need further evaluation and may want to consider comorbidities, for example, by setting differential cut-off values for patients with T2D.

T2D is an important risk factor for CVD events in both men and women. In subjects without T2D, risk of CVD-related outcomes is higher among men compared to women, at all ages. However, T2D presence seems to disfavor women's advantage, which has lead diabetic women to have similar or even higher CVD risk than diabetic men.[41] The combination of hyperuricemia and T2D as risk factors, plus female sex as a susceptible population, agree with our result on diabetic women tending to have higher risks of fatal and nonfatal CVD events, and death from any cause,

associated with higher SUA concentrations, as compared with the groups of diabetic and nondiabetic men, and nondiabetic women. In the same line, it is notable our result on nondiabetic women and diabetic men showing similar strength of the SUA-related higher risks of death from CVD events or from any cause. For the nonfatal outcomes, only the modification by T2D seems to be relevant. This was the case for the association between higher SUA and CVD incidence. On the contrary, higher SUA levels were not associated with either fatal or nonfatal events of MI, in the overall population as well as in any of the stratified groups. The latter differs from a previous study that found significant associations between higher SUA concentrations higher risk of incident MI and ischemic stroke, but not haemorrhagic stroke.[2]

HF develops due to systolic/diastolic dysfunction, which causes enlargement of the heart muscle to compensate for its inefficacy to meet the needs of the body. MI is defined as death of myocardial tissue due to blood supply to the cardiomyocytes. The most common cause of MI is ischemia due to occlusive thrombosis, caused by rupture of a vulnerable atherosclerotic plaque in a coronary artery.[42] Stroke occurs when there is reduction of blood flow to part of the brain. This can occur due to a blocked artery, or leaking/bursting of a blood vessel, resulting in death of brain tissue.[43] Despite their differences, these diseases have common risk factors. The most important ones are hypertension, T2D, obesity, metabolic syndrome, sedentary behaviour, smoking, and other poor lifestyle choices.[44] High salt and high glycaemic diets can induce changes in the metabolism, resulting in increased production of intracellular uric acid.[45,46] It is hypothesized that hyperuricemia may lead to renal vasoconstriction, oxidative stress and ischaemia, and consequent hypertension.[47] Other hypotheses suggest that SUA stimulates the renin-angiotensin system (RAS), causing growth of vascular smooth cells, as well as arterial dysfunction and stiffening; and that SUA up-regulates the enzyme xanthine-oxidase (XO), which elevates oxidative stress, inducing endothelial impairment.[48] Interestingly, studies showed that treatment with allopurinol, an inhibitor of XO, significantly reduced oxidative stress and improved endothelial dysfunction.[49] Further research is needed to understand to what extent the aforementioned molecular mechanisms may differ between women and men.

Previous meta-analyses comparing sex-specific associations of SUA concentrations with risk of all-cause or cardiovascular mortality,[20,50] concluded that pooling the results from both sexes introduces high heterogeneity since the increased risk of developing T2D with higher SUA levels is stronger in women as compared to men;[50] and highlighted the lack of evidence from within-study comparisons of male and female populations, rather than single-sex population studies.[20] Besides, the existing reports from within-study sex comparisons show contradictory results for women

and men across studies, with some of them reporting significantly higher risk of CVD or all-cause mortality for one sex while others did not.[4,21-26] A more recent meta-analysis studied SUA as a risk factor for stroke, coronary heart disease and all-cause mortality only among T2D individuals.[27] The authors concluded that higher SUA levels represent a risk factor for all-cause mortality and stroke, but not for coronary heart disease, among diabetic patients. However, in this meta-analysis, no comparison was made with non-diabetic populations, and authors could not report a dose-response or threshold effect. Further, they warned about the significant limitations of their work, such as the small number of studies included, the different control for confounding variables and high heterogeneity observed across studies, as well as the different study designs.

There are some strengths and limitations to study that need consideration. Our study tackled both main limitations described for the current evidence, by conducting within-study sex and T2D status stratification, and sex and T2D status combined, which was possible due to the large sample size of our study population. With this approach, we were able to replicate previous findings and further provide insights on the interactions of SUA with sex and T2D status, suggesting that both may be effect modifiers in some of the associations with CVD and mortality. In addition, we used a comprehensive adjustment for sociodemographic, lifestyle and health-relates factors, and the exclusion of participants with baseline relevant conditions (CVD, gout and kidney disease), allowing us to account for potential confounding. Among our limitations, SUA concentrations and T2D assessment were carried out at the same time point; therefore, we could not determine whether high SUA levels contributed to the development of T2D status, or vice versa. However, this was not the aim of the present study.

CONCLUSION

Sex and T2D status modify the association of SUA on risk of fatal and nonfatal CVD-related events and all-cause mortality. In general, associations were stronger among women, and women started to show detrimental effects at lower SUA concentrations than men. Moreover, diabetic women had the highest SUA-related risk of dying from any cause or from CVD events; while nondiabetic women showed similar risks of nonfatal outcomes, as compared to diabetic men. Our findings suggest that men and women might benefit differently from prevention and treatment strategies on hyperuricemia and CVD-related morbidity and mortality, and that those strategies should consider the presence of T2D as a comorbidity.

SUPPLEMENTARY MATERIAL

Table S-1: Summary of participants´ baseline cohort characteristics by sex or diabetes status

	Total sample	Women	Men	Non-Diabetics	Diabetics
n	370355	205917	164438	355442	14913
Age, years; mean (SD)	56.36 (8.09)	56.31 (7.99)	56.41 (8.22)	56.21 (8.10)	59.92 (6.87)
Sex, women; n (%)	205917 (55.6)	-	-	200168 (56.3)	5749 (38.6)
Deprivation index; median [IQR]	-2.21 [-3.67, 0.37]	-2.20 [-3.66, 0.33]	-2.22 [-3.69, 0.41]	-2.24 [-3.69,0.31]	-1.42 [-3.25,1.83]
Ethnicity, n (%)					
White	351047 (94.8)	195315 (94.9)	155732 (94.7)	338003 (95.1)	13044 (87.5)
Mixed	5451 (1.5)	3255 (1.6)	2196 (1.3)	5104 (1.4)	347 (2.3)
South Asian	6910 (1.9)	3276 (1.6)	3634 (2.2)	5961 (1.7)	949 (6.4)
Black	5811 (1.6)	3358 (1.6)	2453 (1.5)	5284 (1.5)	527 (3.5)
Chinese	1136 (0.3)	713 (0.3)	423 (0.3)	1090 (0.3)	46 (0.3)
Smoking, n (%)					
Never	205615 (55.5)	123186 (59.8)	82429 (50.1)	198707 (55.9)	6908 (46.3)
Former	126714 (34.2)	64833 (31.5)	61881 (37.6)	120308 (33.8)	6406 (43.0)
Current	38026 (10.3)	17898 (8.7)	20128 (12.2)	36427 (10.2)	1599 (10.7)
Fruits and vegetables consumption, n (%)					
Low	131217 (35.4)	59680 (29.0)	71537 (43.5)	126588 (35.6)	4629 (31.0)
Middle	116553 (31.5)	67940 (33.0)	48613 (29.6)	111929 (31.5)	4624 (31.0)
High	122585 (33.1)	78297 (38.0)	44288 (26.9)	116925 (32.9)	5660 (38.0)
Red meat consumption, portions per week; median [IQR]	1.50 [1.50, 2.50]	1.50 [1.50, 2.50]	2.00 [1.50, 2.50]	1.50 [1.50,2.50]	2.00 [1.50,3.00]
Oily fish, portions per week; median [IQR]	2.00 [1.00, 2.00]	2.00 [1.00, 2.00]	2.00 [1.00, 2.00]	2.00 [1.00,2.00]	2.00 [1.00,2.00]
Processed meat, portions per week; median [IQR]	2.00 [1.00, 3.00]	1.00 [1.00, 2.00]	2.00 [1.00, 2.00]	2.00 [1.00,3.00]	2.00 [1.00,3.00]
Alcohol consumption frequency, n (%)					

Table S-1: Summary of participants´ baseline cohort characteristics by sex or diabetes status (*continued*)

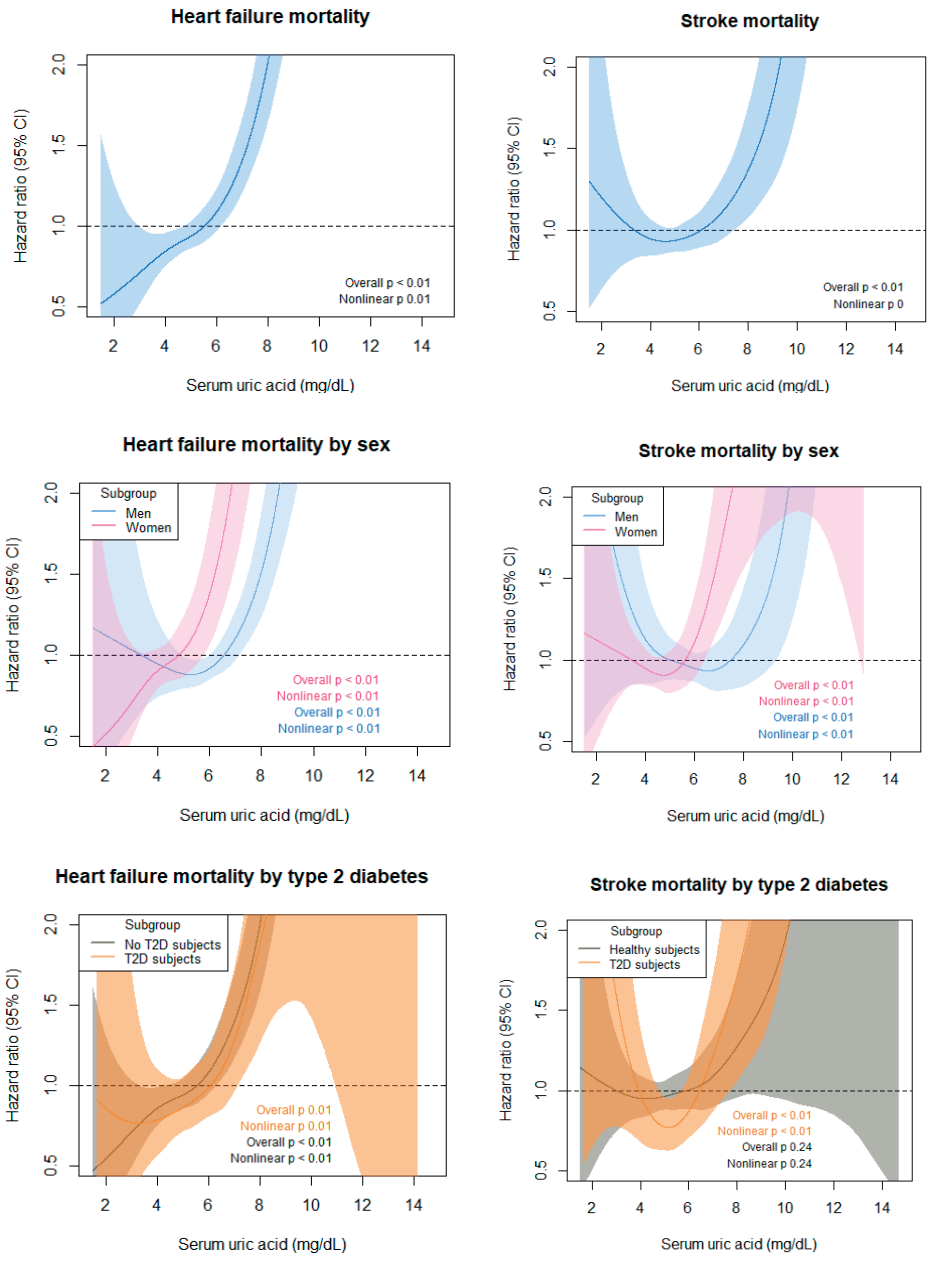
	Total sample	Women	Men	Non-Diabetics	Diabetics
Daily or almost daily	76054 (20.5)	33985 (16.5)	42069 (25.6)	73863 (20.8)	2191 (14.7)
3-4 times a week	86565 (23.4)	42951 (20.9)	43614 (26.5)	84214 (23.7)	2351 (15.8)
Once or twice a week	95949 (25.9)	53227 (25.8)	42722 (26.0)	92475 (26.0)	3474 (23.3)
1-3 times a month	41490 (11.2)	26871 (13.0)	14619 (8.9)	39685 (11.2)	1805 (12.1)
Special occasions only	42035 (11.3)	30330 (14.7)	11705 (7.1)	39256 (11.0)	2779 (18.6)
Never	28262 (7.6)	18553 (9.0)	9709 (5.9)	25949 (7.3)	2313 (15.5)
Total sedentary time, n (%)					
Low	169493 (45.8)	106225 (51.6)	63268 (38.5)	164569 (46.3)	4924 (33.0)
Middle	126216 (34.1)	68490 (33.3)	57726 (35.1)	121022 (34.0)	5194 (34.8)
High	74646 (20.2)	31202 (15.2)	43444 (26.4)	69851 (19.7)	4795 (32.2)
Sleeping time category, n (%)					
Normal 7-9 hours/day	273436 (73.8)	152807 (74.2)	120629 (73.4)	263265 (74.1)	10171 (68.2)
Short sleep <7 hours/day	90843 (24.5)	49538 (24.1)	41305 (25.1)	86682 (24.4)	4161 (27.9)
Long sleep >9 hours/day	6076 (1.6)	3572 (1.7)	2504 (1.5)	5495 (1.5)	581 (3.9)
Body mass index category (BMI), n (%)					
Underweight (BMI < 18.5)	1949 (0.5)	1551 (0.8)	398 (0.2)	1931 (0.5)	18 (0.1)
Normal (BMI 18.5 to 24.9)	124247 (33.5)	81505 (39.6)	42742 (26.0)	122650 (34.5)	1597 (10.7)
Overweight (BMI 25.0 to 29.9)	157663 (42.6)	75656 (36.7)	82007 (49.9)	152460 (42.9)	5203 (34.9)
Obese (BMI ≥ 30.0)	86496 (23.4)	47205 (22.9)	39291 (23.9)	78401 (22.1)	8095 (54.3)
Systolic blood pressure, mmHg; mean (SD)	137.77 (18.62)	135.20 (19.20)	140.98 (17.33)	137.60 (18.66)	141.67 (17.03)
HDL-C, mmol/dL; mean (SD)	1.46 (0.38)	1.60 (0.38)	1.29 (0.31)	1.47 (0.38)	1.20 (0.32)
Total cholesterol, mmol/dL; mean (SD)	5.75 (1.12)	5.90 (1.11)	5.56 (1.10)	5.80 (1.10)	4.49 (1.01)
Uric acid, µm/L; mean (SD)	5.16 (1.32)	4.54 (1.09)	5.93 (1.16)	5.14 (1.31)	5.59 (1.39)
Type 2 diabetes cases; n (%)	14913 (4.0)	5749 (2.8)	9164 (5.6)	-	-

Table 2: Summary of participants´ baseline cohort characteristics by sex and diabetes status

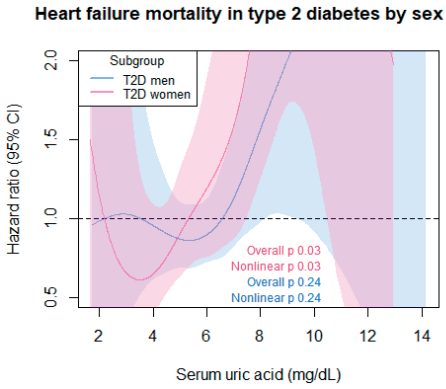
	Non-diabetic women	Diabetic women	Non-diabetic men	Diabetic men
n	200,168	5,749	155,274	9,164
Age, years; mean (SD)	56.21 (7.99)	59.93 (6.77)	56.20 (8.24)	59.92 (6.94)
Deprivation index; median [IQR]	-2.22 [-3.67, 0.28]	-1.24 [-3.13, 1.93]	-2.25 [-3.70, 0.33]	-1.53 [-3.34, 1.76]
Ethnicity, n (%)				
White	190347 (95.1)	4968 (86.4)	147656 (95.1)	8076 (88.1)
Mixed	3089 (1.5)	166 (2.9)	2015 (1.3)	181 (2.0)
South Asian	2948 (1.5)	328 (5.7)	3013 (1.9)	621 (6.8)
Black	3094 (1.5)	264 (4.6)	2190 (1.4)	263 (2.9)
Chinese	690 (0.3)	23 (0.4)	400 (0.3)	23 (0.3)
Smoking, n (%)				
Never	119894 (59.9)	3292 (57.3)	78813 (50.8)	3616 (39.5)
Former	62868 (31.4)	1965 (34.2)	57440 (37.0)	4441 (48.5)
Current	17406 (8.7)	492 (8.6)	19021 (12.2)	1107 (12.1)
Fruits and vegetables consumption, n (%)				
Low	58313 (29.1)	1367 (23.8)	68275 (44.0)	3262 (35.6)
Middle	66107 (33.0)	1833 (31.9)	45822 (29.5)	2791 (30.5)
High	75748 (37.8)	2549 (44.3)	41177 (26.5)	3111 (33.9)
Red meat consumption, portions per week; median [IQR]	1.50 [1.50, 2.50]	2.00 [1.50, 2.50]	2.00 [1.50, 2.50]	2.00 [1.50, 3.00]
Oily fish, portions per week; median [IQR]	2.00 [1.00, 2.00]	2.00 [1.00, 2.00]	2.00 [1.00, 2.00]	2.00 [1.00, 2.00]
Processed meat, portions per week; median [IQR]	1.00 [1.00, 2.00]	2.00 [1.00, 3.00]	2.00 [1.00, 3.00]	2.00 [1.00, 3.00]
Alcohol consumption frequency, n (%)				
Daily or almost daily	33536 (16.8)	449 (7.8)	40327 (26.0)	1742 (19.0)
3-4 times a week	42401 (21.2)	550 (9.6)	41813 (26.9)	1801 (19.7)
Once or twice a week	52120 (26.0)	1107 (19.3)	40355 (26.0)	2367 (25.8)

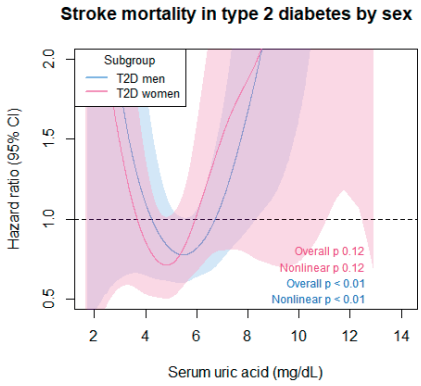
Table 2: Summary of participants´ baseline cohort characteristics by sex and diabetes status (*continued*)

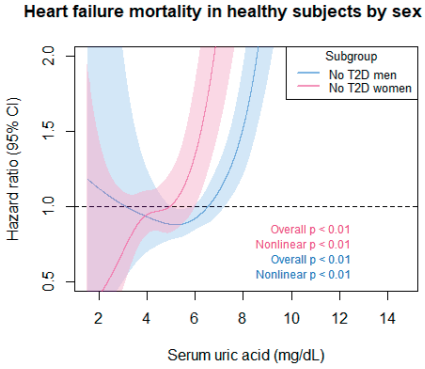
	Non-diabetic women	Diabetic women	Non-diabetic men	Diabetic men
1-3 times a month	26089 (13.0)	782 (13.6)	13596 (8.8)	1023 (11.2)
Special occasions only	28716 (14.3)	1614 (28.1)	10540 (6.8)	1165 (12.7)
Never	17306 (8.6)	1247 (21.7)	8643 (5.6)	1066 (11.6)
Total sedentary time, n (%)				
Low	103929 (51.9)	2296 (39.9)	60640 (39.1)	2628 (28.7)
Middle	66447 (33.2)	2043 (35.5)	54575 (35.1)	3151 (34.4)
High	29792 (14.9)	1410 (24.5)	40059 (25.8)	3385 (36.9)
Sleeping time category, n (%)				
Normal 7-9 hours/day	148984 (74.4)	3823 (66.5)	114281 (73.6)	6348 (69.3)
Short sleep <7 hours/day	47849 (23.9)	1689 (29.4)	38833 (25.0)	2472 (27.0)
Long sleep >9 hours/day	3335 (1.7)	237 (4.1)	2160 (1.4)	344 (3.8)
Body mass index category, n (%)				
Underweight	1542 (0.8)	9 (0.2)	389 (0.3)	9 (0.1)
Normal	80891 (40.4)	614 (10.7)	41759 (26.9)	983 (10.7)
Overweight	74057 (37.0)	1599 (27.8)	78403 (50.5)	3604 (39.3)
Obese	43678 (21.8)	3527 (61.3)	34723 (22.4)	4568 (49.8)
Systolic blood pressure, mmHg; mean (SD)	135.06 (19.23)	140.17 (17.52)	140.88 (17.37)	142.61 (16.64)
HDL-C, mmol/dL; mean (SD)	1.61 (0.37)	1.32 (0.34)	1.30 (0.31)	1.12 (0.28)
Total cholesterol, mmol/dL; mean (SD)	5.93 (1.10)	4.74 (1.01)	5.63 (1.07)	4.34 (0.97)
Uric acid, µm/L; mean (SD)	4.52 (1.08)	5.26 (1.39)	5.94 (1.15)	5.79 (1.35)

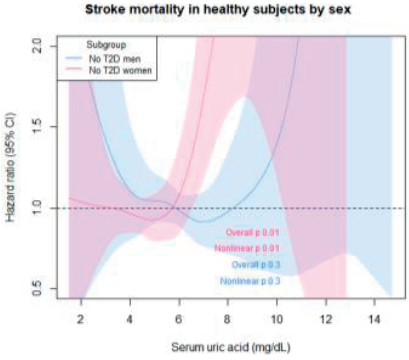


Supplementary Figure S-1: Nonlinear associations of SUA concentrations with fatal events of heart failure and stroke, according to sex and type 2 diabetes status.

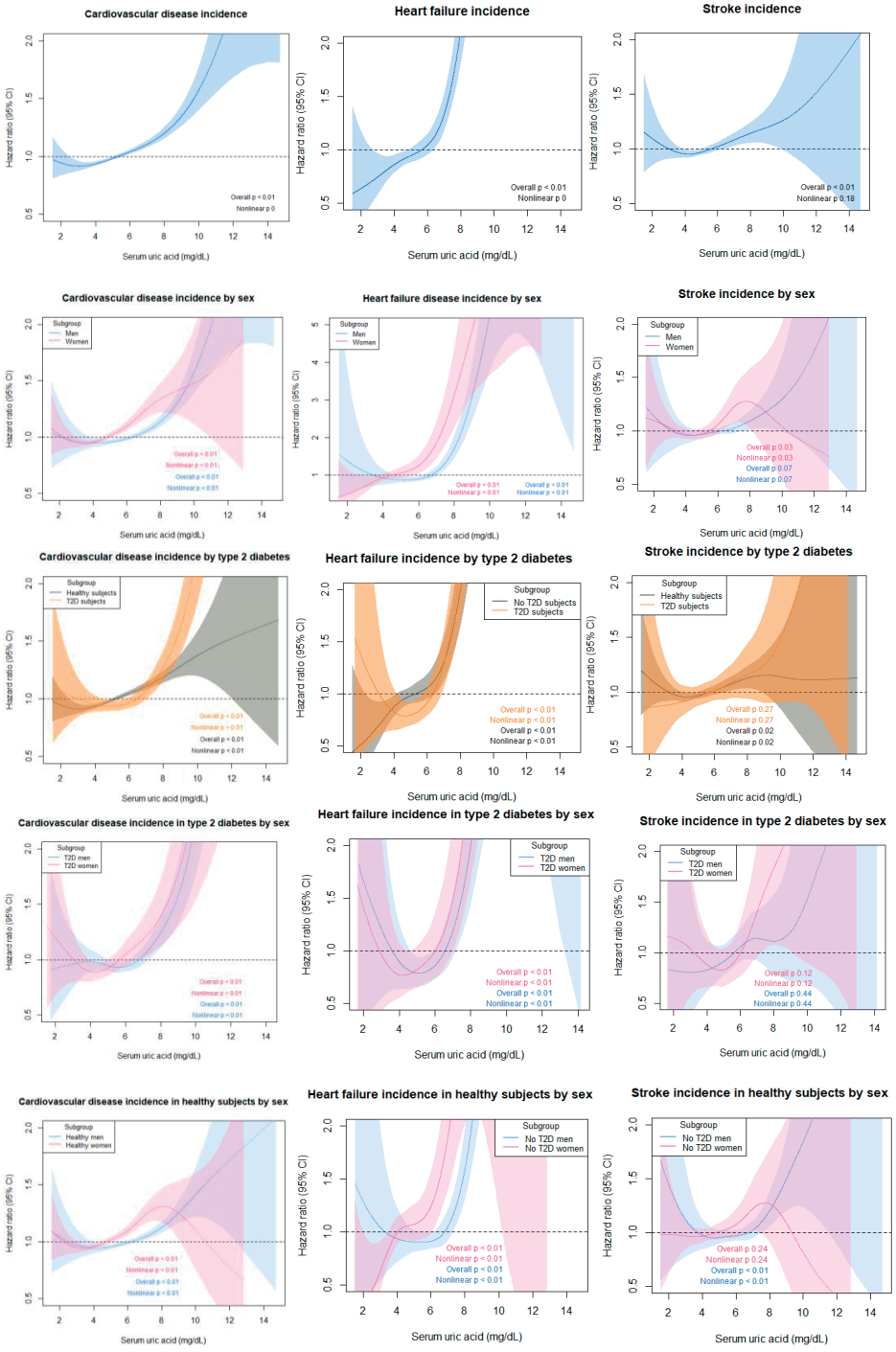








Supplementary Figure S-1: Nonlinear associations of SUA concentrations with fatal events of heart failure and stroke, according to sex and type 2 diabetes status. *(continued)*



Supplementary Figure S-2: Nonlinear associations of SUA concentrations with nonfatal cardiovascular disease, heart failure and stroke events, according to sex and type 2 diabetes status.

REFERENCES

1. Qin L, Yang Z, Gu H, et al. Association between serum uric acid levels and cardiovascular disease in middle-aged and elderly Chinese individuals. BMC Cardiovasc Disord 2014;14:26.
2. Bos MJ, Koudstaal PJ, Hofman A, Witteman JC, Breteler MM. Uric acid is a risk factor for myocardial infarction and stroke: the Rotterdam study. Stroke 2006;37:1503-7.
3. Beard JT, 2nd. Serum uric acid and coronary heart disease. Am Heart J 1983;106:397-400.
4. Cheong E, Ryu S, Lee JY, et al. Association between serum uric acid and cardiovascular mortality and all-cause mortality: a cohort study. J Hypertens 2017;35 Suppl 1:S3-S9.
5. Ong G, Davis WA, Davis TM. Serum uric acid does not predict cardiovascular or all-cause mortality in type 2 diabetes: the Fremantle Diabetes Study. Diabetologia 2010;53:1288-94.
6. Borghi C, Rodriguez-Artalejo F, De Backer G, et al. Serum uric acid levels are associated with cardiovascular risk score: A post hoc analysis of the EURIKA study. Int J Cardiol 2018;253:167-73.
7. Collaborators GBDCoD. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1736-88.
8. Mosca L, Barrett-Connor E, Wenger NK. Sex/gender differences in cardiovascular disease prevention: what a difference a decade makes. Circulation 2011;124:2145-54.
9. Regitz-Zagrosek V, Kararigas G. Mechanistic Pathways of Sex Differences in Cardiovascular Disease. Physiol Rev 2017;97:1-37.
10. Barr WG. Uric Acid. 1990.
11. Li QR, Li XD, Kwong JSW, et al. Diagnosis and treatment for hyperuricaemia and gout: a protocol for a systematic review of clinical practice guidelines and consensus statements. Bmj Open 2017;7.
12. Zhu Y, Pandya BJ, Choi HK. Prevalence of gout and hyperuricemia in the US general population: the National Health and Nutrition Examination Survey 2007-2008. Arthritis Rheum 2011;63:3136-41.
13. Emerging Risk Factors C, Sarwar N, Gao P, et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet 2010;375:2215-22.
14. Bombelli M, Quarti-Trevano F, Tadic M, et al. Uric acid and risk of new-onset metabolic syndrome, impaired fasting glucose and diabetes mellitus in a general Italian population: data from the Pressioni Arteriose Monitorate E Loro Associazioni study. J Hypertens 2018;36:1492-8.
15. van der Schaft N, Brahimaj A, Wen KX, Franco OH, Dehghan A. The association between serum uric acid and the incidence of prediabetes and type 2 diabetes mellitus: The Rotterdam Study. PLoS One 2017;12:e0179482.
16. Chang JB, Chen YL, Hung YJ, et al. The Role of Uric Acid for Predicting Future Metabolic Syndrome and Type 2 Diabetes in Older People. J Nutr Health Aging 2017;21:329-35.
17. Krishnan E, Pandya BJ, Chung L, Hariri A, Dabbous O. Hyperuricemia in young adults and risk of insulin resistance, prediabetes, and diabetes: a 15-year follow-up study. Am J Epidemiol 2012;176:108-16.

18. Bhole V, Choi JW, Kim SW, de Vera M, Choi H. Serum uric acid levels and the risk of type 2 diabetes: a prospective study. Am J Med 2010;123:957-61.
19. Cardona F, Rojo-Martinez G, de la Cruz Almaraz M, Soriguer F, Garcia-Fuentes E, Tinahones FJ. El acido urico es un predictor de desarrollo de diabetes mellitus tipo 2 en la poblacion general. Endocrinologia y Nutricion 2009;56:66-70.
20. Yang Y, Fan Y, Li J, et al. Serum uric acid as a predictor for cardiovascular and all-cause mortality in women versus men. Int J Cardiol 2015;185:125-8.
21. Juraschek SP, Tunstall-Pedoe H, Woodward M. Serum uric acid and the risk of mortality during 23 years follow-up in the Scottish Heart Health Extended Cohort Study. Atherosclerosis 2014;233:623-9.
22. Chen JH, Chuang SY, Chen HJ, Yeh WT, Pan WH. Serum uric acid level as an independent risk factor for all-cause, cardiovascular, and ischemic stroke mortality: a Chinese cohort study. Arthritis Rheum 2009;61:225-32.
23. Sakata K, Hashimoto T, Ueshima H, Okayama A, Group NDR. Absence of an association between serum uric acid and mortality from cardiovascular disease: NIPPON DATA 80, 1980-1994. National Integrated Projects for Prospective Observation of Non-communicable Diseases and its Trend in the Aged. Eur J Epidemiol 2001;17:461-8.
24. Hu P, Seeman TE, Harris TB, Reuben DB. Is serum uric acid level associated with all-cause mortality in high-functioning older persons: MacArthur studies of successful aging? J Am Geriatr Soc 2001;49:1679-84.
25. Culleton BF. Relations of serum uric acid to cardiovascular disease events and mortality: The Framingham heart study. Circulation 1998;98:170-.
26. Freedman DS, Williamson DF, Gunter EW, Byers T. Relation of Serum Uric-Acid to Mortality and Ischemic-Heart-Disease - the Nhanes-I Epidemiologic Follow-up-Study. American Journal of Epidemiology 1995;141:637-44.
27. Shao Y, Shao H, Sawhney MS, Shi L. Serum uric acid as a risk factor of all-cause mortality and cardiovascular events among type 2 diabetes population: Meta-analysis of correlational evidence. J Diabetes Complications 2019:107409.
28. Collins R. What makes UK Biobank special? Lancet 2012;379:1173-4.
29. Palmer LJ. UK Biobank: bank on it. Lancet 2007;369:1980-2.
30. 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.
31. Centre UBC. UK Biobank: Protocol for a large-scale prospective epidemiological resource
32. International Statistical Classification of Diseases and Related Health Problems 10th Revision. ICD-10 Version:2019. (Accessed 25-06-2020, at https://icd.who.int/browse10/2019/en.)
33. Mackenbach JP. Health and Deprivation - Inequality and the North - Townsend,P, Phillimore,P, Beattie,A. Health Policy 1988;10:207-.
34. 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:e6.
35. 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.

36. Celis-Morales CA, Lyall DM, Steell L, et al. Associations of discretionary screen time with mortality, cardiovascular disease and cancer are attenuated by strength, fitness and physical activity: findings from the UK Biobank study. BMC Med 2018;16:77.
37. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organ Tech Rep Ser 2000;894:i-xii, 1-253.
38. Feig DI, Kang DH, Johnson RJ. Uric acid and cardiovascular risk. N Engl J Med 2008;359:1811-21.
39. Adamopoulos D, Vlassopoulos C, Seitanides B, Contoyiannis P, Vassilopoulos P. The relationship of sex steroids to uric acid levels in plasma and urine. Acta Endocrinol (Copenh) 1977;85:198-208.
40. Asllanaj E, Zhang X, Ochoa Rosales C, et al. Sexually dimorphic DNA-methylation in cardiometabolic health: A systematic review. Maturitas 2020;135:6-26.
41. Huebschmann AG, Huxley RR, Kohrt WM, Zeitler P, Regensteiner JG, Reusch JEB. Sex differences in the burden of type 2 diabetes and cardiovascular risk across the life course. Diabetologia 2019;62:1761-72.
42. Frangogiannis NG. Pathophysiology of Myocardial Infarction. Compr Physiol 2015;5:1841-75.
43. Hossmann KA. Pathophysiology and therapy of experimental stroke. Cell Mol Neurobiol 2006;26:1057-83.
44. Rogers C, Bush N. Heart Failure: Pathophysiology, Diagnosis, Medical Treatment Guidelines, and Nursing Management. Nurs Clin North Am 2015;50:787-99.
45. Lanaspa MA, Ishimoto T, Li N, et al. Endogenous fructose production and metabolism in the liver contributes to the development of metabolic syndrome. Nat Commun 2013;4:2434.
46. Lanaspa MA, Kuwabara M, Andres-Hernando A, et al. High salt intake causes leptin resistance and obesity in mice by stimulating endogenous fructose production and metabolism. Proc Natl Acad Sci U S A 2018;115:3138-43.
47. Sanchez-Lozada LG, Rodriguez-Iturbe B, Kelley EE, et al. Uric acid and Hypertension: An Update with Recommendations. Am J Hypertens 2020.
48. Muiesan ML, Agabiti-Rosei C, Paini A, Salvetti M. Uric Acid and Cardiovascular Disease: An Update. Eur Cardiol 2016;11:54-9.
49. Farquharson CA, Butler R, Hill A, Belch JJ, Struthers AD. Allopurinol improves endothelial dysfunction in chronic heart failure. Circulation 2002;106:221-6.
50. Xu YL, Xu KF, Bai JL, et al. Elevation of serum uric acid and incidence of type 2 diabetes: A systematic review and meta-analysis. Chronic Dis Transl Med 2016;2:81-91.

Chapter 6.

General discussion and summary

Chapter 6.1.

General discussion

1. INTEGRATION OF THE FINDINGS

The aims of this thesis were to investigate potential underlying mechanisms explaining associations of protective and adverse dietary and drug risk factors of type 2 diabetes. I focused on the role of epigenetic changes and inflammation processes as potential pathways through the use of original studies and literature reviews approaches.

I studied epigenetic changes involved in promotion of type 2 diabetes development and the feasibility of their use in clinical practice for prediction, diagnosis and prognosis of this disease, as well as whether these epigenetic determinants for type 2 diabetes and cardiometabolic health occur differently in women and men. Next, I explored DNA methylation and inflammatory processes as potential underlying mechanisms linking statin treatment and coffee consumption to type 2 diabetes development. Further, I explored uric acid levels as a risk for all-cause mortality and cardiovascular disease, and how this risk is affected by type 2 diabetes status and sex differences. Finally, in this section the main findings are summarized and some methodological considerations, as well as future perspectives are discussed.

1.1 Epigenetics and type 2 diabetes

In **chapter 2** I extensively reviewed the chromatin landscape and epigenetic biomarkers for clinical diagnosis and prognosis of type 2 diabetes and the sexually dimorphic DNA-methylation patterns in cardiometabolic health. Differential epigenetic patterns of around 77 genes and repetitive and transposable elements have been explored so far in association with glycemic traits and type 2 diabetes. These investigations have used diverse tissues such as whole blood, leukocytes, adipose tissue, skeletal muscle, pancreas, liver, placenta and cord blood. Histone modifications and DNA methylation changes were explored, being the latter the most frequently epigenetic mechanisms studied. Notwithstanding the numerous efforts, evidence is still very limited and results are inconsistent, with some studies showing significant results while some others did not.

According to the reviewed literature, different approaches have been used for the study of DNA methylation signatures: global DNA methylation assessment, DNA methylation in candidate genes, or in a hypothesis free approach like in Epigenome-Wide Association Studies (EWAS). In consequence, the different investigations explored different methylation sites in the genome using different techniques, making consistency in the replication of the findings challenging. Nevertheless, the reported differentially methylated genes seem to be involved in biological processes

that are relevant to type 2 diabetes development, such as insulin secretion, oxidative stress, energy expenditure, glucose transport and inflammation. Studies devoted to histone modifications and type 2 diabetes and glycemic traits are even more limited. The most repeatedly replicated methylation marks have been tested as potential biomarkers for risk prediction and disease prognosis. For this purpose, statistical prediction models have been used to examine the predictive value of methylation at *LINE-1* in type 2 diabetes risk. This showed no significant improvement as compared to traditional type 2 diabetes risk assessment models. *ABCG1* (CpG site cg06500161) and *PHOSPHO1* (CpG site cg02650017) methylation have been suggested as biomarkers but their predictive value has not been tested to date. Additionally, epigenetic modifications have been associated to cardiovascular disease, diabetic nephropathy, diabetic retinopathy and diabetic neuropathy, all of which are part of diabetic complications. However, overall, little evidence is available.

Among the articles reviewed in this chapter, only a few reported sex-stratified associations between DNA methylation and glycemic traits and type 2 diabetes. On the one hand, men-specific hypomethylation was found in *FTO* and *HOOK2*, whereas men-specific hypermethylation was observed in *BCL11A, GCK* and *IGFBP7* in association with type 2 diabetes risk or in newly diagnosed patients versus healthy controls. On the other hand, women-specific differential hypermethylation was reported for *HOOK2, PTPN1*. Finally, X chromosome-wide global DNA methylation was higher in women pancreatic islets. Although there is some evidence linking differential DNA methylation modifications between sexes with risk of cardiometabolic disease including type 2 diabetes and glycemic traits, more studies and consistent results are lacking.

For the available evidence several limitations need to be addressed: the majority of the studies were carried out using a cross-sectional design, therefore causality or temporality cannot be claimed; the sample size of the reviewed studies were limited; most studies are in Caucasians, therefore generalization to other populations is not feasible; different studies used diverse levels of adjustment and apply different thresholds for statistical significance, thus making replication of the findings a very difficult endeavor. All the aforementioned and the high heterogeneity among the reports prevented us from pooling the existing data. Overall, I concluded that DNA methylation signatures associated to glycemic traits and type 2 diabetes may be promising tools to be used as biomarkers of early diagnosis, years before the clinical diagnose, as well as for disease prognosis, however more robust and replicable results from carefully designed studies are required to validate their use in clinical practice for personalized medicine.

1.2 Dissecting the association between environmental factors, epigenetics and risk of type 2 diabetes

Environmental exposures throughout life course have shown to modify epigenetic patterns which in turn may lead to disease development, such as type 2 diabetes. In this thesis I explored the association between statin drugs and type 2 diabetes, and the potential role of DNA methylation as a mediator in this association.

Evidence from observational studies and randomized clinical trials have found a higher risk of type 2 diabetes among patients under statin class of drugs treatment as compared to non-statin users at baseline. However potential mechanisms underlying such association were unknown. In **chapter 3.1** I carried out a detailed population-based study investigating the effect of the different types of statins on type 2 diabetes risk. I explored the relationship of incident type 2 diabetes with current, ever and past statin use as compared to never use in longitudinal analyses; as well as the cross-sectional association of serum fasting glycemic traits with current, ever and past statin use as compared to never use. After comprehensive adjustment for confounders, I found that current statin treatment at baseline was associated with higher serum fasting insulin and HOMA-IR, while the relation with fasting glucose was mediated by BMI and hypertension. Moreover, current and ever statin users, but not former users, showed higher risk of developing type 2 diabetes. Duration of statin treatment but not statin subtype and dosage, was significantly related to higher risk of type 2 diabetes. Cumulative use of statins as well as higher statin dosage were associated with incident impaired fasting glucose. In stratified analyses, the increased risk of type 2 diabetes with statin therapy was significant only among statin users aged 65 years and above, males and overweight/obese.

In **chapter 3.2** I further investigated a potential mechanism underlying the observed associations of exposure to statin therapy with outcomes type 2 diabetes and glycemic traits. I hypothesized that statins may change DNA methylation patters consequently altering expression of relevant genes, thus leading to increased risk of type 2 diabetes. For this purpose, I performed a two-phase Epigenome-wide association analyses, using data from five prospective cohort studies. I identified and replicated five DNA methylation sites differentially methylated among current statin users as compared to non-current users. The identified CpG sites were annotated to genes involved in lipids metabolism (genes *DHCR24* and *SC4MOL),* as well as molecules transport and insulin secretion (*ABCG1*). Moreover, methylation at sites in *DHCR24* and *ABCG1* were associated with lower expression of gene *ABCG1* and with higher levels of serum fasting insulin and HOMA-R and incident type 2 diabetes. Importantly, I provided evidence that methylation at CpG site cg06500161

(*ABCG1*) may have a role partially mediating the association between statins therapy and higher fasting insulin and HOMA-IR using causal mediation analysis. This latter result was independent of blood lipids, excluding type 2 diabetes participants and using a cross-sectional approach. Among statin users, I found that increasing dose of statins was associated with methylation at SC*4MOL.*

Based on our results and previous evidence from population and experimental studies, I hypothesize that statins promote DNA methylation changes that alter gene expression, suggesting an epigenetic regulation of *ABCG1* expression. In turn, this may affect insulin secretion or other mechanisms leading to higher insulin and HOMA-IR.

Previous studies have observed a protective association between coffee consumption and lower risk of type 2 diabetes. The hypothesized potential mechanisms explaining these observations are (i) the action of chlorogenic acid, a phenolic coffee compound which may reduce reducing glucose intestinal uptake through, thus reducing blood glucose levels; (ii) the antioxidant capacity of some coffee compounds that may reduce oxidative stress; (iii) the anti-inflammatory effects on inflammatory markers altered in T2D; (iv) the modulation of the microbiome content and diversity; and (v) the modulation of adenosine receptor signaling which in turn modulates insulin and glucagon signaling. Therefore, in **chapter 4**, I sought to investigate the association between coffee consumption and risk of T2D as well as the potential role of type 2 diabetes related-proinflammatory cytokines and adipokines mediating this association. For this purpose, two large population studies were used, the Rotterdam Study and the UK Biobank. Overall, my work confirmed previous findings on the protective effect of coffee consumption lowering T2D risk, and I further observed beneficial associations between higher coffee consumption and lower insulin resistance over time among non-diabetics, as well as with lower plasma C-reactive protein (CRP), lower complement 3 (C3) and higher adiponectin levels. In addition, I observed an adverse positive association of coffee intake with blood IL-18 levels. Furthermore, I found evidence that CRP levels and adiponectin may mediate part of the observed protective effect of coffee intake. CRP, C3 and IL-18 are pro-inflammatory markers that are more often described to be increased in type 2 diabetes. Although my findings on IL-18 may seem contradictory, previous experimental studies have reported that deficiency of IL-18 may lead to metabolic syndrome due to hyperphagia, obesity and insulin resistance in mice. Thus, the study of the role of IL-18 controlling energy expenditure may high interest for further research. Adiponectin is an adipokine with anti-inflammatory and insulin-sensitization effects, and it has been found to be decreased in obesity and type 2 diabetes. My findings agree with

those suggesting that the beneficial effect of coffee on lower risk of type 2 diabetes may occur through the modulation of the inflammatory response. To the best of my knowledge, this study is the first to suggest that CRP and adiponectin variations due to coffee consumption may be in the causal pathway. Further research should be conducted to determine to what extent coffee consumption should be promoted for T2D prevention in the population.

Type 2 diabetes and high concentrations of serum uric acid have shown to be important risk factors for all-cause mortality and cardiovascular events. Previous studies have also reported a higher risk of all-cause mortality and cardiovascular events among men. In **chapter 5** I explored how diabetes and sex differences modified the effect of high serum uric acid (SUA) concentrations on fatal and nonfatal cardiovascular disease-related events (cardiovascular disease, heart disease, ischemic coronary heart disease, heart failure, myocardial infarction and stroke), as well as all-cause mortality, in a prospective study. First, I observed that higher SUA concentrations were associated with a higher risk of all fatal and nonfatal outcomes, except for myocardial infarction fatal and nonfatal events. Further, I found that prevalent type 2 diabetes status and female sex were factors that showed to increase, or to have a tendency to increase, SUA-related higher risks of the majority, but not all, the outcomes under study. Further research should investigate the biological mechanisms underlying these associations.

2. METHODOLOGICAL CONSIDERATIONS

2.1 Study design in epigenetic studies

Epigenetic marks are prone to changes over time as the product of environmental exposures and the cumulative effect of an epigenetic maintenance system, and are closely related to biological age.[1] These variations may contribute to the late onset of common human diseases as the result of (epi)genome-environment interaction. Such effects are better observed in studies with a prospective design where, by following up a sample of the population it is possible to identify risk factors before the occurrence of the event of interest.[2] Although most of the cohort studies included in the original research described in this thesis (the Rotterdam Study,[3] KORA-F4,[4] LOLIPOP,[5] SHIP-Trend,[6] ESTHER[7]) are prospective cohorts, the majority of these are limited to cross-sectional analyses of DNA methylation and risk factors or diseases status at one time-point, or the assessment of future risk of disease in association with single measurements of epigenetic marks before the event. Thus, among available evidence to date, research on diseases prevalence prevails over disease incidence

investigations. Due to the cross-sectional nature of the associations explored in such studies, conclusions on temporality cannot be derived; therefore, their results must be interpreted carefully.

Repeated measures of epigenetic marks overtime in longitudinal studies could capture the dynamic nature of epigenetic mechanisms, allowing us for the assessment of intra-individual time-dependent variations of the epigenome and their relation with disease development or treatment effect. For example, in the context of type 2 diabetes risk, repeated measurements of DNA methylation at different time points would allow us to investigate to what extent insulin resistance progression towards type 2 diabetes is accompanied by a correlated progressive change in DNA methylation levels; or to determine temporality of the association, in other words, whether insulin resistant predicts DNA methylation variations or vice versa, and to identify the relevant genes; or to what extent some persistent exposure may affect a health outcome via DNA methylation changes over time; or how much time is needed to reverse the changes in some epigenetic mark after exposure cessation, in case it is reversible. To the best of our knowledge, there are only few articles reporting longitudinal intra-individual DNA methylation variations, with a follow-up time ranging from 6 months up to 18 years and used limited sample sizes of specifically selected participants.[2,8-10] Although this kind of evidence from large populations-based cohort studies is yet scarce, such approach have a series of benefits. While cross-sectional approaches cannot capture the dynamic nature of epigenetic patterns and causality cannot be inferred, multiple measurements of DNA methylation marks may contribute providing insights on the involvement of the variations of DNA methylation marks in the progressions of diseases or the effect of a drug treatment or other kind of intervention by using each individual as their own control.[11] Further, repeated measurements increase the statistical power as compared to cross-sectional designs. In addition, the longitudinal design allows the study of epigenetic heritability; monozygotic twins studies suggest there are no distinguishable epigenetic differences between twins in the early years of life, however in older ages twins show notable differences in the overall content and distribution of 5-methylcytosine DNA and histone acetylation, which in turn may affect their gene-expression pattern.[12,13] Therefore, in future epigenetic studies efforts should be put in the analysis of the longitudinal trajectories of DNA methylation patterns.[14] Eventually, this evidence could be used as an additional tool for clinicians when designing strategies for the early identification of individuals at high risk of developing type 2 diabetes before the onset of the disease, as well as strategies to monitor the evolution of the health status of a person in response to a given exposure over time.

2.2 Tissue specificity and confounding in epigenetic studies

Gene regulations, including DNA methylation changes, seem to occur in a tissue-specific manner.[15,16] Thus, the trait or disease of interest should determine the selection of a suitable target tissue.[17] For the study of diabetes, drugs, and biomarkers of inflammation and metabolism, samples from pancreas, liver and adipose tissue seem most appropriate. However, there is limited availability of such specific tissues from human studies. The majority of the published reports from large cohort studies, as well as the original studies included in this thesis, have used peripheral blood as tissue to measure biomarkers and DNA methylation signatures. Saliva samples are also often used. In the context of type 2 diabetes pathophysiology, measurements in such non-specific tissues might not be representing the true effect of the exposure under investigation, leading to underestimation. Researchers are aware of such limitations, and in order to facilitate the study of gene regulation, expression, epigenomics and histology across specific less accessible tissues, large open access databases and web-based tools have been created based on data from studies in which these various tissues were available. Examples of these large scale efforts are genotype-tissue expression GTEx (https://gtexportal.org/home/), the Encyclopedia of DNA Elements ENCODE,[18] Blood and Brain Correlation BECon,[19] and the blood-brain DNA methylation comparison tool (http://epigenetics.essex.ac.uk/bloodbrain/). Further joint efforts are needed in order to broad the range of tissues and enlarge the samples sizes available on these platforms.

As the epigenome can change in response to environmental factors, epigenetic studies are also prone to confounding. Confounding is a source of bias and arises when there are additional factors, other than the exposure under study, that are involved in either a risk or protective effect on the outcome, but do not take part in the pathway of the exposure of interest. If confounding is not properly addressed, it may lead to incorrect conclusions. Some confounders in observational studies, are often demographic and environmental factors. I have discussed in this thesis that changes in the epigenome are especially sensitive to various environmental stimuli and consequent disease development throughout the life course. Therefore, unlike genetic studies, epigenetic studies, are highly prone to confounding. An additional source of confounding commonly present in epigenetic studies comes from the fact that the epigenetic signatures, such as DNA methylation, are more often measured in leukocytes from peripheral whole blood and less frequently in specific target tissues, as discussed above. Leukocytes comprise an admixture of heterogeneous cell populations, with different functions and their own epigenetic profile, and whose presence in blood could vary during the occurrence of some disease. In this regard, to account for inter-individual cell heterogeneity, epigenetic epidemiologi-

cal models adjust for measured or estimated white blood cells proportion. There has been a concern in EWAS research related to the potential confounding effect of this heterogeneity across leukocytes' DNA methylation signatures. Under this scenario, the observed associations of DNA methylation in blood with some trait might reflect differences in leukocyte composition rather than true changes in the methylation sites. Nonetheless, one study concluded that this kind of confounding is not an issue in DNA methylation studies in blood.[20] I my EWAS investigation, I addressed confounding through analysis, by performing a comprehensive adjustment for potential confounding variables, such as age, sex, lifestyle factors (smoking, diet, alcohol consumption, physical activity), comorbidities, medications and measured or estimated white blood cells proportion. However, despite adjustments, there is always a risk of residual confounding. An additional way to overcome confounding in epigenetic studies would be to carry out the measurements in specific tissues according to the disease or health condition of interest, to avoid the confounding due to blood leukocytes proportion; however, as discussed above, this is often not possible. Further, statistical tools could be implemented to deal with confounding, like the use of genetic instruments, like the case of Mendelian randomization analyses. Nevertheless, I did not apply this approach in this thesis.

2.3 Ethnic heterogeneity in epigenetic studies

The vast majority of the population studies have been performed in Caucasian ancestry and so are the majority of the repositories with publicly available data on genetics and epigenetics. Further, several studies have reported ethnic differences in cardiovascular disease, diabetes and other, but not all, cardiometabolic risk factors.[21,22] Therefore, results obtained from studies in Caucasians cannot be extended to other ethnic groups. A study showed important differences in DNA methylation profile in naive CD4+ T in healthy African-Americans as compared to healthy European-American Caucasians.[23] The authors of this study suggested that these differences would confer differential susceptibility to autoimmune diseases. Similarly, ethnic disparities in DNA methylation marks have been observed in cancer prognosis and survival, as well as in healthy controls from cancer studies, albeit the direction of the change is contradictory.[24] For example, one study found higher levels of global DNA methylation among healthy non-Caucasians relative to healthy Caucasians;[25] while other study observed higher global DNA methylation among non-Hispanic whites as compared Hispanics.[26] In cancer, several studies have reported ethnicity differences in methylation status of relevant genes for cancer or global DNA methylation in prostate tumors, breast, colorectal and lung cancer, among Americans, Hispanics, Caucasians, as well as European-, American-, Asian-, African- or Israeli-born Jews and Israeli-born Arabs.[24] In contrast, an EWAS on depressive

symptoms in whole blood provided evidence that DNA methylation patterns may be similar across various ethnicities such as African Americans and European-American Caucasians.[27] Possible factors contributing to the divergence in epigenetic patterns between ethnicities may be differences in allele frequencies, epigenetic heritability, or in gene-environment interactions.[28] Nonetheless, in epigenetic studies the effect of different ethnicities might be diluted due to shared environmental exposures, such as diet and pollution,[24] given that epigenetic patterns are influenced by the environment.

Thus, there is a need for more studies investigating epigenetic mechanisms in subjects with ethnic backgrounds other than only Caucasians. These studies may contribute not only to identify the potential underlying epigenetic mechanisms linking gene-environment differences among ethnicities, but also epigenetic mechanisms may constitute a potential tool for rapid adaptation to environmental exposures and thus for evolution.[29] Nonetheless, this hypothesis requires further research.

2.4 Nutrition assessment in population studies

A tool often used in populations studies to collect information on participants' food consumption are food frequency questionnaires (FFQs). In FFQs subjects report the habitual intake of a given list of food and beverage items, specifying frequency and portion size consumed during a certain period of time, usually the past year. However, since FFQs are self-reports, they are prone to information bias and lead to measurement error, due to the retrospective nature of the dietary assessment. However, misclassification of the exposure is assumed to be non-differential, thus random and not related to the outcome of interest. However, misclassification of the exposure could also occur due to individuals either underreporting or over reporting their food consumption, depending on the type of food and specific participants' characteristics, such as age, sex, ethnicity, socioeconomic status and body mass index (BMI). For example, healthy foods such as fruits and vegetables are often over reported,[30] while individuals with higher BMI tend to underreport absolute energy, protein, sodium and potassium intake.[31] Additionally, differences in the number and type of items included in the questionnaires may also be a source of measurement error. More detailed FFQs including a higher number of items may be prone to overestimations of food consumption and total energy.[30] This may constitute a limitation in studies where different populations using different questionnaires are compared. In our study in **chapter 4** on coffee consumption, I included data on coffee intake as well as foods to build a diet quality score, that were collected through paper-based 170-item and 389-item FFQs, as well as in home interviews in the Rotterdam Study; while coffee consumption data in the UKBB was collected

through a web-based FFQ containing a total of 29 questions on dietary habits. The aforementioned sources of measurement error may affect the associations with the outcome of interest when foods constitute the main exposure of the study. In an effort to address the differences in the dietary assessment methods, the FFQs in both cohorts, the Rotterdam Study and the UK Biobank, were validated against 24-hours recall questionnaires[32,33] and additional 9-days[34] and 4-weeks[35] dietary record in the Rotterdam Study. Overall, the aforesaid FFQs validation studies showed moderate to strong correlation and reliable reproducibility of the dietary assessments using the different questionnaires.

An additional source of bias faced by nutrition studies is confounding, as also discussed in section 2.2. In an effort to diminish this issue, our study included adjustments for several potential confounding factors, such as sociodemographic and health-related factors, as well as other lifestyle factors, including physical activity, smoking status, and body mass index. Another way to deal with confounding is through stratification of the study population, in order to produce groups within which the confounding factor does not vary. Next, these subgroups are analyzed independently. Thus, the potential confounder cannot confound the association given that it does not differ across the exposure-outcome among those individuals. Different results between the non-stratified versus the stratified analyses indicate that confounding is likely. In **chapter 4** of this thesis, additional to the comprehensive adjustment for potential confounders, I also addressed latent residual confounding by carrying out independent analyses among participants in each category of smoking behavior (never, former and current smokers), normal weight and overweight/obese participants and sex. For example, I found that higher coffee intake was significantly associated with lower risk of T2D among never and former smokers. On the contrary, this was not observed among current smokers. Despite the adjustment for smoking in my primary analysis, this inconsistency observed across smoking categories suggested a residual confounding effect of smoking for the group of current smokers, for whom higher coffee consumption did not show any significant effect on lower T2D risk. Notwithstanding all the efforts, in an observational study, residual confounding can never be ruled out.

3. FUTURE PERSPECTIVES

3.1 Microarray technology for DNA methylation assessment

The development of new technology for the study of the epigenome has produced exponential increase in the knowledge we have about epigenetics and its involve-

ment in biological processes not only in the field of type 2 diabetes, but also cardiometabolic health, cancer, neuroscience, and many others. The more accessible arrays technology, have allowed cohort studies to perform massive profiling of the human epigenome among thousands of individuals. Regarding the assessment of DNA methylation signatures, arrays like Illumina Infinium HumanMethylation27 BeadChip facilitated the interrogation of 27,578 CpG sites genome-wide; whereas Infinium HumanMethylation450 BeadChip became the most popular, targeting more than 450,000 CpGs. The technology keeps evolving and currently studies are moving forward the use of Infinium MethylationEPIC BeadChip, interrogating more than 850,000 CpG sites across the genome. Despite the differences between the Infinium 450K BeadChip and the Infinium EPIC BeadChip, the latter covers approximately 90% of the same sites on the 450K array. The design of future BeadChips shall aspire to further increase the amount of CpGs interrogated. However, scientists are still studying only a small part of the epigenome. There are approximately 28 million CpG dinucleotides across the genome. In this process an additional challenge should be considered: the new DNA methylation arrays shall include technology compatible with the existing BeadChips. Currently, both the Infinium 450K BeadChip and the Infinium EPIC BeadChip use the same technology and showed overall per-sample correlations of $r > 0.99$,[36] which encourages the use of combined data. Future array development should therefore keep aiming for the technologies compatibility, to make allow the comparability of the future studies with the already available evidence.

3.2 Use of epigenetic marks in clinical practice

Notwithstanding the many mechanisms and pathways that still are to be unraveled, current evidence points out epigenetics as promising tools in clinical practice. Eventually, individuals' methylation patterns may be used in precision medicine to guide clinicians to work out the most adequate prevention strategy or tailored treatment at the right time for that particular subject. For example, based on the methylation levels at relevant CpGs for type 2 diabetes development, a person at risk of developing the disease could be identified at a very early stage, before the clinical manifestation of the disease and the occurrence of the associated vascular damage. Since the robust replication of the current findings in this field is still limited, further research with comparable technology is needed to identify and validate epigenetic marks with potential use in clinical practice. Similarly, if suitable epi-biomarkers have been identified, they could serve as a tool to help monitoring the effect of medicine or other treatment intervention, as well as the disease progression.

Furthermore, the identification of epigenetic mechanisms involved in disease has allowed the development of the pharmacoepigenomics field. This emerging discipline is being led by cancer research. Aberrant epigenetic changes are intimately associated to the genesis of cancer.[37] Cancer cells suffer epimutations that modify the structure and stability of the genome, which it is suggested to contribute to propagate carcinogenesis. Therefore, epigenetic reprogramming constitutes a promising approach for cancer treatment as well as fighting cancer drug resistance.[37] With several ongoing trials, efforts are focused on the design of the so called epigenetic drugs, or *epidrugs*. Epidrugs are chemical compounds targeting epigenetic marks such as the enzymes DNA methyltransferases (DNMTs) and histone deacetylases (HDACs). These enzymes are responsible for the maintenance and establishment of epigenetic changes. Several epigenetic therapies have been proven to effectively inhibit DNMT to modulate DNA methylation levels, thus restoring the function of genes that had been aberrantly silenced in cancer; or to promote histone acetylation by blocking the catalytic sites of HDACs and causing growth arrest and death in a wide range of cancer transformed cells. Such therapies have therefore been approved by the United States Food and Drug Administration (FDA) for use in cancer treatment and also new ones are in evaluation in pre-clinical and clinical trials.[37] In regard to type 2 diabetes research, clinical trials may also provide valuable evidence on epigenetic changes related to type 2 diabetes and cardiometabolic health. To the best of our knowledge, only a few clinical trials on epidrugs for the treatment of diabetes and its complications have been conducted or have recently started.[38] Some of those studies investigated antioxidant compounds such as resveratrol and curcumin, which may also have DNA methyltransferase and histone deacetylase inhibition or activation activities. However, the field of pharmacoepigenomics in the context of type 2 diabetes is still in its infancy and further efforts from clinical trials are required to gather enough evidence for the validation and safety and efficacy assessment of the compounds with potential epidrug use.

The inclusion of these novel preventive, treatment and monitoring strategies in the field of type 2 diabetes will positively impact the quality of life of the individuals at high risk of disease or newly diagnosed. However, further population research in the field of diabetes and cardiometabolic health is needed to robustly identify and replicate relevant epigenetic markers for their potential use as clinical biomarkers. These efforts may succeed based on collaborative work, for which the use of the similar technologies across the participating studies is key in order to allow replication and validation of the findings. Also, a stronger inclusion of trans-ethnic analyses and studies among populations with different environmental exposures, such as dietary habits, are imperative for the correct interpretation of the findings

and the better understanding of the role of the epigenetic mechanisms underlying the observed associations and their translation into clinical practice.

4. CONCLUSION

The aims of this thesis were to study potential underlying mechanisms explaining associations of protective and adverse dietary and drug risk factors of type 2 diabetes. Through the use of original studies and literature reviews, I emphasized the role of epigenetic changes, diet and inflammation processes.

Overall, from the work presented in this thesis, I can conclude that epigenetics may constitute a mechanism involved in type 2 diabetes onset, and that those epigenetic changes may occur in women and men differently. Further, among the environmental factors promoting changes in the epigenome I could find statin medication, whose diabetogenic effect could be partially mediated by differential DNA methylation. On the side of protective factors against type 2 diabetes, I found that coffee consumption may contribute modulating the inflammatory response, consequently decreasing type 2 diabetes risk.

REFERENCES

1. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol 2013;14:R115.
2. Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. Nat Rev Genet 2006;7:812-20.
3. Ikram MA, Brusselle G, Ghanbari M, et al. Objectives, design and main findings until 2020 from the Rotterdam Study. Eur J Epidemiol 2020;35:483-517.
4. Rathmann W, Haastert B, Icks A, et al. High prevalence of undiagnosed diabetes mellitus in Southern Germany: target populations for efficient screening. The KORA survey 2000. Diabetologia 2003;46:182-9.
5. Chahal NS, Lim TK, Jain P, Chambers JC, Kooner JS, Senior R. Does subclinical atherosclerosis burden identify the increased risk of cardiovascular disease mortality among United Kingdom Indian Asians? A population study. Am Heart J 2011;162:460-6.
6. Volzke H, Alte D, Schmidt CO, et al. Cohort profile: the study of health in Pomerania. Int J Epidemiol 2011;40:294-307.
7. Raum E, Rothenbacher D, Low M, Stegmaier C, Ziegler H, Brenner H. Changes of cardiovascular risk factors and their implications in subsequent birth cohorts of older adults in Germany: a life course approach. Eur J Cardiovasc Prev Rehabil 2007;14:809-14.
8. Bjornsson HT, Sigurdsson MI, Fallin MD, et al. Intra-individual change over time in DNA methylation with familial clustering. JAMA 2008;299:2877-83.
9. Tan Q, Heijmans BT, Hjelmborg JV, Soerensen M, Christensen K, Christiansen L. Epigenetic drift in the aging genome: a ten-year follow-up in an elderly twin cohort. Int J Epidemiol 2016;45:1146-58.
10. Sandovici I, Leppert M, Hawk PR, Suarez A, Linares Y, Sapienza C. Familial aggregation of abnormal methylation of parental alleles at the IGF2/H19 and IGF2R differentially methylated regions. Hum Mol Genet 2003;12:1569-78.
11. Nustad HE, Almeida M, Canty AJ, LeBlanc M, Page CM, Melton PE. Epigenetics, heritability and longitudinal analysis. BMC Genet 2018;19:77.
12. Fraga MF, Ballestar E, Paz MF, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A 2005;102:10604-9.
13. Talens RP, Christensen K, Putter H, et al. Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. Aging Cell 2012;11:694-703.
14. Staley JR, Suderman M, Simpkin AJ, et al. Longitudinal analysis strategies for modelling epigenetic trajectories. Int J Epidemiol 2018;47:516-25.
15. Lokk K, Modhukur V, Rajashekar B, et al. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. Genome Biol 2014;15:r54.
16. Sonawane AR, Platig J, Fagny M, et al. Understanding Tissue-Specific Gene Regulation. Cell Rep 2017;21:1077-88.
17. Summaries for patients. Coffee drinkers at lower risk for type 2 diabetes. Ann Intern Med 2004;140:I17.
18. Davis CA, Hitz BC, Sloan CA, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res 2018;46:D794-D801.
19. Edgar RD, Jones MJ, Meaney MJ, Turecki G, Kobor MS. BECon: a tool for interpreting DNA methylation

findings from blood in the context of brain. Transl Psychiatry 2017;7:e1187.
20. Heiss JA, Brenner H. Impact of confounding by leukocyte composition on associations of leukocyte DNA methylation with common risk factors. Epigenomics 2017;9:659-68.
21. Bennett PC, Gill PS, Silverman S, Blann AD, Balakrishnan B, Lip GY. Ethnic/racial differences in circulating markers of angiogenesis and their association with cardiovascular risk factors and cardiovascular disease. Int J Cardiol 2013;167:1247-50.
22. Kurian AK, Cardarelli KM. Racial and ethnic differences in cardiovascular disease risk factors: a systematic review. Ethn Dis 2007;17:143-52.
23. Coit P, Ognenovski M, Gensterblum E, Maksimowicz-McKinnon K, Wren JD, Sawalha AH. Ethnicity-specific epigenetic variation in naive CD4+ T cells and the susceptibility to autoimmunity. Epigenetics Chromatin 2015;8:49.
24. Xia YY, Ding YB, Liu XQ, et al. Racial/ethnic disparities in human DNA methylation. Biochim Biophys Acta 2014;1846:258-62.
25. Hsiung DT, Marsit CJ, Houseman EA, et al. Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. Cancer Epidemiol Biomarkers Prev 2007;16:108-14.
26. Zhang FF, Cardarelli R, Carroll J, et al. Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. Epigenetics 2011;6:623-9.
27. Record Owner NLM. DNA Methylation Signatures of Depressive Symptoms in Middle-aged and Elderly Persons: Meta-analysis of Multiethnic Epigenome-wide Studies.
28. Fraser HB, Lam LL, Neumann SM, Kobor MS. Population-specificity of human DNA methylation. Genome Biol 2012;13:R8.
29. Giuliani C, Sazzini M, Bacalini MG, et al. Epigenetic Variability across Human Populations: A Focus on DNA Methylation Profiles of the KRTCAP3, MAD1L1 and BRSK2 Genes. Genome Biol Evol 2016;8:2760-73.
30. Amanatidis S, Mackerras D, Simpson JM. Comparison of two frequency questionnaires for quantifying fruit and vegetable intake. Public Health Nutr 2001;4:233-9.
31. Park Y, Dodd KW, Kipnis V, et al. Comparison of self-reported dietary intakes from the Automated Self-Administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers. Am J Clin Nutr 2018;107:80-93.
32. 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:e6.
33. Klipstein-Grobusch K, den Breeijen JH, Goldbohm RA, et al. Dietary assessment in the elderly: validation of a semiquantitative food frequency questionnaire. Eur J Clin Nutr 1998;52:588-96.
34. Goldbohm RA, van den Brandt PA, Brants HA, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. Eur J Clin Nutr 1994;48:253-65.
35. Feunekes GI, Van Staveren WA, De Vries JH, Burema J, Hautvast JG. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. Am J Clin Nutr 1993;58:489-96.
36. Solomon O, MacIsaac J, Quach H, et al. Comparison of DNA methylation measured by Illumina 450K and EPIC BeadChips in blood of newborns and

14-year-old children. Epigenetics 2018;13:655-64.
37. Miranda Furtado CL, Dos Santos Luciano MC, Silva Santos RD, Furtado GP, Moraes MO, Pessoa C. Epidrugs: targeting epigenetic marks in cancer treatment. Epigenetics 2019;14:1164-76.
38. Sommese L, Zullo A, Mancini FP, Fabbricini R, Soricelli A, Napoli C. Clinical relevance of epigenetics in the onset and management of type 2 diabetes mellitus. Epigenetics 2017;12:401-15.

Chapter 6.2.

Summary

SUMMARY

Chapter 1 provides a general introduction on the role of epigenetics, medications and diet as protective or adverse determinants of type 2 diabetes, as well as a description of the studies on which this thesis is based.

In **Chapter 2,** I discuss the available evidence on epigenetics of type 2 diabetes and other cardiometabolic traits, including epigenetic differences between sexes. Specifically, **Chapter 2.1** is a narrative review of the current literature on epigenetic alterations, such as DNA methylation and histone modifications, involved in glucose homeostasis, insulin metabolism and diabetes onset; as well as the potential use of the identified epigenetic patterns as biomarkers in diagnosis and prognosis in clinical practice. Studies on DNA methylation alterations in type 2 diabetes and glycemic traits have used global DNA methylation measurements, candidate gene approaches, and epigenome-wide studies (EWAS) to measure DNA methylation changes in peripheral whole blood, lymphocytes and target tissues such as skeletal muscle, adipose and pancreatic tissue, as well as placenta. Epigenetic signatures, especially on DNA methylation, have been found in association with glucose homeostasis, insulin metabolism and prevalent type 2 diabetes, although they are often inconsistent across different tissues. Histone modifications in association with glycemic traits and type 2 diabetes is an understudied research topic. Evidence on the predictive value of epigenetic biomarkers in diabetes management is limited. However, some studies suggest that epigenetic marks, such as DNA methylation changes, may be a promising tool to complement the current practice on prediction, diagnosis and prognosis of type 2 diabetes. In **Chapter 2.2,** I systematically reviewed sex-specific associations of DNA methylation with intermediate cardiometabolic traits and incident cardiovascular disease. A total of 35 articles based on 30 unique studies and involving a total of 14,020 participants were identified. Overall, they reported sex differences in the associations of epigenome-wide and gene-candidate DNA methylation with blood lipids, stroke risk and cardiometabolic diseases. The identified genes were *PLA2G7, BCL11A, KDM6A, LIPC, ABCG1, PLTP, CETP, ADD1, CNN1B, HOOK2, GFBP-7, PTPN1, GCK, PTX3, ABCG1, GALNT2, CDKN2B, APOE, CTH, GNASAS, INS, PON1, TCN2, CBS, AMT, KDMA6A, FTO, MAP3K13, CCDC8, MMP-2 and ER-α.* Pathway analysis showed that these genes are involved in vitamin B12 metabolism, statin pathway, plasma lipoprotein, plasma lipoprotein assembly, remodeling and clearance and cholesterol metabolism.

Chapter 3 aimed to dissect the association between statin therapy and the incidence of the type 2 diabetes. Particularly, **Chapter 3.1** shows the results of my

study investigating the association between statin treatment and glycemic traits and incident type 2 diabetes, in a prospective population-based cohort study, the Rotterdam Study. Baseline use of statins was associated with higher serum fasting insulin concentrations and insulin resistance, while ever use of statins increased the risk of developing type 2 diabetes by 38%, independent of statin type and dosage, and with higher risk of incident impaired fasting glucose by 9%. The statin-related increased risk of type 2 diabetes was higher among overweight/obese individuals, as compared with subjects with normal body mass index. **Chapter 3.2** presents the findings of on my study aimed to identify DNA methylation signatures associated with statin therapy, and their potential role as mediators of the diabetogenic effect of statins use, using a two-phase epigenome-wide association study (EWAS). Using five cohort studies, I discovered and replicated five DNA methylation sites associated with statin use, annotated to genes *DHCR24, ABCG1* and *SC4MOL*. Higher methylation at *ABCG1* was associated with adverse plasma glycemic traits profile and prevalent type 2 diabetes, as well as with down-regulation of this gene. Evidence from mediation analysis suggested that higher *ABCG1* methylation partially mediates the effect of statin use on insulin and insulin resistance.

Chapter 4 shows the results of the study of coffee dedicated to investigate the associations between coffee consumption, insulin resistance and biomarkers of inflammation, as well as adipokines, and to further explore whether subclinical inflammation response is implicated in the observed association between coffee and type 2 diabetes. Using data from two large population-based cohort studies, I observed that an increase in 1 cup per day in coffee consumption was associated with lower insulin resistance, lower C-reactive (CRP) plasma concentration, as well as with a 4 to 6% lower risk of type 2 diabetes incidence. Mediation analysis showed that coffee-related changes in concentrations of CRP and adiponectin may mediate part of the beneficial effect of higher coffee consumption and lower risk of type 2 diabetes. Stronger associations were observed among women, never smokers, those with overweight and consumers of ground coffee as compared to instant coffee.

In chapter 5 presents the findings of the study on the association of serum uric acid (SUA) levels with all-cause and specific cardiovascular-related mortality, as well as with cardiovascular (CVD)-related nonfatal events, and the effect modification role of type 2 diabetes and sex. The CVD-related fatal and nonfatal events included mortality and incidence of CVD, heart failure (HF), myocardial infarction (MI) and stroke. Among the participants of the UK Biobank, higher SUA concentrations were associated with higher risk of all CVD-related fatal and nonfatal outcomes and all-cause mortality, except for MI. Most of these associations followed a J-shape. Higher

risk of most fatal outcomes tended to be stronger among women than men, and stronger among diabetics than non-diabetics; with the strongest association found among diabetic women. For nonfatal outcomes, SUA-related higher risk of CVD events also tended to be stronger among both diabetic women and men as compared to nondiabetics.

Chapter 6.3.

Nederlandse samenvatting

NETHERLANDSE SAMENVATTING

Hoofdstuk 1 geeft een algemene inleiding over de rol van epigenetica, medicatie en voeding als beschermende of ongunstige determinanten van diabetes type 2 evenals een beschrijving van de studies waarop dit proefschrift is gebaseerd.

In **Hoofdstuk 2** bespreek ik het beschikbare bewijs over de epigenetica van diabetes type 2 en andere cardiometabole kenmerken, inclusief epigenetische verschillen tussen geslacht. **Hoofdstuk 2.1** is een review van de huidige literatuur over epigenetische veranderingen, zoals DNA-methylatie en histon modificaties, betrokken bij glucose-homeostase, insuline metabolisme en het ontstaan van diabetes; evenals het mogelijke gebruik van de geïdentificeerde epigenetische sites als biomarkers voor diagnose en prognose in de klinische praktijk. Studies over DNA-methylatie veranderingen bij diabetes type 2 en glycemische eigenschappen hebben gebruik gemaakt van globaal DNA-methylatie, kandidaat-gen en epigenoom-brede studies (EWAS) om DNA-methylatie veranderingen te meten in perifeer volbloed, lymfocyten en gerelateerd weefsel zoals skeletspieren, vet- en pancreasweefsel evenals de placenta. Epigenetische kenmerken, vooral DNA-methylatie, zijn gevonden in verband met glucose-homeostase, insuline metabolisme en prevalentie van diabetes type 2, echter vaak inconsistent in verschillende weefsels. Histon veranderingen in verband met glycemische eigenschappen en diabetes type 2 is een te weinig onderzocht onderzoeksonderwerp. Bewijs voor de voorspellende waarde van epigenetische biomarkers bij diabetes management is beperkt. Sommige studies suggereren echter dat epigenetische kenmerken, zoals veranderingen in DNA-methylatie, een veelbelovend hulpmiddel kunnen zijn als aanvulling van de huidige praktijk op het gebied van voorspelling, diagnose en prognose van diabetes type 2.

In **Hoofdstuk 2.2** heb ik systematisch de geslacht specifieke associaties van DNA-methylatie met intermediaire cardiometabole kenmerken en incidentele cardiovasculaire aandoeningen bestudeerd. In totaal werden 35 artikelen geïdentificeerd op basis van 30 unieke studies met in totaal 14.020 deelnemers. Over het algemeen rapporteerden ze geslacht specifieke verschillen in de associaties van epigenoom-brede en kandidaat-gen DNA-methylatie met bloedlipiden, het risico op beroerte en cardiometabole ziekten. De geïdentificeerde genen waren *PLA2G7*, *BCL11A*, *KDM6A*, *LIPC*, *ABCG1*, *PLTP*, *CETP*, *ADD1*, *CNN1B*, *HOOK2*, *GFBP-7*, *PTPN1*, *GCK*, *PTX3*, *GALNT2*, *CDKN2B*, *APOE*, *CTH*, *GNASAS*, *INS*, *PON1*, *TCN2*, *CBS*, *AMT*, *KDMA6A*, *FTO*, *MAP3K13*, *CCDC8*, *MMP-2* en *ER-α*. Pathway-analyse toonde aan dat deze genen betrokken zijn bij vitamine B12-metabolisme, de statine-pathway, plasma lipoproteïnen, plasma lipoproteïne assemblage, her-modellering en klaring en cholesterol metabolisme.

Hoofdstuk 3 had als doel de associatie tussen statinetherapie en de incidentie van diabetes type 2 te ontleden. **Hoofdstuk 3.1** laat de resultaten zien van mijn onderzoek naar de associatie tussen statine behandeling en glycemische eigenschappen en diabetes type 2 incidenten, in een prospectieve op populatie-gebaseerde cohortstudie, de Rotterdam Studie. Gebruik van statines bij baseline was geassocieerd met hogere nuchtere insuline concentraties in serum en met insulineresistentie. Daarnaast was ooit gebruik van statines geassocieerd met een 38% verhoogd risico op het ontwikkelen van diabetes type 2 onafhankelijk van het statine type en de dosering en met een 9% hoger risico op incidenteel verminderd nuchtere glucose. Het statine gerelateerde verhoogde risico op diabetes type 2 was hoger bij personen met overgewicht en obesitas, in vergelijking met personen met een normale body mass index.

Hoofdstuk 3.2 presenteert de bevindingen van mijn studie gericht op het identificeren van DNA-methylatie patronen geassocieerd met statinetherapie en hun mogelijke rol als mediatoren van het diabetogene effect van statinegebruik met behulp van een twee-fase epigenoom-brede associatiestudie (EWAS). Met behulp van vijf cohortstudies identificeerde en repliceerde ik vijf DNA-methylatie sites die verband houden met het gebruik van statines en geannoteerd zijn aan de genen *DHCR24*, *ABCG1* en *SC4MOL*. Hogere methylering van *ABCG1* was geassocieerd met een ongunstig plasma glycemische profiel en prevalente diabetes type 2, evenals met downregulatie van dit gen. De resultaten van mediatie-analyse suggereerde dat hogere *ABCG1*-methylatie gedeeltelijk het effect van statine gebruik op insuline en insulineresistentie medieert.

Hoofdstuk 4 toont de resultaten van de koffie studie die is gericht op het onderzoeken van de associaties tussen koffieconsumptie, insulineresistentie en biomarkers van ontsteking, evenals adipokines. Daarnaast om verder te onderzoeken of subklinische ontstekingsreacties betrokken zijn bij de waargenomen associatie tussen koffie en diabetes type 2. Met behulp van de gegevens van twee grote op populatie-gebaseerde cohortstudies, observeerde ik dat een toename van 1 kopje koffie per dag was geassocieerd met een lagere insulineresistentie, een lager C-reactieve (CRP) plasmaconcentratie, evenals met een 4 tot 6% lager risico op de incidentie van diabetes type 2. Mediatie-analyse toonde aan dat koffie gerelateerde veranderingen in de concentraties van CRP en adiponectine een deel van het gunstige effect van een hogere koffieconsumptie en een lager risico op diabetes type 2 kunnen mediëren. Er werden sterkere associaties waargenomen bij vrouwen, niet-rokers, mensen met overgewicht en consumenten van gemalen koffie in vergelijking met oploskoffie.

In **hoofdstuk 5** worden de bevindingen gepresenteerd van de studie naar de associatie van serum urinezuur (SUA)-concentraties met alle oorzaken en specifieke cardiovasculaire gerelateerde sterfte, evenals met cardiovasculair gerelateerde niet-fatale gebeurtenissen en de effect modificerende rol van diabetes type 2 en geslacht. De cardiovasculair gerelateerde fatale en niet-fatale voorvallen omvatten mortaliteit en incidentie van cardiovasculaire aandoeningen, hartfalen, myocardinfarct en beroerte. Onder de deelnemers van de UK Biobank waren hogere SUA-concentraties geassocieerd met een hoger risico op alle cardiovasculair gerelateerde fatale en niet-fatale uitkomsten en met mortaliteit door alle oorzaken, behalve myocardinfarct. De meeste van deze associaties volgden een J-vorm. Een hoger risico op de meeste dodelijke aflopen waren over het algemeen sterker bij vrouwen dan bij mannen en sterker bij diabetici dan bij niet-diabetici; met de sterkste associatie bij vrouwen met diabetes. Voor niet-fatale uitkomsten was het SUA-gerelateerde hogere risico op cardiovasculaire-voorvallen ook sterker bij zowel diabetische vrouwen als mannen in vergelijking met niet-diabetici.

Chapter 7.

Appendices

LIST OF MANUSCRIPTS

Carolina Ochoa-Rosales*, Eralda Asllanaj*, Marija Glisic, Taulant Muka, Oscar H. Franco. *Chromatin landscape and Epigenetic biomarkers for clinical diagnosis and prognosis of Type 2 Diabetes Mellitus*, Prognostic Epigenetics Volume 12, 1st Edition, Elsevier. https://doi.org/10.1016/B978-0-12-814259-2.00012-1

Eralda Asllanaj, **Carolina Ochoa-Rosales***, Xiaofang Zhang*, Jana Nano, Wichor Bramer, Eliana Portilla, Kim Braun, Valentina González-Jaramillo, Wolfgang Ahrens, Arfan Ikram, Mohsen Ghanbari, Trudy Voortman, Oscar Franco, Taulant Muka, Marija Glisic. (2020). Sexually Dimorphic DNA-methylation in Cardiometabolic Health: A Systematic Review. Maturitas 2020;135:6-26. 10.1016/j.maturitas.2020.02.005.

Fariba Ahmadizar, **Carolina Ochoa-Rosales**, Marja Glisic, Oscar Franco, Taulant Muka, Bruno Stricker. (2019). Associations of statin use with glycaemic traits and incident type 2 diabetes. Br J Clin Pharmacol 2019; 85:993-1002. Doi:85. 10.1111/bcp.13898.

Carolina Ochoa-Rosales, Eliana Portilla, Jana Nano, Rory Wilson, Benjamin Lehne, Pashupati Mishra, Xu Gao, Mohsen Ghanbari, Oscar Rueda-Ochoa, Diana Juvinao-Quintero, Marie Loh, Weihua Zhang, Jaspal Kooner, Hans Grabe, Stephan Felix, Ben Schöttker, Yan Zhang, Christian Gieger, Martina Müller-Nurasyid, Margit Heier, Annette Peters, Terho Lehtimäki, Alexander Teumer, Hermann Brenner, Melanie Waldenberger, M. Arfan Ikram, Joyce B.J. van Meurs, Oscar H. Franco, Trudy Voortman, John Chambers, Bruno H. Stricker, Taulant Muka. (2020). *Epigenetic Link Between Statin Therapy and Type 2 Diabetes*. Diabetes Care 2020; 43:875-84. dc191828. 10.2337/dc19-1828.

Carolina Ochoa-Rosales, Niels van der Schaft, Kim Braun, Frederick K. Ho, Fanny Petermann-Rocha, Jill P. Pell, M. Arfan Ikram, Carlos A. Celis-Morales*, Trudy Voortman*. *C-reactive protein partially mediates the inverse association between coffee consumption and risk of type 2 diabetes*. Submitted for publication.

Carolina Ochoa-Rosales, Niels van der Schaft, Frederick K. Ho, Jill P Pell, M. Arfan Ikram, Carlos A. Celis-Morales*, Trudy Voortman* *Type 2 diabetes and sex differences in the association between serum uric acid and risk of fatal and nonfatal cardiovascular related outcomes*. Manuscript.

Jiantao Ma, Casey M Rebholz, Kim V E Braun, Lindsay M Reynolds, Stella Aslibekyan, Rui Xia, Niranjan G Biligowda, Tianxiao Huan, Chunyu Liu, Michael M Mendelson, Roby Joehanes, Emily A Hu, Mara Z Vitolins, Alexis C Wood, Kurt Lohman, **Carolina Ochoa-Rosales**, Joyce van Meurs, Andre Uitterlinden, Yongmei Liu, Mohamed A Elhadad, Margit Heier, Melanie Waldenberger, Annette Peters, Elena Colicino, Eric A Whitsel, Antoine Baldassari, Sina A Gharib, Nona Sotoodehnia, Jennifer A Brody, Colleen M Sitlani, Toshiko Tanaka, W David Hill, Janie Corley, Ian J Deary, Yan Zhang, Ben Schöttker, Hermann Brenner, Maura E Walker, Shumao Ye, Steve Nguyen, Jim Pankow, Ellen W Demerath, Yinan Zheng, Lifang Hou, Liming Liang, Alice H Lichtenstein, Frank B Hu, Myriam Fornage, Trudy Voortman, Daniel Levy. *Whole Blood DNA Methylation Signatures of Diet Are Associated with Cardiovascular Disease Risk Factors and All-cause Mortality*. Circ Genom Precis Med. 2020;10.1161/CIRCGEN.119.002766. doi:10.1161/CIRCGEN.119.002766

Rajiv Chowdhury, Kevin Heng, Md Shajedur Rahman Shawon, Gabriel Goh, Daisy Okonofua, **Carolina Ochoa-Rosales**, Valentina Gonzalez-Jaramillo, Abbas Bhuiya, Daniel Reidpath, Shamini Prathapan, Sara Shahzad, Christian L Althaus, Nathalia Gonzalez-Jaramillo, Oscar H Franco, Global Dynamic Interventions Strategies for COVID-19 Collaborative Group. *Dynamic interventions to control COVID-19 pandemic: a multivariate prediction modelling study comparing 16 worldwide countries*. Eur J Epidemiol 35, 389–399 (2020). doi.org/10.1007/s10654-020-00649-w

Carolina Ochoa-Rosales, Nathalia González-Jaramillo, Aldo Vera-Calzaretta, Oscar H. Franco. Impacto de diferentes medidas de mitigación en el curso de la pandemia de COVID-19 en Chile: proyección preliminar para el período del 14 de abril al 14 de mayo. Revista de Salud Pública, 22(2):1-6, abr. 2020. doi.org/10.15446/rsap.V22n2.86380

Giovanni Fiorito, Cathal McCrory, Oliver Robinson, Cristian Carmeli, **Carolina Ochoa-Rosales**, et al. *Socioeconomic position, lifestyle habits and biomarkers of epigenetic aging: a multi-cohort analysis.* Ageing 2019 vol 11, 2045-2070. Doi: 10.18632/aging.101900

Marija Glisic, Natyra Kastrati, Juna Musa, Jelena Milic, Eralda Asllanaj, Eliana Portilla Fernandez, Jana Nano, **Carolina Ochoa Rosales**, Masoud Amiri, Bledar Kraja, Arjola Bano, Wichor M Bramer, Anton J M Roks, A H Jan Danser, Oscar H Franco, Taulant Muka. *Phytoestrogen supplementation and body composition in postmenopausal women: A systematic review and meta-analysis of randomized controlled trials*. Maturitas. 2018 Sep;115:74-83. doi: 10.1016/j.maturitas.2018.06.012. Epub 2018 Jun 22

Diana L. Juvinao-Quintero, Riccardo E. Marioni, **Carolina Ochoa-Rosales**, Tom C. Russ, Ian J. Deary, Trudy Voortman, Marie-France Hivert, Gemma C. Sharp, Caroline L. Relton, Hannah R. Elliott. *DNA methylation of blood cells is associated with prevalent type 2 diabetes in a meta-analysis of four European cohorts. Submitted for publication.*

Tianxiao Huan, Steve Nguyen, Elena Colicino, **Carolina Ochoa-Rosales**, et al. *An integrative analysis of clinical and epigenetic biomarkers of mortality. Submitted for publication.*

Oscar L. Rueda-Ochoa, Lyda Rojas, Marija Glisic, **Carolina Ochoa-Rosales**, et al. *Mendelian Randomization of DHEAS and NT-Pro-BNP. Submitted for publication.*

* Denotes equal contribution

PHD PORTFOLIO

Name PhD student:	Carolina Patricia Ochoa Rosales
Erasmus MC Department:	Epidemiology
Research School:	NIHES
PhD period:	September 2016 - October 2020
Promotor:	Prof. Dr. Arfan Ikram
Co-promotor:	Dr. Trudy Voortman

1. PhD training	Year	ECTs
General Courses		
Master of Science in Epidemiology, NIHES	2016-2018	70
Specific courses		
EndNote, Medical Library, Erasmus MC	2017	0.3
Introduction to Medical Writing, Erasmus MC	2018	2
Scientific Integrity, Erasmus MC	2019	2
Mendelian Randomization, Cambridge	2019	2
Seminars and meetings		
Nutrition and Lifestyle Epi meetings, Erasmus MC	2016-2020	1
Cardiometabolic Epi meetings, Erasmus MC	2016-2020	1
MolEpi meetings, Erasmus MC	2016-2020	1
Seminars at the Department of Epidemiology, Erasmus MC	2016-2020	1
Seminars at Institute of Cardiovascular and Medical Sciences, University of Glasgow	2019-2020	0.5
Conferences		
Epigenetic of common diseases, Cambridge	2017	1
American Diabetes Association 78th Scientific Sessions, Orlando	2018	1
CHARGE Consortium conference, Rotterdam	2018	0.3
Annual Dutch Diabetes Research Meeting, Oosterbeek	2018	0.3
EuroPrevent, Lisbon	2019	1
ESC Congress 2020	2020	0.3
EASD 56th annual meeting	2020	1

Presentations		
Epigenetic of common diseases, Cambridge – Poster presentation	2017	0.3
American Diabetes Association 78th Scientific Sessions, Orlando – Poster presentation		0.3
Annual Dutch Diabetes Research Meeting, Oosterbeek – Oral presentation	2018	0.3
EuroPrevent, European Society of Cardiology – Oral presentation	2019	0.3
Scholarships and awards		
Albert Renold Travel Fellowship, European Foundation for the Study of Diabetes.	2019	
Young Investigator Award, European Society of Cardiology	2019	
Award to the best MSc and DSc thesis of period 2017-2018, The Netherlands Institute of Health Sciences (NIHES)	2018	
Scholarship Program Doctorado Becas Chile, National Agency for Research and Development (ANID), Chile	2016	
Other		
Reviewer of international peer-reviewed journals: Clinical Nutrition, Revista Española de Nutrición Humana y Dietética, PLOS ONE	2018-2020	1
Vising scientist at Institute of Cardiovascular and Medical Sciences, University of Glasgow	2019-2020	1
2. Teaching		
Supervising students		
Amber Meulenbeld, MSc thesis	2020	2
Rianne Smit, MSc thesis	2020	2
Supervising exercises		
Teaching assistant in Biostatistical Methods I: basic principles, NIHES	2018, 2020	1

ABOUT THE AUTHOR

Carolina Patricia Ochoa Rosales was born on October 29th 1986 in Valparaíso, Chile. She studied Biochemistry at Universidad de Concepción. In this period, she completed internships in the Regional Hospital of Concepción and at the Genetics and Biochemistry units at the Forensic laboratory at the Forensic Medical Service in Concepción. She obtained her degree as Biochemist in 2012. After that, Carolina worked as a teacher in a Chemistry and Biology at Universidad de las Américas, and next as a research assistant on healthy lifestyles and metabolic health at the Centro de Vida Saludable, Universidad de Concepción. In this period, she also obtained a Master degree on Health Administration and Public Health at the Universidad del Desarrollo, in 2016.

Carolina's greatest research interest is to combine Genetic and Molecular Biology sciences with Population Studies, to contribute with evidence for future health policies. In order to pursuit this interest, Carolina decided to continue with a Master of Science in Genetic Epidemiology at the Institute of Health Sciences (NIHES), at the Erasmus University Medical Center (EMC), Rotterdam, the Netherlands, which she completed in 2018. She expanded her work at the Department of Epidemiology of EMC as a PhD student, conducting research on Epigenomics, Nutrition and Lifestyle, Cardiometabolic health and Molecular Epidemiology. During the last year of her PhD program, Carolina worked as a visiting scientist at the Institute of Cardiovascular and Medical Sciences, University of Glasgow, Scotland, United Kingdom.

The results of her PhD research, entitled *"Disentangling the underlying mechanisms linking epigenetic, metabolic and environmental determinants of type 2 diabetes"*, are presented in this thesis.

DANKWOORD

Thanks to my Promotors and Co-promotors, the current and the past. Thanks to Dr. Trudy Voortman, Dr. Taulant Muka, Prof. Dr. Oscar Franco and Prof. Dr. Arfan Ikram, for leading me, for being so supportive and for encouraging me to go further, for being my mentors. I have learnt so much from each one of you and I have so much to thank you for. I am happy I have the opportunity to use these pages to express my gratitude. Dear Oscar, thanks for accepting me and welcoming me at EMC in the first place, for the guidance and support during the time we worked together at EMC and after that; not only to me, but to Sr. Bateman as well. Dear Taulant, thanks for teaching me how to be a researcher, for always being there for me, replying emails in evenings and weekends in order to accompany me and support me in my first steps. Thanks for trusting my work, for pushing me to go to conferences and for making me build self-confidence in my own work and for teach me how to become and independent researcher. Dear Trudy, thanks for receiving me in your research group and being my co-promotor. Thanks for letting me be, in terms of research and analyses and supervision of master students, for that's a sign that you trust me as a researcher and my work, hence I feel now more self-confident. Thanks for supporting me in every challenge I wanted to go for and for open up opportunities for me to grow as a researcher. Thanks for being my reference of a successful woman researcher, not only for your great professional accomplishments, but also for being so caring, understanding and committed to your students, even if it means working on evenings and weekends. Dear Arfan, thanks for your support on expanding my opportunities in the research path. It will for sure positively impact my future academic career, and I very appreciate it. Dear Dr. Carlos and dear prof. Dr. Jill, thanks for welcoming me into their research group at the University of Glasgow and for taking part in my PhD research and defense process. Thanks Carlos for being so supporting and for showing that enthusiasm and willingness to set collaborations.

Thanks to my Paranymphs Silvana and Banafsheh, being such a great support throughout this process, but also for your honest friendship and care. You have been not only the most awesome, helpful and organized Paranymphs a candidate can have, but also the greatest fiends.

Thanks to Mirjam, for your help, kindness and commitment, always willing to assist me and anyone who needed it. Thanks to Nano, for your help every time I had an IT problem.

Thanks to all the participants of the Rotterdam Study, the staff at the Ergocenter and at EMC who make the continuity of the cohort possible and take care of the data so we can work, specially to Jolande and Frank. Without your work and the whole team, the development of this thesis and thesis of all students would not have been possible.

Many researchers were involved in the work that gave rise to this thesis. I want to thank to my coauthors: Eralda, Niels, Rory, Mohsen, Kim, Jana, Fariba, Fanny, Fred, Arfan, Oscar, Joyce, Carlos, Jill, Bruno, Taulant, Trudy, and so many more who I cannot fit in this page. Your collaboration was key in the development of this thesis; your comments made the papers become 10 times better. You have taught me a lot and I highly appreciate that. But collaborations did not only involve the Rotterdam Study. Several other cohorts participated in the development of this thesis. Thanks to the participants, staff and researchers from the UK Biobank, KORA, EPIC, LOLIPOP, ESTHER and ALSPAC. Also thanks to those who invited me to collaborate in their own projects, such as those in CHARGE consortium.

Thanks to the teachers during the master courses, specially to prof. Oscar Franco, prof. Arfan Ikram, prof. Dimitris Rizopoulous, Dr. Najav Amin. I had no background in epidemiology when I just started the program, but thanks to your lectures I was able to understand and conduct the research that led to this thesis. It was an honor to be at your classes and learn from you, who are recognized as references in your respective fields worldwide.

Thanks to all my colleagues at the department, for I have learnt something from all of you, in work related things but also as a person; you even helped me improving my English skills. We shared group meetings, cookie breaks, chats at the coffee corner, Christmas dinners, lunch at Het park and drinks after work. We have also laughed and cried together, supported each other during this journey, in the good and the bad, work and personally related. I know I have found friends for life. For this I, and I'm sure Ale also, want to specially thank you Silvana, Irma, Banafsheh, Arash, Cony, Oscar, Lyda, Hamid, Anh Nhi, Alice, Palo and Dino, for your friendship and the great moments shared.

No puedo dejar de hacer una mención especial al grupo latinoamericano en EMC, colegas que muy pronto se transformarían en grandes amigos. Gracias por acogernos y hacernos sentir como en casa. Gracias a Oscar, Lyda, Cony, Eliana, Magda, Humberto, Palo, Dino y Valentina por el apoyo en horas de trabajo y en horas de fiesta. Gracias a Oscar y Magda por otorgarme el honor de ser su paraninfa. Gracias a todo

el team Colombia por la bandeja paisa, buñuelos, arepas, plátano frito, chicharrón, chocolate caliente, buen café y por bailar la mayonesa.

También quiero agradecer a mi mentora y colega en la UdeC, la profesora Dra. Natalia Ulloa Muñoz. Profe Naty, gracias por el cariño, por estar todo este tiempo pendiente y por incentivarme para el desarrollo de mi carrera académica.

Last but not least, un gran GRACIAS a mi familia, así con mayúsculas. A mi más grande amigo y compañero, Alejandro, alias Ale, Janufer, Alejo o Sr. Bateman. Gracias por ser el más apañador, comprensivo y regaloneador; por el ser el mejor compañero en todas las aventuras y desafíos, y por ser mi más grande apoyo durante el recorrido de este doctorado. Este logro también te pertenece. Gracias a mis padres, Patricia y Mario, por su apoyo incondicional desde siempre, por hacerme sentir segura de mí misma y por repetirme que todo lo que me proponga lo puedo lograr. Gracias por apoyarme en todos los desafíos académicos que me he puesto por delante; y esta vez, aunque separados por un océano, no fue la excepción.

Once more, thanks to all!

ACKNOWLEDGMENTS

Carolina Ochoa-Rosales' research received funding from the National Agency for Research and Development (ANID) / Scholarship Program Doctorado Becas Chile/2016 – 72170524. Further financial support was provided by the European Foundation for the Study of Diabetes (EFSD), via the Albert Renold Travel Fellowship Programme.

The Rotterdam Study is supported by grants from the Municipality of Rotterdam, the NESTOR program for research in the elderly (Ministry of Health and Ministry of Education), the Netherlands Prevention Fund, the Netherlands Organization for scientific research (NWO), the Netherlands Heart Foundation and the Rotterdam Medical Research Foundation (ROMERES).

PROPOSITIONS

1. Epigenetic signatures could serve as biomarkers for early diagnosis or disease progression in type 2 diabetes, and as novel targets for the design of new therapies (this thesis).

2. Sex-specific epigenetic changes may explain sex differences in cardiometabolic health (this thesis).

3. Statins treatment increase type 2 diabetes risk, which is partially mediated by methylation changes at gene *ABCG1* (this thesis).

4. The protective effect of coffee consumption against development of type 2 diabetes is partially explained by modulation of the inflammatory response (this thesis).

5. High uric acid levels are a risk factor for premature mortality and cardiovascular disease, especially for women and type 2 diabetes patients (this thesis).

6. A person's unique epigenome explains what genes alone cannot tell. Therefore, it should play a key role in precision medicine.

7. Through lifestyle changes, we have the power to lead our genes.

8. The observed health benefits of coffee consumption suggest that coffee should be a preferred beverage to take us throughout the day and to keep us healthy throughout life.

9. "Science and everyday life should not be separated." Rosalind Franklin, chemist who provided evidence of the molecular structure of DNA.

10. "There is no problem in science that can be solved by a man that cannot be solved by a woman." Vera Rubin, astrophysicist who discovered evidence of dark matter.

11. "Coming together is the beginning. Keeping together is progress. Working together is success." Henry Ford, developer of the assembly line technique of mass production.