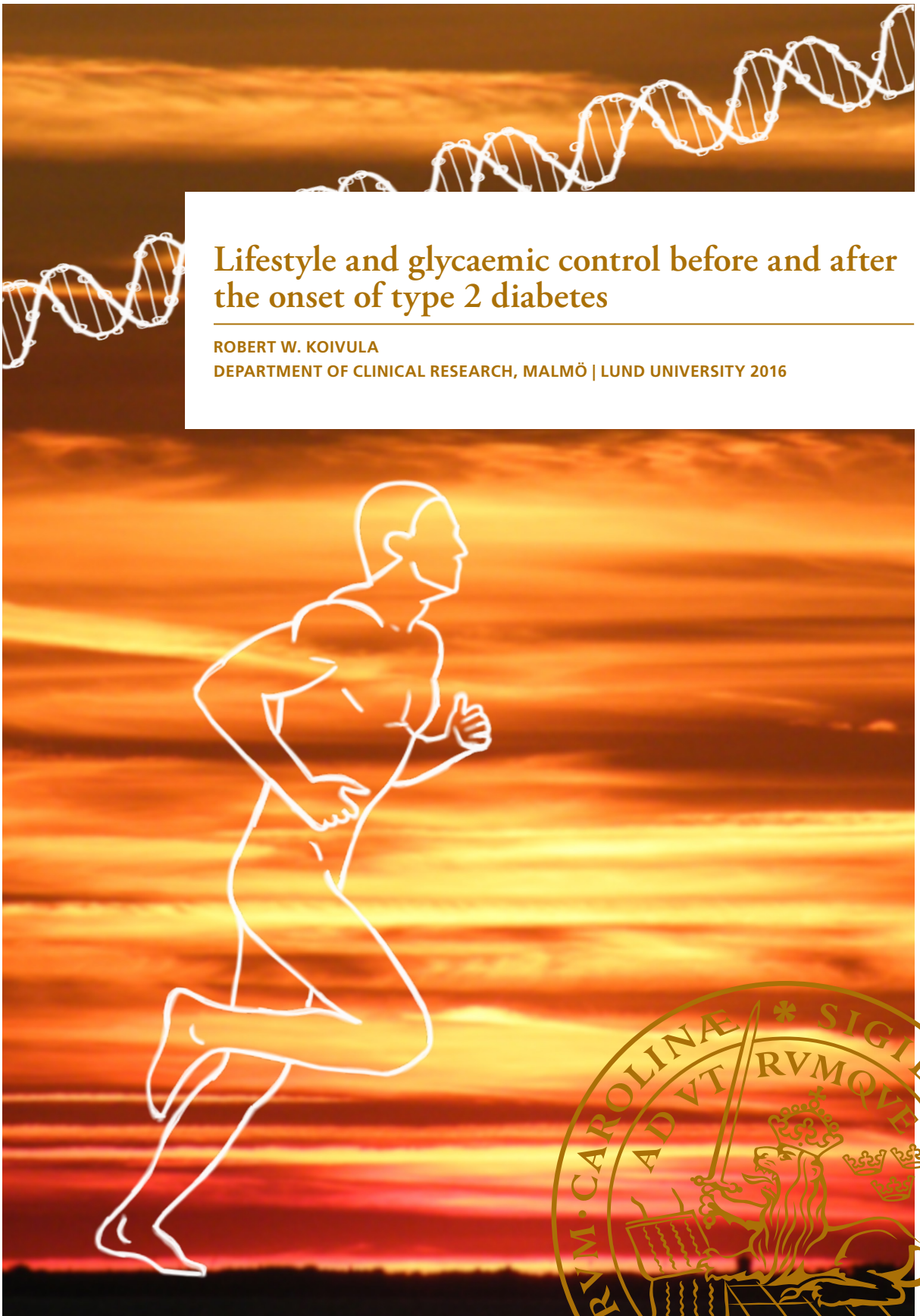
Lifestyle and glycaemic control before and after the onset of type 2 diabetes

Koivula, Robert

2016

Document Version:
Publisher's PDF, also known as Version of record



Lifestyle and glycaemic control before and after the onset of type 2 diabetes

ROBERT W. KOIVULA
DEPARTMENT OF CLINICAL RESEARCH, MALMÖ | LUND UNIVERSITY 2016

Lifestyle and glycaemic control before and after the onset of type 2 diabetes

Robert Wilhelm Koivula

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty Medicine, Lund University, Sweden.
To be defended at 'Main Lecture Hall' Clinical Research Centre, Skåne University Hospital Malmö. Date 23rd November at 13:00

Faculty opponent
Professor Roy Taylor, Newcastle University, UK.

Organization LUND UNIVERSITY	Document name: DOCTORAL DISSERTATION
	Date of issue: November 23rd 2016
Author(s): Robert Wilhelm Koivula	Sponsoring organization: IMI-DIRECT, Novo Nordisk

Title and subtitle: Lifestyle and glycaemic control before and after the onset of type 2 diabetes

Abstract:

Type 2 diabetes (T2D) is a complex disease with widespread physiological insults to the regulation of metabolic homeostasis, above all glycaemic regulation. The pathogenesis of T2D and its progression is broadly understood to be through a gradual decrease in peripheral insulin sensitivity, a compensatory rise in insulin secretion, and a gradual decline in beta-cell function, resulting in glycaemic dysregulation and eventual T2D. Unhealthful lifestyle factors such as low physical activity (PA), energy dense diets with poor nutritional value, and chronic positive calorie balance are associated with an accelerated decline in glycaemic control and obesity. Single nucleotide polymorphisms (SNPs) discovered in genome wide association studies (GWAS) have demonstrated a genetic susceptibility to T2D (65 SNPs), fasting glucose (36 SNPs), 2-hr glucose (9 SNPs), and obesity (97 SNPs).

In Paper 1, in 3,444 adults in northern Sweden, we compared the predictive ability of lifestyle factors and the aforementioned SNPs with T2D and obesity for incidence of T2D, impaired fasting glucose (IFG), impaired glucose tolerance (IGT) and obesity over ~10 years. We found that lifestyle and genetic factors had broadly the same predictive ability (by ROC AUC) for incidence in T2D (75% vs. 74%), IFG (63% vs. 66%), IGT (64% vs. 61%) and obesity (68% vs. 73%). With exception of IGT, adding genetic factors to lifestyle models improved predictive ability with resulting continuous net reclassification improvements of 58%, 36% and, 64% for T2D, IFG and obesity, respectively.

In paper 2 and 3, we overview the rationale, design and baseline results from two new prospective cohort studies within the DIRECT (Diabetes Research into Patient Stratification) Consortium. These cohort studies aim to improve prevention and treatment of T2D by discovering new biomarkers for glycaemic deterioration before (Cohort 1, N =2,335) and after the onset of T2D (Cohort 2, N =830). The cohorts are comprehensively assessed at follow-up visits at 18, 36 (Cohort 2) and 48 (Cohort 1) months. Aside from standard measurements, clinical assessments include: Beta-cell function, insulin sensitivity, glycaemia, objective PA, diet, abdominal MRI, genomic, transcriptomic, metabolomic, proteomic and microbiomic assessments.

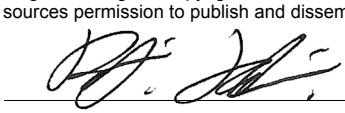
A potential mechanism behind the pathogenesis of T2D has been hypothesised in a 'twin-cycle' model. In the first 'liver cycle', peripheral insulin resistance, in combination with a positive caloric balance, leads to accumulating liver fat. The hyperinsulinaemia from insulin resistance leads to an increase in hepatic de-novo lipogenesis, further increasing liver fat. This leads to reduced insulin-mediated suppression of gluconeogenesis, further increasing glycaemia, insulinaemia and accelerating the liver cycle. The increased very low lipoprotein secretion from lipogenesis eventually increases ectopic triglycerides in surrounding tissue, such as the pancreas. This feeds a 'pancreatic cycle' where lipotoxicity and glucotoxicity reduces beta-cell function and postprandial glucose, further accelerating the liver-cycle.

In paper 4, I test the 'twin-cycle' hypothesis and if the association between PA and glycaemic control is mediated by parameters within the model. Using structural equation modelling in newly gathered DIRECT baseline dataset, I find that: most of the relationships in the 'twin-cycle' hypothesis are observed as hypothesised in both prediabetes and T2D; that PA is associated with most parameters in the model; and, that the association of PA with glycaemic control and liver fat is mostly mediated by whole body insulin sensitivity.

Key words: Type 2 Diabetes, Prediabetes, Obesity, Beta-cell Function, Insulin Sensitivity, Physical Activity, Exercise, Diet, Lifestyle, Ectopic Fat, Genetic Epidemiology, Genetic Risk, Biomarker, Precision Medicine

Classification system and/or index terms (if any)

Supplementary bibliographical information		Language: English
ISSN and key title: 1652-8220		ISBN: 978-91-7619-365-5
Recipient's notes	Number of pages: 102	Price
	Security classification	

Signature  Date : 19th Oct 2016

Lifestyle and glycaemic control before and after the onset of type 2 diabetes

Robert Wilhelm Koivula

LUND
UNIVERSITY

Cover photo:

‘Gym guy’ in the sunset, Västra Hamnen, Malmö by Robert Koivula, 2016

Faculty of Medicine
Department of Clinical Research, Malmö

ISSN 1652-8220
ISBN 978-91-7619-365-5
Lund University, Faculty of Medicine Doctoral Dissertation Series 2016:138

Printed in Sweden by Media-Tryck, Lund University
Lund 2016

Don't worry about genius and don't worry about not being clever. Trust rather to hard work, perseverance, and determination. The best motto for a long march is "Don't grumble. Plug on."

You hold your future in your own hands. Never waver in this belief. Don't swagger. The boy who swaggers – like the man who swaggers – has little else that he can do. He is a cheap-Jack crying his own paltry wares. It is the empty tin that rattles most. Be honest. Be loyal. Be kind. Remember that the hardest thing to acquire is the faculty of being unselfish. As a quality it is one of the finest attributes of manliness.

Love the sea, the ringing beach and the open downs.

Keep clean, body and mind.

-Sir Frederick Treves, Bart, KCVO, CB Serjeant Surgeon to HM the King, Serjeant Surgeon to HRH Prince of Wales, written at 6 Wimpole Street, Cavendish Square, London on 2 September 1903, on the occasion of the twenty-fifth anniversary of the Boy's Own Paper.

To my parents, Bettina and Ulf Koivula

Content

Thesis articles

Paper 1: Poveda A, **Koivula RW**, Ahmad S, Barroso I, Hallmans G, Johansson I, Renström F, Franks PW. Innate biology versus lifestyle behaviour in the aetiology of obesity and type 2 diabetes: the GLACIER Study. Diabetologia. 2016 Mar;59(3):462-71. doi: 10.1007/s00125-015-3818-y. Epub 2015 Dec 1. PubMed PMID: 26625858; PubMed Central PMCID: PMC4742501

Paper 2: **Koivula RW**, Heggie A, Barnett A, Cederberg H, Hansen TH, Koopman AD, Ridderstråle M, Rutters F, Vestergaard H, Gupta R, Herrgård S, Heymans MW, Perry MH, Rauh S, Siloaho M, Teare HJ, Thorand B, Bell J, Brunak S, Frost G, Jablonka B, Mari A, McDonald TJ, Dekker JM, Hansen T, Hattersley A, Laakso M, Pedersen O, Koivisto V, Ruetten H, Walker M, Pearson E, Franks PW; DIRECT Consortium. Diabetologia. 2014 Jun;57(6):1132-42. doi: 10.1007/s00125-014-3216-x. Epub 2014 Apr 4. PubMed PMID: 24695864; PubMed Central PMCID: PMC4018481.

Paper 3: **Koivula RW**, Forgie I, Heggie A, Hansen Tue, Hudson M, Koopman A, Rutters F, Siloaho M, Brage S, Dawed AY, Ford H, Groves CJ, Mahajan A, Perry MH, Rauh SP, Ridderstråle M, Teare HJA, Tura A, Vestergaard H, White T, Dekker J, Hansen Torben, Hattersley A, Laakso M, Pedersen O, Bell J, Brunak S, Froguel P, Frost G, Gupta R, Jablonka B, McDonald TJ, Pavo I, Mari A, Walker M, McCarthy MI, Ruetten H, Pearson E, Franks PW, for the IMI DIRECT consortium. Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: A description of the baseline data from the epidemiological studies of the IMI DIRECT Consortium. (Manuscript)

Paper 4: **Koivula RW**, Atabaki-Pasdar N, Mari A, Kurbasic A, White T, Frost G, Brage S, Bell J, Walker M, Ruetten H, Pearson E, Pavo I, Franks PW, for the IMI DIRECT consortium. The role of physical activity in metabolic homeostasis before and after the onset of type 2 diabetes: an IMI DIRECT study (Manuscript)

Articles not included in thesis

Koivula RW, Tornberg AB, Franks PW. Exercise and diabetes-related cardiovascular disease: systematic review of published evidence from observational studies and clinical trials. Curr Diab Rep. 2013 Jun;13(3):372-80. doi: 10.1007/s11892-013-0373-0. Review. PubMed PMID: 23494754.

Ahmad S, Rukh G, Varga TV, Ali A, Kurbasic A, Shungin D, Ericson U, **Koivula RW**, Chu AY, Rose LM, Ganna A, Qi Q, Stančáková A, Sandholt CH, Elks CE, Curhan G, Jensen MK, Tamimi RM, Allin KH, Jørgensen T, Brage S, Langenberg C, Aadahl M, Grarup N, Linneberg A, Paré G; InterAct Consortium; DIRECT Consortium, Magnusson PK, Pedersen NL, Boehnke M, Hamsten A, Mohlke KL, Pasquale LT, Pedersen O, Scott RA, Ridker PM, Ingelsson E, Laakso M, Hansen T, Qi L, Wareham NJ, Chasman DI, Hallmans G, Hu FB, Renström F, Orho-Melander M, Franks PW. Gene × physical activity interactions in obesity: combined analysis of 111,421 individuals of European ancestry. PLoS Genet. 2013;9(7):e1003607. doi: 10.1371/journal.pgen.1003607. Epub 2013 Jul 25. PubMed PMID: 23935507; PubMed Central PMCID: PMC3723486.

Varga TV, Sonestedt E, Shungin D, **Koivula RW**, Hallmans G, Escher SA, Barroso I, Nilsson P, Melander O, Orho-Melander M, Renström F, Franks PW. Genetic determinants of long-term changes in blood lipid concentrations: 10-year follow-up of the GLACIER study. PLoS Genet. 2014 Jun 12;10(6):e1004388. doi: 10.1371/journal.pgen.1004388. eCollection 2014 Jun. PubMed PMID: 24922540; PubMed Central PMCID: PMC4055682.

Renström F, **Koivula RW**, Varga TV, Hallmans G, Mulder H, Florez JC, Hu FB, Franks PW. Season-dependent associations of circadian rhythm-regulating loci (CRY1, CRY2 and MTNR1B) and glucose homeostasis: the GLACIER Study. Diabetologia. 2015 May;58(5):997-1005. doi: 10.1007/s00125-015-3533-8. Epub 2015 Feb 24. PubMed PMID: 25707907.

Kato N, Loh M, Takeuchi F, Verweij N, Wang X, Zhang W, Kelly TN, Saleheen D, Lehne B, Mateo Leach I, Drong AW, Abbott J, Wahl S, Tan ST, Scott WR, Campanella G, Chadeau-Hyam M, Afzal U, Ahluwalia TS, Bonder MJ, Chen P, Dehghan A, Edwards TL, Esko T, Go MJ, Harris SE, Hartiala J, Kasela S, Kasturiratne A, Khor CC, Kleber ME, Li H, Mok ZY, Nakatochi M, Sapari NS, Saxena R, Stewart AF, Stolk L, Tabara Y, Teh AL, Wu Y, Wu JY, Zhang Y, Aits

I, Da Silva Couto Alves A, Das S, Dorajoo R, Hopewell JC, Kim YK, **Koivula RW**, Luan J, Lyytikäinen LP, Nguyen QN, Pereira MA, Postmus I, Raitakari OT, Bryan MS, Scott RA, Sorice R, Tragante V, Traglia M, White J, Yamamoto K, Zhang Y, Adair LS, Ahmed A, Akiyama K, Asif R, Aung T, Barroso I, Bjonnes A, Braun TR, Cai H, Chang LC, Chen CH, Cheng CY, Chong YS, Collins R, Courtney R, Davies G, Delgado G, Do LD, Doevendans PA, Gansevoort RT, Gao YT, Grammer TB, Grarup N, Grewal J, Gu D, Wander GS, Hartikainen AL, Hazen SL, He J, Heng CK, Hixson JE, Hofman A, Hsu C, Huang W, Husemoen LL, Hwang JY, Ichihara S, Igase M, Isono M, Justesen JM, Katsuya T, Kibriya MG, Kim YJ, Kishimoto M, Koh WP, Kohara K, Kumari M, Kwek K, Lee NR, Lee J, Liao J, Lieb W, Liewald DC, Matsubara T, Matsushita Y, Meitinger T, Mihailov E, Milani L, Mills R, Mononen N, Müller-Nurasyid M, Nabika T, Nakashima E, Ng HK, Nikus K, Nutile T, Ohkubo T, Ohnaka K, Parish S, Paternoster L, Peng H, Peters A, Pham ST, Pinidiyapathirage MJ, Rahman M, Rakugi H, Rolandsson O, Rozario MA, Ruggiero D, Sala CF, Sarju R, Shimokawa K, Snieder H, Sparsø T, Spiering W, Starr JM, Stott DJ, Stram DO, Sugiyama T, Szymczak S, Tang WH, Tong L, Trompet S, Turjanmaa V, Ueshima H, Uitterlinden AG, Umemura S, Vaarasmaki M, van Dam RM, van Gilst WH, van Veldhuisen DJ, Viikari JS, Waldenberger M, Wang Y, Wang A, Wilson R, Wong TY, Xiang YB, Yamaguchi S, Ye X, Young RD, Young TL, Yuan JM, Zhou X, Asselbergs FW, Ciullo M, Clarke R, Deloukas P, Franke A, Franks PW, Franks S, Friedlander Y, Gross MD, Guo Z, Hansen T, Jarvelin MR, Jørgensen T, Jukema JW, Kähönen M, Kajio H, Kivimaki M, Lee JY, Lehtimäki T, Linneberg A, Miki T, Pedersen O, Samani NJ, Sørensen TI, Takayanagi R, Toniolo D; BIOS-consortium; CARDIo GRAMplusCD; LifeLines Cohort Study; InterAct Consortium, Ahsan H, Allayee H, Chen YT, Danesh J, Deary IJ, Franco OH, Franke L, Heijman BT, Holbrook JD, Isaacs A, Kim BJ, Lin X, Liu J, März W, Metspalu A, Mohlke KL, Sanghera DK, Shu XO, van Meurs JB, Vithana E, Wickremasinghe AR, Wijmenga C, Wolffenbuttel BH, Yokota M, Zheng W, Zhu D, Vineis P, Kyrtopoulos SA, Kleinjans JC, McCarthy MI, Soong R, Gieger C, Scott J, Teo YY, He J, Elliott P, Tai ES, van der Harst P, Kooner JS, Chambers JC. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. Nat Genet. 2015 Nov;47(11):1282-93. doi: 10.1038/ng.3405. Epub 2015 Sep 21. PubMed PMID: 26390057; PubMed Central PMCID: PMC4719169.

Grøntved A, **Koivula RW**, Johansson I, Wennberg P, Østergaard L, Hallmans G, Renström F, Franks PW. Bicycling to work and primordial prevention of cardiovascular risk: a cohort study among Swedish men and women. J Am Heart Assoc. 2016 (in press)

Abbreviations

BMI:	Body mass index
CFI:	Comparative Fit Index
cIGR:	Combined impaired glucose regulation
CVD:	Cardiovascular Disease
DanFunD:	Danish Functional Disability study
DIRECT:	Diabetes Research on Patient Stratification Consortium
DPP:	Diabetes prevention program (American DPP)
DPS:	Finnish diabetes prevention study (Finnish DPP)
ENMO:	Euclidean norm minus one (a direction agnostic estimate of physical activity intensity)
FFA:	Free fatty acid
FFQ:	Food Frequency Questionnaire
GGG:	Gut, Grain and Greens study
GLACIER:	Gene-Lifestyle interactions And Complex traits Involved in elevated disease Risk
GRS:	Genetic Risk Score
GWAS:	Genome-wide association study
HbA1c:	Glycated haemoglobin A1c
HD:	Healthy Diet Score
HMS:	Hoorn Meal Study
hpfVM:	high-pass-filtered vector magnitude (a direction agnostic estimate of physical activity intensity)
IA1c:	Impaired Hba1c
IFG:	Impaired fasting glucose

IGT:	Impaired glucose tolerance
iIA1c:	isolated impaired A1c
iIFG:	Isolated impaired fasting glucose
iIGT:	Isolated impaired glucose tolerance
MDC:	Malmö Diet and Cancer study
METSIM:	Metabolic Syndrome in Men study
MFS:	Malmö feasibility study (Swedish DPP)
NGR:	Normal glucose regulation
NHS:	New Hoorn Study
NNR:	Nordic Nutrition Recommendation
OGTT:	Oral glucose tolerance test (‘fs’ denotes frequently sampled)
PA:	Physical Activity
PCA:	Principal Components Analysis
RMSEA:	Root mean standardised error of approximation
SEM:	Structural equation model
SNP:	Single nucleotide polymorphism
SRMR:	Standardised root mean square residual
T2D:	Type 2 diabetes
TLI:	Tucker-Lewis Index
VHS:	Västerbotten Health Survey
VIP:	Västerbotten Intervention Program

Introduction

Epidemiology is the study of health and disease outcomes relating to potential causal factors in a population setting. Discoveries of epidemiology are used to guide further experimental research, improve aetiological understanding of diseases, as well as to guide public health policy. Famously, John Snow found that the centre of the London cholera epidemic in 1854 was likely to be the water pump on Broad street (see Figure 1). He came to this conclusion after carefully interviewing local residents and finding an association with the incidence of cholera and the proximity to, and later use of water from that public pump. When the authorities, informed by his findings, later removed the handle of the pump, the incidence of cholera subsided. For this, he is often considered the 'father' of modern epidemiology[1]. Since then, methods in epidemiology have advanced but the principal remains the same: to make causal inferences about exposures and outcomes in defined populations. In the case of this thesis, the main outcome of interest is type 2 diabetes (T2D), the exposures are lifestyle, adiposity, and genetic factors, with the populations under investigation being cohorts from northern Europe.

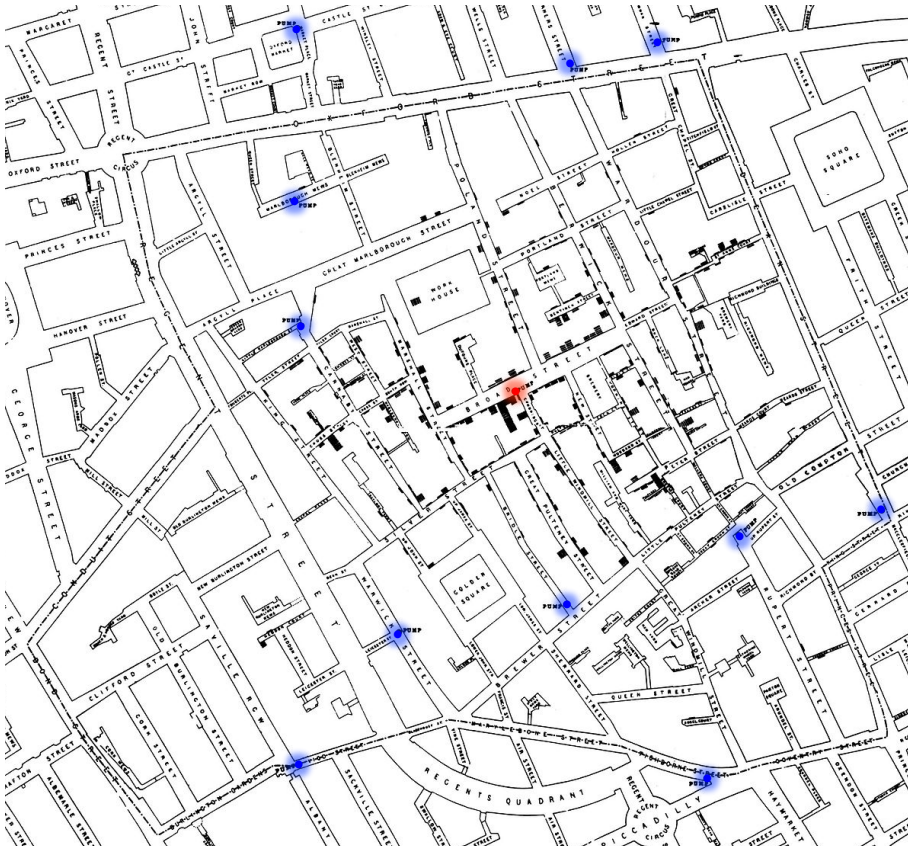


Figure 1. London 1854, Cholera incidence surrounding Broad Street.
Red dot: Broad street water pump, blue dots: other water pumps, black shapes: properties with incident cholera. Orginial map published by John Snow[2].

Background

T2D is a 'complex' disease with many identifiable physiological insults that change with time and vary between people.

T2D is diagnosed by its primary symptom of elevated blood glucose. T2D is characterised by a gradual decline in insulin sensitivity, which (for a while, at least) allows maintenance of glycaemic control by a compensatory rise in endogenous insulin secretion[3]. Type 1 diabetes on the other hand, is characterised by an autoimmune destruction of pancreatic beta-cells leading to an inability to endogenously secrete insulin but is not in the scope of this thesis. T2D is also characterised by a gradual loss of endogenous insulin secretion (beta-cell

function), which eventually leads to the loss of glycaemic control and the diagnosis of T2D requiring treatment. The first line of treatment, according to guidelines from the International Federation of Diabetes which reflect common practice[4], is 'lifestyle measures'. If glycaemic control is not achieved by lifestyle changes, the next level is treatment with various pharmaceutical agents aimed to increase either insulin sensitivity or boost endogenous insulin production. If this does not effectively improve glycaemic control, insulin treatment with exogenous insulin is required. Indeed, 25% of newly diagnosed T2D cases go on to exogenous insulin treatment within six years[5]. The aim of such treatment (with the exception of lifestyle interventions) is primarily symptom-, rather than primordial-cause orientated; i.e. pharmaceutical treatment attempts to restore identifiable physiological insults to glycaemic control rather than addressing the underlying causes of those insults.

We are exposed to a vast number of potential primordial causal factors, often referred to as the 'exposome'[6]. Coupled with numerous individual genetic differences, it seems unlikely that there is a single primordial cause for T2D but rather a combination of causal factors[7,8]. This is an important point because whatever causes are at the origin of the pathogenesis of T2D, they clearly affect multiple factors, which in turn affect other factors (and so on), leading to a heterogeneous, complex disease. So complex in fact that, despite several subtypes of diabetes already identified, T2D makes up 90% of them. This reflects the fact that T2D is essentially the diagnosis reached when all other distinct causes of diabetes subtypes have been ruled out.

Globally, the prevalence of T2D today is estimated at around 8.8% of adults, about 415 million people[9]. The prevalence of diagnosed diabetes in the USA in 1980 was estimated to be around 0.6%, 5.4% and 9.7% and 8.6% in people aged 0-44, 45-65, 65-74, and 75+ years, respectively. These numbers increased to 1.5%, 12.0%, 21.5% and 19.2% in 2014[10]. The rising prevalence has been attributed to the parallel increase in obesity[11] and 'westernised' lifestyles, comprising of excessive consumption of energy-dense diets with poor nutritional content, in combination with inadequate levels of habitual physical activity. However, a major driver of this increase in prevalence is likely to be better disease detection and that people now live longer with diabetes. The increasing prevalence of T2D has led to a spiralling economic and social burden of treating diabetes complications, early loss of life and reduced quality of life. With no highly effective means to prevent or treat T2D, and no cure (in part, due to our lack of knowledge about the disease's aetiology and mechanisms of action), there is a clear and urgent need for research that addresses these insufficiencies.

The 'exposome' mentioned earlier is a universal collection of exposures; the focus of this thesis, as the title suggests, will be exposure to 'lifestyle factors'.

'Lifestyle' includes all factors describing our habitual behaviour on a day-to-day basis. In this thesis however, I will focus primarily on physical activity (PA) and to a lesser extent diet. I do not consider the two in isolation because they are so closely linked in our daily lives. For example, the 'inactivity' of TV watching is related to snacking[12]. Similarly, participants who are more physically active also eat more[13]. The Lancet Physical Activity Series Working Group estimated that worldwide, physical inactivity accounted for 7% of the burden of T2D[14], accruing around 53.8 billion US$ in healthcare costs in 2013[15].

Below is a brief background describing the role of lifestyle factors in the genetic epidemiology of T2D. I shall then describe the role of lifestyle and adiposity in the epidemiology of T2D and their role in experimental studies of T2D.

Role of lifestyle factors in genetic epidemiology of type 2 diabetes

Family and twin studies have noted a significant genetic component to the susceptibility of T2D through heritability studies, where 'broad sense heritability' is defined as: H^2 = total genetic variance ÷ total phenotypic variance and 'narrow sense heritability' is h^2 = additive genetic variance ÷ total phenotypic variance. Heritability studies in twins have estimated the heritability h^2 to be ~72% and sibling-based heritability estimates have been reported h^2 to be ~69%[16,17]. However, in older adults this may be less, as a study of ~5800 Finns from around ~940 families estimated a h^2 of ~69% in 35-60 years, but only ~31% when ages up to 75 years were include[18].

From 2005, a major advance in the field of genetic epidemiology was the genome-wide association study (GWAS). In these studies, population cohorts were used to discover single nucleotide polymorphisms (SNPs) associated with an outcome of interest. These studies give insight into which proteins (depending on the location of the SNP) might be involved in underlying biology. To date in European study populations, 65 variants have been robustly associated with T2D susceptibility, 36 with fasting glucose, and 9 with 2-hr glucose at a genome-wide level of statistical significance ($P < 5\times10^{-8}$, a P-value threshold of 1 million tests, reflecting an estimated number of SNPs above 5% minor allele frequency in the whole genome of 3.3bn pairs)[19-21]. Collectively however, these SNPs only describe 5.7% of the variance in T2D risk[20]. However, common (minor allele frequency >5 %) and rare (minor allele frequency <5 %) SNPs explain about 50% of the heritability estimates by the Falconer's method[17,22].

As to why so much of the heritable fraction of T2D is *not* explained by established gene variants (a phenomenon termed 'missing heritability') is widely speculated. For example, Manolio et al.[23] suggested the following reasons for missing heritability: i) rare SNPs not being included (although later analyses including rare

variants have not had a major impact on missing heritability[17]); ii) multiple causal SNPs being inherited together (i.e. in linkage disequilibrium) which would mask individual effects; and iii) shared family environments, such as lifestyle (in genetics it is common to use the term 'environment' to mean any non-genetic exposure which can affect the phenotype of interest).

Analyses that seek to account for missing heritability by including higher resolution genotyping (where entire regions of the genome are sequenced for each individual, or higher resolution reference panels are used to impute rare variants) have been undertaken. However, a landmark study of this sort found that rare variants added little information[24].

A recent study[17] used structural equation modelling (SEM), which is a multivariate analysis method allowing the inclusion of 'latent' variables, factors that themselves cannot be measured but are 'manifested' as other measurable factors. In this study, the authors used SEM to estimate the proportion of heritability explained by SNPs and shared familial environment, and then compared this to the heritability estimates gleaned from an established approach (Falconer's method). They hypothesised that heritability is often overestimated because the shared familial environment is included in the estimate of heritability, and thus SNPs would explain a smaller proportion of heritable fraction of disease. Using SEM, they estimated h^2 heritability of T2D to be ~50%, which meant that SNPs now explained ~70% of the total heritability. It is noteworthy that studies attempting to explain missing heritability usually do so by using SNPs discovered in conventional T2D GWAS, which focus on variants ranked by *P*-value and not by effect size. Thus, any variants that have large, yet heterogeneous effects, might not be picked up in conventional GWAS and are, therefore, not used in missing heritability analyses.

The evidence clearly suggests that there is a genetic component to the susceptibility of T2D while the magnitude of this compared with the environment is unclear. Still, as our genetic constitution at a population level has not changed substantially in the last 40 000 years[25], it is thought that genetic susceptibility to T2D is not independent of environmental factors, such as lifestyle, which have changed over time[8,26]. Thus, when studying the effects of genomic variation in glycaemic traits, understanding the effects of lifestyle factors on these traits is essential.

Role of lifestyle and adiposity in epidemiology of type 2 diabetes

In a landmark epidemiological study including 85 000 nurses, unhealthy lifestyle (defined as a combination of diet and PA factors) was shown to be associated with 91% of the incidence of T2D that occurred during 16 years of follow-up[27].

In another landmark study on PA and T2D incidence, each additional 500kcal/day expended through PA was estimated to result in a 6% reduction in the relative risk of T2D, up to at least 3500kcal/day of PA energy expenditure[28]. The authors estimated that the protective effect of PA was strongest in those at high risk for T2D (by history of hypertension, high BMI, and family history of T2D).

High levels of PA (compared with low) and obesity (compared with normal weight) have been found to be associated with reduced incidence of T2D[29]. Furthermore, they were found to do so even when adjusted for each other, indicating that part of their effect is likely to be independent, i.e. acting through alternate mechanisms[29]. Notably, this same study found that PA attenuated the risk for T2D in obesity.

The relationships of body fat distribution, as well as type of adipose tissue with insulin resistance and glycaemic control is well established[3,30,31]. A more recent study by Marinou et al. found that deep subcutaneous adipose tissue and visceral adipose tissue were especially detrimental towards whole body- and liver-specific insulin sensitivity[32]. Differentiation between lower and upper body subcutaneous adipose tissue and insulin sensitivity has also been shown; lower body subcutaneous adipose tissue seems protective, whereas upper body subcutaneous adipose tissue seems detrimental to insulin sensitivity[33].

Visceral adipose tissue has also been shown to be inversely associated with PA energy-expenditure, assessed objectively with combined accelerometry and heart rate sensing[34]. Visceral adipose tissue, in turn, is known to be deleterious to our capacity to maintain metabolic homeostasis[35].

Collectively these studies indicate a complex relationship between body fat distribution, glycaemic control and lifestyle factors where some relationships are independent of other ones. For example, the beneficial effects of PA on glycaemic control may be through multiple mechanisms some dependent on adiposity and others not.

Role of lifestyle and adiposity in experimental studies of type 2 diabetes

There is epidemiological evidence linking PA, diet and adiposity factors with T2D. However, this is based on observational data and, as such, is prone to bias and confounding. A natural follow-up step is to test hypotheses generated from

epidemiology in an experimental setting. In this section, I will describe experimental evidence of the role lifestyle has in the aetiology of T2D and how adiposity relates to this. I will focus on *in-vivo* experiments and trials in humans rather than *in-vitro* experiments. In trials, rather than observing associations from naturally occurring exposures, these are experimentally introduced where other exposures are minimised and controlled.

The landmark lifestyle intervention trials

Three seminal lifestyle intervention trials (combining diet and PA components) have shown beneficial effects on glycaemic and adiposity traits: The Malmö feasibility study (MFS)[36], the forerunner to the well-known Finnish Diabetes Prevention study (DPS)[37] and the Diabetes Prevention Program (DPP) based in the USA[38]. The MFS illustrated the feasibility and effectiveness of a large scale (about 7000 participants), long-term (5-year protocol) intervention in reducing T2D incidence and glycaemic control in a T2D, impaired glucose tolerance (IGT) and control arm. In MFS, 90% of participants completed the protocol where estimated maximal oxygen uptake (an estimate of physical fitness) was improved by 10-14% in the intervention arm compared to a reduction of 5-9% in the control arm. The trial showed that more than half of the participants in the T2D and IGT groups were in remission and normalised, respectively after the mean follow-up of six years. The DPS intervened on 522 participants at risk of developing T2D over a mean follow-up time of around three years. Compared to the control arm, the intervention arm showed a 58% reduction in risk for T2D. The DPP trial randomized 3234 participants with impaired fasting glucose (IFG) and/or IGT and randomised participants to either a placebo arm (standard care), metformin arm or lifestyle intervention arm. Over a mean follow up of around three years, they observed a 58% reduction in the incidence of T2D in the lifestyle intervention arm, and a 31% reduction in the metformin arm compared with the placebo.

Physical activity, insulin sensitivity and oxidative capacity

There is evidence suggesting that PA improves glycaemic control by improving peripheral insulin sensitivity and skeletal muscle fat oxidative capacity[39,40]. A six month supervised aerobic exercise intervention was performed in 17 adult pre-menopausal women. Total and regional adipose tissue depots were measured with MRI before and after the intervention. The participants displayed a preferential decrease of visceral adipose tissue, despite no change in BMI and bodyweight following the intervention. Measures of glycaemic control were not included in the study protocol[41]. Despite the participants undergoing an exercise intervention, no difference in caloric intake (assessed by a 7-day diet diary) was observed, suggesting that the participants likely achieved a negative caloric balance. Further, 10 participants with T2D and 10 matched non-diabetics took part in an 8-week PA intervention trial designed to increase walking by 45 minutes per day. An increase

in lipid oxidation and improved fasting and 2-hr glucose was observed after eight weeks of the intervention in T2D participants only[42]. The non-diabetic group, however, showed lower levels of PA from objective PA monitors, which could be the underlying cause for the reported differences.

None of these studies were randomised controlled trials (RCT), and not all accounted for diet. Findings from non-RCTs are prone to bias and confounding, as it may be that there are other exposures acting on the participants than those introduced by the trial. Without randomly assigning a control group, this potential effect is difficult to account for. Similarly, the intervention may unwittingly introduce a change of diet, which could confound the results. However, a 3-month light PA RCT (supervised walking for 60 minutes, 3 times per week), with objectively measured PA in 78 middle-aged Finnish men at risk of developing T2D, showed improvements in insulin sensitivity, 2-hr glucose, low density lipoprotein cholesterol and visceral fat, independent of weight change; this study did account for diet, which showed no significant differences[43]. In the same study, dose-response effects of walking on low-density lipoprotein and visceral adipose tissue were observed and investigators conclude that the beneficial effects of PA are likely to be, in part, independent of adiposity.

PA can take place in various modes, intensities, durations and frequencies, and may have different effects on T2D risk. A study compared the effect of several protocols with varying combinations of intensities and durations in a cross-over efficacy trial. This showed that a high amount of moderate intensity PA was effective at improving glucose tolerance despite a small (2kg) weight change[44]. In recent years, high intensity/low volume exercise has been suggested to be a potentially time-efficient way to exercise for health benefits[45]. In 2013, a cross-over trial in 10 overweight/obese men (aged 26.9±6 years) tested the effects of repeated bouts of high intensity exercise (4×30s sprints, with 4.5 min rest) compared with a single extended exercise bout (extended sprint), matched for energy expenditure and including a control of no exercise over three days in random order. Results showed improved fat oxidation in the day following the tests, by 63% and 38% in the repeated short sprints and extended sprints respectively, vs. the control[46].

Collectively these studies suggest that the beneficial effect of PA on glycaemic regulation is at least in part through improved insulin sensitivity and fat oxidative capacity.

The 'twin-cycle' hypothesis

There is also evidence suggesting that the mechanisms through which lifestyle factors affect insulin sensitivity and other glycaemic parameters are mediated by adiposity. This is primarily through ectopic fat deposition in liver, but also through

fat deposition in the pancreas. The 'twin cycle' hypothesis[30,31] suggests that caloric balance, mediated by liver fat, affects hepatic insulin sensitivity and eventually beta-cell function, through increased levels of pancreatic fat, see Figure 2.

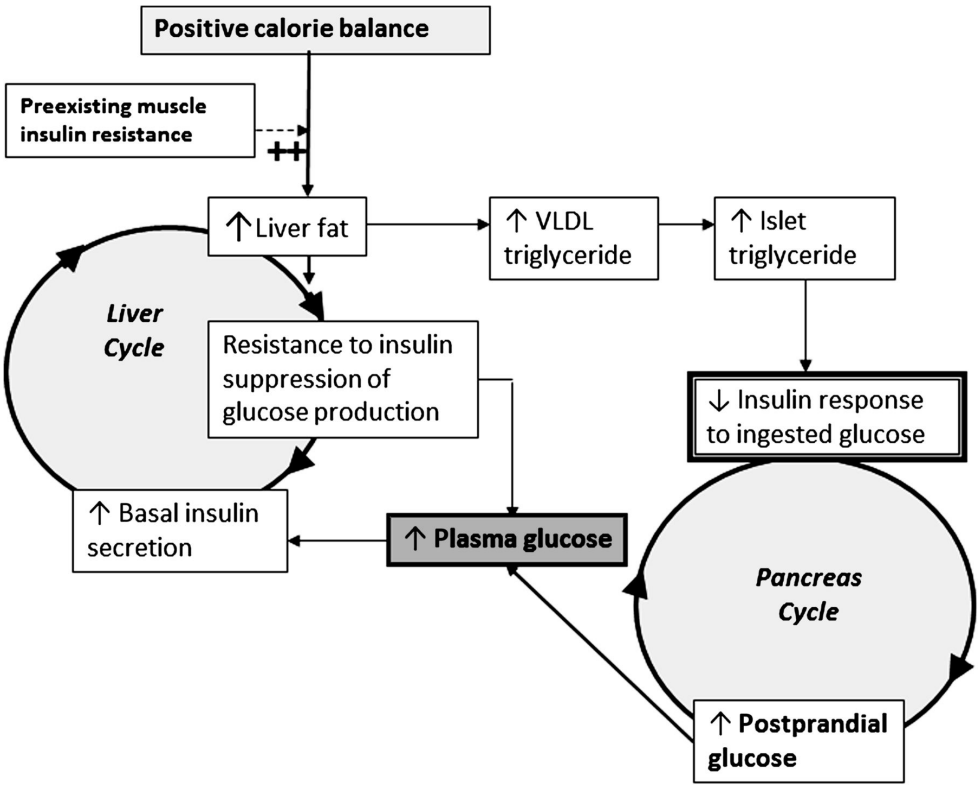


Figure 2. The twin cycle hypothesis of the aetiology of type 2 diabetes.
Figure from Taylor R., 2013, Diabetes[31].

In the twin-cycle hypothesis, a positive caloric balance in combination with peripheral insulin resistance leads to accumulating liver fat. Accumulating liver fat is exacerbated by increased *de-novo* lipogenesis, stimulated by the high levels of insulin (due to higher insulin resistance). Liver fat accumulation leads to hepatic insulin resistance, to inhibition of hepatic gluconeogenesis, which raises glucose- and (subsequently) insulin-levels further. The up-regulated *de-novo* lipogenesis leads to higher circulating levels of very low-density lipoprotein, which raises triglyceride levels in surrounding tissue such as the pancreas. This, in combination with the high levels of glucose, leads to glucotoxicity and lipotoxicity, which

reduces beta-cell function. Reduced beta-cell function leads to an inability to compensate for increased insulin resistance, which leads to impaired glycaemic regulation.

The 'twin cycle' hypothesis is based on various bodies of evidence, some of which are discussed above. Taylor's hypothesis does not consider PA independently of diet but rather as a component of caloric balance. The hypothesis has been tested in trials conducted by investigators from the same research unit[47,48]. However, a recent review, specifically focusing on the effects of exercise on liver fat, concluded that exercise can indeed reduce liver fat through a reduction in circulating free fatty acids (FFA), caused by increased skeletal muscle uptake of FFA[49]. This hypothesis is consistent with the '*athlete paradox*' where endurance-trained athletes are insulin sensitive despite increased intra myocellular lipids, i.e. the relationship between intra myocellular lipids and insulin sensitivity is in direct contrast to that seen in non-athletes.

Taken together, the existing evidence suggests a number of potential mechanisms by which lifestyle factors can affect parameters of glycaemic control and the pathogenesis of T2D, in particular those in the twin-cycle hypothesis. These mechanisms have not, however, been shown in population settings under a single unified analysis framework.

Aims

The existing evidence strongly supports the beneficial effects of PA in T2D pathophysiology. This includes beneficial effects on adiposity-related traits (e.g. obesity, visceral fat and liver fat) and glycaemic control (e.g. HbA1c, fasting glucose, 2-hr glucose and insulin sensitivity). The evidence also suggests that these relationships may interrelate in cyclic pathways.

Improving our understanding of these relationships to find underlying causal mechanisms (such as those hypothesised in the twin cycle model) is a major research focus worldwide, including the GLACIER and IMI DIRECT studies, within which the projects covered in this thesis are nested.

The overall objective of this thesis is to investigate the role of lifestyle with an emphasis on PA in glycaemic control, before and after the onset of T2D.

The specific aims of the four papers included in this thesis are:

- In Paper 1, we compare the predictive ability of established genetic susceptibility loci and established lifestyle factors in the incidences of obesity, impaired glycaemic control and T2D.
- In Paper 2, we overview the design and rationale of the two IMI DIRECT glycaemic deterioration prospective cohorts.
- In Paper 3, we describe the baseline characteristics of the two IMI DIRECT glycaemic deterioration prospective cohorts, in order to provide context for those subsequently analysing and reading articles based on data from these cohorts.
- In Paper 4, we test the hypothesised relationships of the key factors involved in the regulation of glucose homeostasis, proposed by Taylor et al.[31] in the mechanistic 'twin cycle model'. The aim is also to determine if the association of PA with glycaemic control is mediated by parameters in the model, and whether such relationships differ in a prediabetic cohort compared with a diabetic cohort.

Materials and Methods

Two population cohorts are used in this thesis, the Gene-Lifestyle interactions And Complex traits Involved in elevated disease Risk (GLACIER) study and two new cohorts within the Diabetes Research on Patient Stratification Consortium (DIRECT). Paper 1 is a GLACIER paper, whilst Papers 2-4 are DIRECT papers. In this section, I give an overview of each cohort and describe the materials and methods pertaining to the results presented in this thesis. More detailed information about the materials and methods used in these papers are available in a study by Kurbasic et al.[50] for GLACIER, and Papers 2-3 for the DIRECT cohorts.

GLACIER

The GLACIER study is nested within the Västerbotten Health Survey (VHS), which is sometimes called the Västerbotten Intervention Program[51] (VIP) coordinated in Umeå, Sweden (see Figure 3).

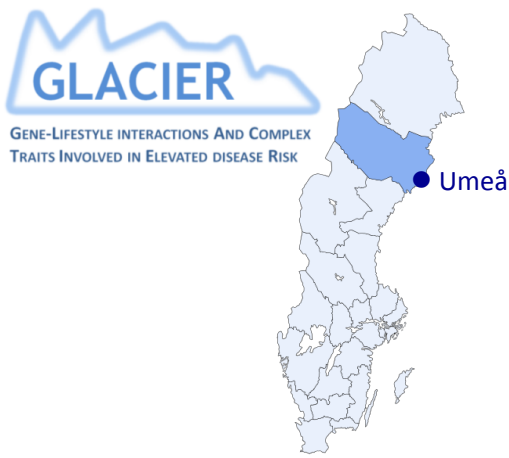


Figure 3
GLACIER cohort study site.

The VHS/VIP began in the 1985 in response to high cardiovascular disease (CVD) mortality in the region at the time. The study is a community-based intervention where participants from Västerbotten county were invited for detailed screening and lifestyle counselling. The visit took place at their primary healthcare facility in the year of their 30th, 40th, 50th and 60th birthdays. The detailed health screening consisted of standard clinical parameters, blood pressure, anthropometrics, fasting blood samples, 75g oral glucose tolerance test, and validated lifestyle assessment questionnaires. VHS/VIP consists of around 90,000 participants, of these around 19,000 were enrolled into GLACIER. Within GLACIER, around 6000 participants have been genotyped using the Illumina Cardio-MetaboChip array, and 5000 participants have a follow-up visit (at around 10 years). Around 3500 of the genotyped participants have a follow-up visit.

IMI DIRECT

The DIRECT study was launched in 2012 under the banner of the Innovative Medicines Initiative (IMI), a joint undertaking of the European Union (EU), European academic institutions and pharmaceutical companies that form part of the Seventh Framework Programme (FP7). There are 21 academic partners and 5 pharmaceutical industry partners in DIRECT. The overall aim of the DIRECT consortium is to discover and validate biomarkers that: predict glycaemic deterioration before and after T2D onset; predict therapeutic response; and help stratify T2D into subclasses for more efficient prevention and treatment. Within the DIRECT consortium, we established two new prospective cohort studies aimed at identifying biomarkers for glycaemic deterioration, one before the onset of T2D (Cohort 1, 'Prediabetes'), and one after the onset of T2D (Cohort 2, 'T2D'). The study centres are located at 7 academic sites across Europe, see Figure 4.

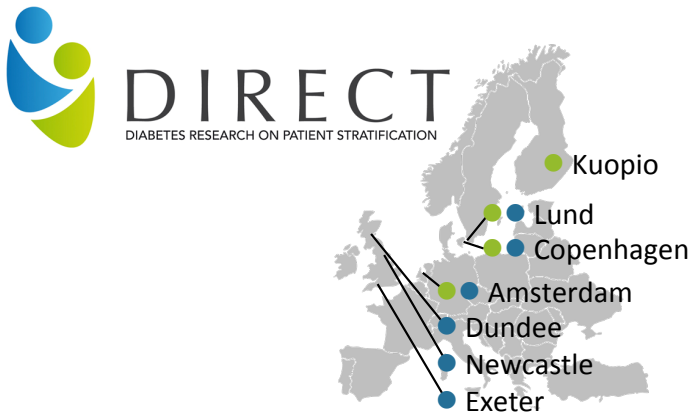


Figure 4.

IMI DIRECT glycaemic deterioration cohort study sites. Green dot: Cohort 1 ('Prediabetes'), Blue dot: Cohort 2 ('T2D')

Three of the seven study centres (Lund/Malmö, Copenhagen, and Amsterdam) recruit for both Cohort 1 and Cohort 2. The Kuopio study centre, only recruits for Cohort 1 and the three centres in the UK (Dundee, Newcastle, and Exeter) recruit only for Cohort 2. The overall design of the cohorts is illustrated in Figure 5.

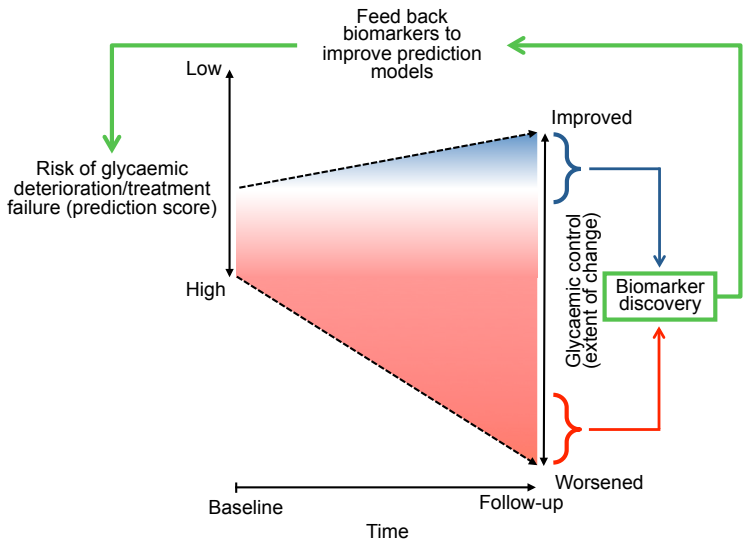


Figure 5.

Overview of the biomarker discovery strategy in the two glycaemic deterioration prospective cohorts of the DIRECT consortium. Persons at high risk of glycaemic deterioration before (Cohort 1) or soon after (Cohort 2) the onset of T2D are enrolled and followed for between 36 and 48 months. Biomarkers are discovered for the rate of glycaemic deterioration. Discovered biomarkers are subsequently fed back to improve risk-prediction models, which will be validated in other epidemiological studies and clinical trials organised by the DIRECT Consortium and its Partners. Adapted from Paper 1.

The two cohorts in DIRECT for glycaemic deterioration are prospective cohorts designed to capture the change in glycaemic control over time, before and after the onset of T2D. At recruitment, participants are expected to represent a range of risk for glycaemic deterioration and, over time, this deterioration is expected to be heterogeneous between the participants. Discovery of biomarkers that predict the extent of decline in glycaemic regulation are fed back to improve prediction models and shall be validated in subsequent DIRECT clinical trials.

Recruitment

Participants in Cohort 1 were recruited using a risk prediction algorithm, DIRECT-DETECT based on the DETECT-2DM algorithm[52]. The DIRECT-DETECT algorithm was developed by DIRECT scientists for this purpose. Two models were developed, one for use where HbA1c was not available (Model 1) and another with HbA1c (Model 2). The models are outlined below. Note that for 'if' arguments, true = 1, false = 0.

Model 1 for men:

HbA1c at follow-up = (5502 + 53 × (if age between 45.0 and 54.99) + 91 × (if age between 55.0 and 64.99) + 188 × (if age ≥ 65.0) – 28 × (if BMI between 25.0 and 29.99) + 32 × (if BMI ≥30.0) + 65 × (if waist between 94.0 and 101.99) + 115 × (if waist ≥ 102.0) + 50 × (if person uses antihypertensive medication) + 133 × (if current smoker) + 69 × (if parent(s) have/had T2D)) / 1000

Model 1 for women:

HbA1c at follow-up = (5398 + 173 × (if age between 45.0 and 54.99) + 213 × (if age between 55.0 and 64.99) + 307 × (if age ≥ 65.0) + 33 × (if BMI between 25.0 and 29.99) + 108 × (if BMI ≥30.0) + 2 × (if waist between 94.0 and 101.99) + 106 × (if waist ≥ 102.0) + 50 × (if person uses antihypertensive medication) + 93 × (if current smoker) + 71 × (if parent(s) have/had T2D)) / 1000

Model 2 for men:

HbA1c at follow-up = (2211 + 4 (if age between 45.0 and 54.99) + 19 (if age between 55.0 and 64.99) + 81 (if age ≥ 65.0) + 70 (if waist between 94.0 and 101.99) + 128 (if waist ≥ 102.0) - 109 (if former smoker) - 9 (if current smoker) + 90 (if parent(s) have/had T2D) + 604 × HbA1c) / 1000

Model 2 for women:

HbA1c at follow-up = (2300 - 24 (if BMI between 25.0 and 29.99) + 39 (if BMI ≥30.0) + 31 (if waist between 80.0 and 87.99) + 81 (if waist ≥ 88.0) + 600 × HbA1c) / 1000

These models were applied to datasets from parent cohorts at 'prediabetes' study centres (Cohort 1). Participants were then ranked based on the scores from these models; recruitment began from those with the highest scores. A timeline overview of the visit protocols for both cohorts is illustrated in Figure 6.

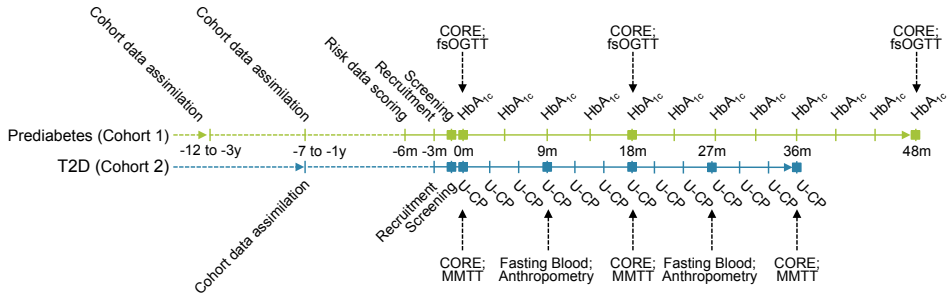


Figure 6.
Overview of the timeline of the DIRECT glycaemic deterioration cohorts 1 and 2 protocols. Core assessments (CORE) are: anthropometry; fasting blood; MRI*; faecal microbiome; urine; physical activity; diet; quality of life; diabetes family history; medication history. Dashed lines indicate data assimilated from existing cohorts and registers. m, months; U-CP, urinary C-peptide; y, years. For Cohort 2, the 9m and 27m* visits are minor study visits. *carried out in a subset of the sample population. Squares, deep-phenotype study visit. Adapted from Paper 1.

Protocols for each study are prospective with main visits at baseline, 18 months and 36 months (Cohort 2), and 48 months (Cohort 1). Baseline visit recruitment for both cohorts began in December 2012 and the final 48m visits are expected to take place towards the end of the second quarter in 2018. Both cohorts are comprehensively phenotyped with assessments for: Standard clinical characteristics, glycaemic control, and beta-cell function modelling from frequently samples oral glucose tolerance test / mixed meal tolerance test (fsOGTT/MMTT), abdominal adiposity by MRI, self-reported diet, objective PA by tri-axial accelerometry, and multiple omics including genomic, transcriptomic, metabolomic, proteomic and faecal microbiome.

Cohort specific methods

Glycaemia

GLACIER

Capillary blood was drawn after an overnight fast of at least 8 hours. 12% at baseline, and 2% at follow-up reported having fasted for less than 8 hours. Fasting status was therefore included as a covariate in glycaemic models. A second blood

sample was drawn 2 hours after a 75g oral glucose load. Plasma glucose concentrations were measured using a Reflotron bench-top analyser (Roche Diagnostics Scandinavia, Umeå, Sweden). T2D was determined either by self-report or from fasting and 2-hr glucose values determined during the visit.

DIRECT

Fasting venous blood samples were drawn after a 10-hour overnight fast from a cannula inserted into a forearm vein. Non-fasted participants were rescheduled or excluded from the study, so all participants were fasted for a minimum of 10 hours.

In Cohort 1, a finger-prick fasting capillary glucose sample with a HemoCue Glucose 201 monitor or similar was taken before the glucose load. Participants with a fasting capillary glucose above >11mmol/l were excluded from the study. After this, an fsOGTT was conducted where a baseline blood sample was drawn and a standard 75g oral glucose load was consumed within 2-5 minutes. Blood samples were subsequently drawn at 15, 30, 45, 60, 90, and 120 minutes.

In Cohort 2, an MMTT was carried out instead of an fsOGTT. A baseline sample was taken and following this, participants consumed a 250 ml Fortisip liquid meal (18.4 g carbohydrate per 100 ml) over a period of 2–5 min. Blood samples were drawn at 30, 60, 90, and 120 minutes post load.

Based on glucose, C-peptide and insulin concentrations from the blood samples at the above time points, β-cell function was parameterised using a mathematical model described in detail elsewhere[53]. Briefly, the model describes the relationship between insulin secretion and glucose concentration as the sum of two components. The first component is the dose-response relationship between glucose and insulin secretion (this parameter is termed *glucose sensitivity*). This is modulated by a potentiation factor, which accounts for the enhancement of the insulin secretion during the descending late phase of the glucose concentration following the load (this parameter is termed *potentiation fraction ratio* or PFR). The second component of the model reflects the dependency of insulin secretion on the rate at which the glucose concentration increases during the ascending early phase of the glucose curve (this parameter is termed *rate sensitivity*), and is indicative of early insulin release[54].

Physical Activity

GLACIER

A modified version of the validated International Physical Activity Questionnaire (IPAQ) was used[55,56]. For the analysis in Paper 1, only leisure time PA was considered. Participants were asked "In the past 3 months, how often have you exercised in training gear?" with the possibility to answer: 'never, occasionally, 1–2 times/week, 2–3 times/week or >3 times/week'. 'Never' and 'occasionally' were combined into a physically 'inactive' category and '1–2 times/week', '2–3 times/week' and '>3 times/week' to 'physically active'.

DIRECT

We objectively assessed PA using a tri-axial accelerometer (ActiGraph GT3X+/GT3X+w/GT3X+bt; Actigraph, LLC, Pensacola, FL, USA). The monitor was worn on the non-dominant wrist continuously for 10 days. It was fastened with the manufacturer's disposable (single use, non-removable) hospital band-type strap to allow comfortable, uninterrupted measures of both sleep and PA. The monitor was set to record at 30 Hz, with the LED-light off and the manufacturer's sleep-mode disabled. Participants were instructed only to remove the device if they were going to undertake water-based activities deeper than 1 meter and lasting for more than 30 minutes. To analyse the rawest possible measures, manufacturer raw data files (.gt3x) were converted to comma-separated value (.csv) using ActiLife 6 (version 6.11.5, ActiGraph Co, Pensacola, USA). The raw (.csv) data was then used to perform quality control steps and calculate the PA parameters using PAMPRO (version uploaded 2015-10-21, MRC Epidemiology unit, Cambridge, UK), a custom open source software available under public license (https://github.com/Thomite/pampro). Each axis of acceleration was auto-calibrated to local gravity[57]. Non-wear was rare due to the method of fastening; it was not possible to refasten the accelerometer if removed. Nonetheless, non-wear was inferred as a 60-min consecutive period of vector magnitude (direction agnostic acceleration intensity) standard deviation of less than 4mgs. Measures were adjusted intra-individual differences due to non-wear removal by diurnal rhythm[58]. The main PA estimates used in the analyses of both Paper 3 and 4 are Euclidean norm minus one (ENMO) and high-pass-filtered vector magnitude (hpfVM), both of which infer intensity of participants' movement in any direction ,at any given time (or average values during a defined period). Custom analyses of the raw accelerometry data are an important feature of these PA measures, as the methods are entirely transparent and can be comparable when applied to data from other monitors and other studies[59].

Diet

GLACIER

Diet was assessed using a validated Food Frequency Questionnaire (FFQ), designed to capture habitual diet over the previous year[60-62]. A nine-point frequency scale was used for participants to report how often they consumed food or beverages. Answers ranged from 'never' to 'four or more per day'. Average portion size for meat and fish, vegetables, potatoes, rice and pasta were also assessed. Total energy intake was calculated based on the National Food Administration database (www.slv.se). The FFQ version used was adjusted for, because the 84-food item FFQ (used between 1985-1996) was reduced to a 66-food item FFQ (used from 1996 onwards) after combining several questions related to similar foods items. In addition to this, if ≥10% of the FFQ was missing, or total energy intake was estimated to be <500 or >4500kcal/day, the data were excluded. Three diet scores were calculated from the parameters: the Healthy Diet (HD) score, the Nordic Nutrition Recommendation (NNR)[63] score, and a principal components analysis (PCA) score were all constructed from the available diet data. These are described in more detail in Paper 1. Briefly, the HD score was validated elsewhere[64] and is based on intakes of eight food groups. Whole grains, fish, fruits and vegetables were designated as 'favourable' foods, whereas red and processed meats, desserts and sweets, sugar-sweetened beverages and fried potatoes were designated as 'unfavourable'. A higher HD score indicates a healthier diet. The validated NNR[63] score assigns one point for every Nordic nutrition criteria which is met, thus a higher NNR score indicates a healthier diet. A PCA including all the macronutrients (i.e. carbohydrate, protein, total fat, saturated fat, monounsaturated fatty acids, polyunsaturated fatty acids, essential fatty acids and fibre) and adjusted for total energy intake was conducted to create the PCA score. One factor was retained which contrasted carbohydrate and fibre intake against fat intake; this accounted for 54% of the variance of all macronutrients.

DIRECT

A 24-hr multi-pass dietary record was used to assess diet. The assessment was made the day before the baseline visit. The method has been validated for total energy intake using double labelled water with good reproducibility[65-68].

The method is composed of three question 'passes': the first document is a usual day's meal; the second pass gives the participant time to reflect and add to the first pass; the final third pass aims to record portion size. In conjunction with the 24-hr diet record, participants also answered a food habit questionnaire covering the overall quality of their diet against guidelines. Questionnaires were translated into the local language of each study centre and back translated by native speakers at

the analysing academic Partner site. The data was then manually entered and computationally analysed using Dietplan-6, a comprehensive food analysis programme (version 6.70.43, 2013; Forestfield Software, Horsham, UK). Under/over reporting of energy intake is assessed using Goldberg's equation[69,70].

Magnetic Resonance Imaging (MRI)

DIRECT

We assessed abdominal adiposity factors using Magnetic Resonance Imaging (MRI). MRIs were undertaken on about half of Cohort 1 and all of Cohort 2 for the baseline visit. Protocols across the study centres were standardised as far as possible given the equipment available. Scans were made at 1.5 and 3.0T field strengths, depending on equipment. The scanners used at each centre were: Siemens Trio 3T at Dundee, Philips Intera 1.5T at Exeter, Siemens Espree 1.5T at Newcastle, Philips Achieva 3T at Copenhagen and Siemens Avanto 1.5T at Kuopio and Amsterdam.

Participants are scanned in the prone position with arms extended above their head. T1-weighted images are taken from diaphragm to the acetabulum using the maximum field of view (during free breathing) with a slice thickness and gap of 10 mm. Pancreatic volume was assessed from scans in suspended breathing. Three-dimensional T1-weighted scans with fat suppression is placed over the pancreas to cover the whole organ. 50 to 90 slices with thickness ranging from 1.2 to 2mm is used. Once the pancreas is identified, axial images are performed during suspended breathing. These are used to position a single-slice multiecho sequence through the pancreas using a surface coil. An identical axial slice is acquired through the liver. These methods have been described and validated elsewhere [71]. Images are converted to an analysable format using Image J (Image; National Institutes of Health, Bethesda,MD). An automated pixel-by-pixel analysis is then carried out to for colour-coded maps of the entire pancreas and liver using Matlab version 7.7 (Mathworks, Natick, MA, USA). The proportion of fat and water within the liver and pancreas are then estimated from these.

Genotyping

GLACIER

Extraction of DNA was carried out using methods described elsewhere [72,73]. Genotyping was performed using the Illumina Metabochip array (Illumina, San Diego, CA, USA)[74] at the Wellcome Trust Sanger Institute, UK.

DIRECT

Extraction of DNA was carried out using Maxwell 16 Blood DNA purification kits and Maxwell 16 semi-automated nucleic acid purification system (Promega). Genotyping was conducted using the Illumina HumaCore array (HCE24 v1.0) and genotypes were called using Illumina's GenCall algorithm. Additional details on genotyping in the DIRECT cohorts can be found in Paper 3.

Paper Specific Methods

Paper 1

Glycaemic strata

Glycaemic strata were defined according to the American Diabetes Association criteria from 2003[75]. For T2D, these were: fasting glucose ≥7.0 mmol/l or a 2-hr glucose concentration ≥11.1 mmol/l. IFG was defined as: fasting glucose ≥6.1 and <7.0 mmol/l. IGT was defined as 2-hr glucose ≥7.8 and <11.1 mmol/l. Normal glucose regulation (NGR) was defined as having fasting glucose <6.1 mmol/l and 2hr-glucose <7.8 mmol/l. Incident cases of IFG and IGT were defined as participants changed from NGR to IFG or IGT during follow-up.

Genotypic variables

We extracted: 65 T2D associated SNPs[20], 36 fasting glucose associated SNPs[21], 9 2-hr associated SNPs[21], and 97 BMI associated SNPs[76]. We coded genotypes according to number of effect alleles (alleles associated with higher trait values), as reported in the source studies[20,21,76]. We used proxy loci for 26 SNPs which were not available on MetaboChip. Missing genotypes were imputed using the mean imputation method[77] by replacing each missing genotype with its mean value obtained from the rest of the cohort having genotypic information. Missing rate was ≤0.07 per participant and ≤0.007 per SNP.

We observed no significant deviations from Hardy–Weinberg equilibrium (No SNPs with P<0.0001). We generated unweighted Genetic Risk Scores (GRS) to examine the cumulative effects of the SNPs. This was done by summing up the number of effect alleles from each trait associated SNP, thus, the minimum value for each score is 0 and the maximum twice the number of SNPs in the score. These are t2d-GRS for T2D, fg-GRS for fasting glucose 2hg-GRS for 2-hr glucose and ob-GRS for obesity.

Statistical analyses

We carried out logistic regressions to assess the predictive ability of genetic and lifestyle factors on the incidence of T2D, IFG, IGT, Obesity and weight gain ≥10%. Participants who were classified as T2D, IFG, IGT, or Obese at baseline were excluded from analyses. We used three models:

- Model 1 (genetic) included age, age^2, follow-up duration, fasting status (for glycaemic traits), sex and trait-specific SNPs as independent variables
- Model 2 (lifestyle) included age, age^2, follow-up duration, fasting status (for glycaemic traits), sex, FFQ version, education, smoking status, PA and intakes of total energy, alcohol, salt, sucrose, macronutrients, vitamins and minerals
- Model 3 (combined) included all variables in Models 1 and 2.

The predictive ability of the models was assessed by calculating the area under the receiver operator curve (ROC AUC) and the predictive ability of the different models were compared using a method described in detail elsewhere[78]. To assess the gain in predictive accuracy of adding the genetic factors to the lifestyle factors in the combined model, we estimated the continuous net reclassification improvement (cNRI)[79]. Model calibration was assessed by Akaike's information criterion (AIC) and the Hosmer–Lemeshow test[80].

To assess the association of separate lifestyle factors and genetic factors, we combined the genotypes into GRSs and diet variables into diet scores as described above. This was important to reduce multicollinearity, which effects individual factor estimates[81].

We investigated differences in genetic and lifestyle factors related to the traits by calculating quartiles of each lifestyle factor/score (except alcohol) and comparing the top and bottom quartiles.

Analyses were carried out in PLINK (version 1.07)[82], R (version 3.1.1)[83] and SAS (version 9.4)[84].

Paper 3

Glycaemic impairment was defined according to the American Diabetes Association criteria from 2011[85]. NGR was defined as having fasting glucose <5.6 mmol/l and 2hr-glucose <7.8 mmol/l. Impaired A1c (IA1c) was defined ≥5.6 (37mmol/mol) and <6.5% (47.5mmol/mol). IFG was defined as: fasting glucose ≥5.6 and <7.0 mmol/l. IGT was defined as 2-hr glucose ≥7.8 and <11.1 mmol/l. T2D was defined as fasting glucose ≥7.0 mmol/l or a 2-hr glucose concentration

≥11.1 mmol/l. Based on these criteria, participants in Cohort 1 were stratified into: NGR, isolated impaired HbA1c (iIA1c), isolated impaired fasting glucose (iIFG), isolated impaired glucose tolerance (iIGT) and combined impaired glucose regulation (cIGR). NGR was defined as having HbA1c, fasting glucose and 2-hr glucose values within normal ranges. iIA1c, iIFG, iIGT were defined as having only IA1c, IFG or IGT, respectively. cIGR was defined as having a combination of IA1c, IFG and/or IGT. Cohort 2 was stratified into treatment categories: lifestyle advice only (LS) or metformin and receiving lifestyle advice (Met+LS). These strata were chosen to reflect established strata according to current guidelines.

We calculated means adjusted for standard putative confounders (age, sex and centre) using least square means. We used general linear models to compare differences between strata in continuous variables. Statistical significance is defined as $P < 0.05$. All analyses were carried out in R (version 3.2.3)[83]. Version 'preliminary release 1 (direct_11-02-2016)' of the DIRECT data release was used for this analysis.

Paper 4

Continuous variables were rank normally transformed where the mean = 0 and standard deviation = 1. Variables were regressed on age, sex, study centre, energy intake, carbohydrate-, fat-, and protein-intake (and metformin treatment in Cohort 2), and the residuals from these regressions extracted. These residuals, which reflect the deviation from the mean after adjustment for the confounders, were then used in further analyses.

To assess the relationship between parameters in the hypothesised twin-cycle hypothesis, we used variables in the dataset to define a model as close as possible to the original hypothesised model[31] (see Figure 2 in introduction). Based on this, we defined a SEM using only manifest (measured) variables to fit a twin cycle hypothesis model on the DIRECT baseline data; we included edges from PA on all parameters in the model. This SEM definition is illustrated in Figure 7.

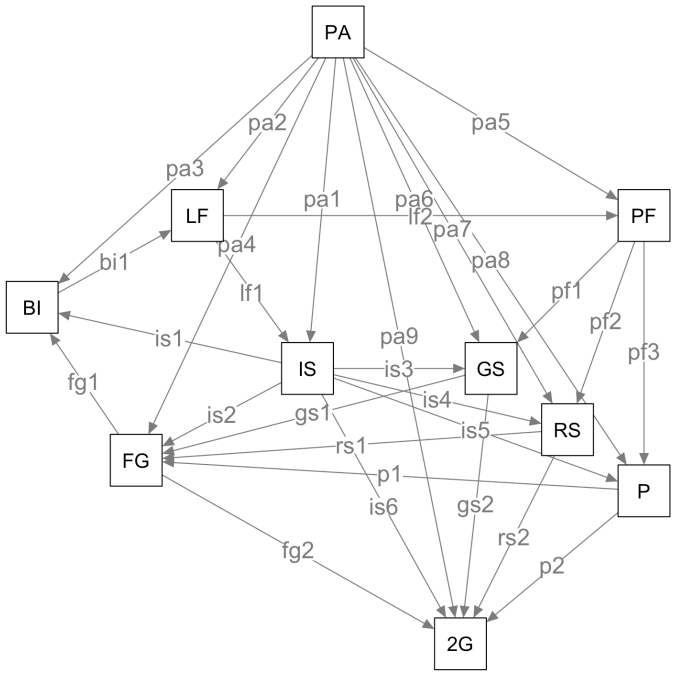


Figure 7
Structural equation model definition diagram for a hypothesised model for the role of physical activity and liver fat in glycaemic control. The diagram illustrates the model definitions, where manifest nodes are represented as squares and arrows indicate regression coefficients pointing towards an outcome of a respective regression (note several arrows pointing to the same node indicates a multiple regression). All continuous variables normally transformed and adjusted for age, sex, metformin treatment (Cohort 2), study centre, total energy-, carbohydrate-, fat-, and protein-intake. PA: MeanVMhpf (physical activity), FG: Fasting Glucose, 2G: 2-hr Glucose, IS: Oral Glucose Insulin Sensitivity, LF: Liver Fat, PF: Pancreatic Fat, BI: Fasting insulin secretion rate, GS: Glucose Sensitivity (insulin secretion per glucose), RS: Rate Sensitivity (early insulin secretion enhancement), P: Potentiation Fraction Ratio (late insulin secretion enhancement).

The model differs from the original twin-cycle model primarily in the use of insulin sensitivity. As a measure of resistance to insulin suppression of hepatic glucose production is not available in our dataset, we used 2-hr oral glucose insulin sensitivity (OGIS) instead. Our model also differs in the direction of the relationship modelled between 2-hr/postprandial glucose and fasting glucose.

We assessed pairwise relationships using Pearson correlations, which we report in a correlation matrix. To assess relationships within the twin-cycle model, we estimated whole model fit, direct effects, and pathway effects (from PA to various outcomes). Pathway/mediation effects were estimated using the coefficient product method[86] where mediation is defined using the approach described by Baron and Kenny (Baron & Kenny, 1986). Relative model fit was assessed using the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI). Absolute fit

was assessed using Root Mean Square Error of Approximation (RMSEA), and Standardised Root Mean Square Residuals (SRMR). Use of these to categorically determine adequacy of fit is controversial; however, it is common practice to report them[87-89].

Multiple testing adjustments were not required, as the pairwise associations were separate tests done largely in replication of hypothesised associations. Estimates within the SEM are nested within a single model, thus do not require multiple testing adjustment. Statistical significance was defined as $P < 0.05$. All statistics were computed using R (version 3.2.3)[83]. SEMs were fitted using R-package lavaan (version 0.5-20)[90]. Models were plotted using semPlot (version 1.0.1)[91]. The DIRECT data release version used for the analyses in this article is 'preliminary release 1 (direct_11-02-2016)'.

Results and Discussion

Innate biology vs. lifestyle factors in predicting glycaemic deterioration and diabetes (Paper 1)

The purpose of the analysis was to conduct a head-to-head comparison of the full array of established genetic markers (from GWAS on trait susceptibility) and established lifestyle risk factors for IFG, IGT, T2D and obesity incidence.

Main findings

Of 5,726 participants from the GLACIER cohort with the necessary phenotypic and genotypic data for the current analyses, 3,444 participants had follow-up data available for a median of 9.9±0.4 years. In this period, the incidence of T2D was 192, IFG 563, IGT 613, and obesity 264. Mean age of the cohort at baseline was 45.2±6.7 years (presented as mean±SD), BMI was 25.1±3.7 kg/m^2, fasting glucose was 5.3±0.7 mmol/l, and 2-hr glucose was 6.5±1.4.

T2D

The predictive ability of the *genetic* and *lifestyle* models was similar for T2D incidence (AUC 74% v. 75%; $P_{\text{difference}} = 0.47$). The *combined* model (lifestyle model + genetic factors) had a significantly higher predictive ability (AUC 80% $P_{\text{difference}} = 0.0003$) than the lifestyle model alone. The net reclassification improvement of the *combined* model compared to the lifestyle model was 58% ($P < 0.0001$). We also investigated the association of genetic and lifestyle factors with incidence of T2D (with separate models for each diet score). All lifestyle factors, except alcohol intake, demonstrated statistically significant associations with incidence of T2D (see Table 1).

Table 1.
ORs for prediction of incident type 2 diabetes according to lifestyle risk factors and GRSs.

	Regression model (OR, 95% CI)								
	NNR Score			HD Score			PCA Score		
Variables	**Genetic** (n= 2,017)	**Lifestyle** (n= 2,017)	**Combined** (n= 2,017)	**Genetic** (n= 2,087)	**Lifestyle** (n= 2,087)	**Combined** (n= 2,087)	**Genetic** (n= 2,017)	**Lifestyle** (n= 2,017)	**Combined** (n= 2,017)
Smoking status	-	**1.59** (1.06-2.37)	**1.55** (1.04-2.32)	-	**1.55** (1.05-2.30)	**1.53** (1.03-2.27)	-	**1.66** (1.11-2.48)	**1.62** (1.08-2.43)
Education	-	**0.57** (0.34-0.96)	**0.58** (0.34-0.98)	-	**0.56** (0.34-0.93)	**0.57** (0.34-0.95)	-	**0.55** (0.33-0.92)	**0.55** (0.33-0.94)
Alcohol intake	-	0.87 (0.52-1.45)	0.86 (0.51-1.44)	-	0.93 (0.56-1.53)	0.91 (0.55-1.51)	-	0.86 (0.52-1.44)	0.85 (0.51-1.43)
Physical activity	-	**0.51** (0.33-0.80)	**0.51** (0.33-0.79)	-	**0.54** (0.35-0.83)	**0.53** (0.35-0.82)	-	**0.49** (0.32-0.76)	**0.48** (0.31-0.75)
T2d-GRS	**1.84** (1.16-2.92)	-	**1.84** (1.15-2.94)	**1.9** (1.20-3.00)	-	**1.9** (1.20-3.02)	**1.84** (1.16-2.92)	-	**1.82** (1.14-2.91)
NNR Score	-	1 (0.63-1.59)	0.99 (0.62-1.57)	-	-		-	-	-
HD Score	-	-	-	-	1.26 (0.77-2.06)	1.22 (0.74-2.01)	-	-	-
PCA Score	-	-	-	-	-	-	-	**0.5** (0.31-0.83)	**0.52** (0.31-0.85)

Genetic model: age, age^2, sex, follow up years and ob-GRS. Lifestyle model: age, age^2, sex, FFQ type, follow up years, education, smoking status, alcohol intake, physical activity and diet scores. Combined model: Lifestyle model + T2d-GRS. Statistically significant values (P<0.05) are marked in bold. Smoking status = non-smokers vs current smokers; education = school vs university education; physical activity= inactive vs active; alcohol intake= 1st vs 4th quartiles; ob-GRS=1st vs 4th quartiles; NNR Score= 1st vs 4th quartiles; HD Score= 1st vs 4th quartiles; PCA Score= 1st vs 4th quartiles.

Notably, active participants had a lower incidence of T2D, independent of diet scores, alcohol intake, smoking status and education. The PCA score was the only diet score to be associated with incidence of T2D.

IFG

The predictive ability of the *genetic* model tended to be higher than that of the *lifestyle* model for IFG incidence, but this difference was not statistically significant (AUC 66% v. 63%; $P_{\text{difference}} = 0.05$). The predictive ability of the *combined* model (lifestyle model + genetic factors) was significantly higher (AUC 69% $P_{\text{difference}} < 0.0001$) than that of the *lifestyle* model alone. The net reclassification improvement of the *combined* model compared to the *lifestyle* model was 36% ($P < 0.0001$). None of the lifestyle factors were significantly associated with incidence of IFG (see Table 2).

Table 2.
ORs for prediction of incident IFG according to lifestyle risk factors and GRSs.

	Regression model (OR, 95% CI)								
	NNR Score			HD Score			PCA Score		
Variables	**Genetic** (n= 2,778)	**Lifestyle** (n= 2,778)	**Combined** (n= 2,778)	**Genetic** (n= 2,882)	**Lifestyle** (n= 2,882)	**Combined** (n= 2,882)	**Genetic** (n= 2,778)	**Lifestyle** (n= 2,778)	**Combined** (n= 2,778)
Smoking status	-	1.17 (0.91-1.50)	1.18 (0.91-1.51)	-	1.16 (0.91-1.48)	1.16 (0.91-1.49)	-	1.16 (0.91-1.49)	1.17 (0.91-1.50)
Education	-	0.84 (0.63-1.12)	0.85 (0.63-1.14)	-	0.89 (0.66-1.18)	0.9 (0.68-1.21)	-	0.83 (0.62-1.12)	0.85 (0.63-1.14)
Alcohol intake	-	1.16 (0.85-1.59)	1.15 (0.83-1.57)	-	1.18 (0.87-1.61)	1.16 (0.85-1.58)	-	1.16 (0.84-1.59)	1.14 (0.83-1.57)
Physical activity	-	0.92 (0.73-1.15)	0.93 (0.74-1.17)	-	0.93 (0.74-1.16)	0.94 (0.75-1.18)	-	0.91 (0.73-1.15)	0.93 (0.74-1.17)
fg-GRS	**1.67** (1.25-2.24)	-	**1.66** (1.24-2.23)	**1.73** (1.30-2.31)	-	**1.72** (1.29-2.30)	**1.67** (1.25-2.24)	-	**1.66** (1.24-2.23)
NNR Score	-	1.01 (0.77-1.34)	1.01 (0.76-1.33)	-	-	-	-	-	-
HD Score	-	-	-	-	0.87 (0.64-1.17)	0.87 (0.64-1.17)	-	-	-
PCA Score	-	-	-	-	-	-	-	1.09 (0.81-1.47)	1.1 (0.82-1.48)

Genetic model: age, age^2, sex, follow up years and fg-GRS. Lifestyle model: age, age^2, sex, FFQ type, follow up years, education, smoking status, alcohol intake, physical activity and diet scores. Combined model: Lifestyle model + fg-GRS. Statistically significant values (P<0.05) are marked in bold. Smoking status = non-smokers vs current smokers; education = school vs university education; physical activity= inactive vs active; alcohol intake= 1st vs 4th quartiles; ob-GRS=1st vs 4th quartiles; NNR Score= 1st vs 4th quartiles; HD Score= 1st vs 4th quartiles; PCA Score= 1st vs 4th quartiles.

IGT

The ability of the *genetic* model to predict IGT incidence was significantly lower than that of the *lifestyle* model (AUC 61% v. 64%; $P_{\text{difference}}$ =0.03). The predictive ability of the *combined* model (*lifestyle* model + genetic factors) was significantly better (AUC 65% $P_{\text{difference}}$ <0.03) than for the *lifestyle* model alone. The net reclassification improvement of the *combined* model compared to the *lifestyle* model was not statistically significant (8%, P =0.08). PA was significantly associated with incidence of IGT after adjustment for diet, smoking status, alcohol intake and education (see Table 3).

Table 3.
ORs for prediction of incident IGT according to lifestyle risk factors and GRSs.

	Regression model (OR, 95% CI)								
	NNR Score			HD Score			PCA Score		
Variables	**Genetic** (n= 2,420)	**Lifestyle** (n= 2,420)	**Combined** (n= 2,420)	**Genetic** (n= 2,509)	**Lifestyle** (n= 2,509)	**Combined** (n= 2,509)	**Genetic** (n= 2,420)	**Lifestyle** (n= 2,420)	**Combined** (n= 2,420)
Smoking status	-	0.92 (0.72-1.18)	0.91 (0.71-1.17)	-	0.91 (0.71-1.16)	0.9 (0.70-1.15)	-	0.94 (0.73-1.21)	0.93 (0.72-1.20)
Education	-	0.79 (0.59-1.05)	0.79 (0.59-1.05)	-	0.8 (0.60-1.06)	0.79 (0.59-1.05)	-	0.78 (0.58-1.04)	0.77 (0.58-1.03)
Alcohol intake	-	1.26 (0.92-1.71)	1.26 (0.93-1.72)	-	1.27 (0.94-1.72)	1.27 (0.94-1.73)	-	1.28 (0.94-1.75)	1.29 (0.95-1.76)
Physical activity	-	**0.69** (0.55-0.86)	**0.69** (0.55-0.86)	-	**0.74** (0.59-0.92)	**0.74** (0.59-0.93)	-	**0.67** (0.54-0.85)	**0.67** (0.54-0.85)
2hg-GRS	**1.44** (1.11-1.87)	-	**1.46** (1.13-1.90)	**1.45** (1.12-1.87)	-	**1.46** (1.13-1.89)	**1.44** (1.11-1.87)	-	**1.46** (1.13-1.90)
NNR Score	-	1.05 (0.80-1.39)	1.05 (0.80-1.38)	-	-	-	-	-	-
HD Score	-	-	-	-	0.94 (0.71-1.26)	0.94 (0.71-1.26)	-	-	-
PCA Score	-	-	-	-	-	-	-	**0.71** (0.53-0.95)	**0.71** (0.53-0.95)

Genetic model: age, age^2, sex, follow up years and 2hg-GRS. Lifestyle model: age, age^2, sex, FFQ type, follow up years, education, smoking status, alcohol intake, physical activity and diet scores. Combined model: Lifestyle model + 2hg-GRS. Statistically significant values (P<0.05) are marked in bold. Smoking status = non-smokers vs current smokers; education = school vs university education; physical activity= inactive vs active; alcohol intake= 1st vs 4th quartiles; ob-GRS=1st vs 4th quartiles; NNR Score= 1st vs 4th quartiles; HD Score= 1st vs 4th quartiles; PCA Score= 1st vs 4th quartiles.

For dietary factors, only the PCA score was significantly associated with the incidence of IGT.

Obesity

The difference in predictive ability between *genetic* and *lifestyle* models on obesity incidence was not statistically significant (AUC 68% v. 73%; $P_{\text{difference}}$ =0.08). The *combined* model (*lifestyle* model + genetic factors) significantly improved predictive ability (AUC 79% $P_{\text{difference}}$ <0.0001) of the *lifestyle* model. The net reclassification improvement of the *combined* model compared to the *lifestyle* model was 64% (P < 0.0001). Education and alcohol intake were significantly associated with obesity incidence, independent of diet, and PA (see Table 4). Notably, neither diet nor PA demonstrated statistically significant associations with obesity incidence.

Table 4.
ORs for prediction of incident obesity according to lifestyle risk factors and GRSs.

	Regression model (OR, 95% CI)								
	NNR Score			HD Score			PCA Score		
Variables	**Genetic** (n=1,511)	**Lifestyle** (n= 1,511)	**Combined** (n= 1,511)	**Genetic** (n=1557)	**Lifestyle** (n= 1,557)	**Combined** (n= 1,557)	**Genetic** (n= 1,511)	**Lifestyle** (n= 1,511)	**Combined** (n= 1,511)
Smoking status	-	1.36 (0.95-1.93)	1.37 (0.96-1.96)	-	1.34 (0.94-1.90)	1.35 (0.95-1.92)	-	1.36 (0.96-1.94)	1.38 (0.97-1.96)
Education	-	**0.58** (0.37-0.89)	**0.58** (0.37-0.90)	-	**0.57** (0.37-0.89)	**0.57** (0.37-0.89)	-	**0.57** (0.37-0.89)	**0.57** (0.37-0.89)
Alcohol intake	-	**0.62** (0.40-0.96)	**0.61** (0.39-0.95)	-	**0.61** (0.40-0.94)	**0.61** (0.39-0.94)	-	**0.62** (0.40-0.96)	**0.61** (0.39-0.95)
Physical activity	-	1.05 (0.77-1.44)	1.03 (0.75-1.41)	-	1.03 (0.76-1.41)	1.02 (0.75-1.39)	-	1.05 (0.77-1.43)	1.03 (0.75-1.41)
Ob-GRS	**2.02** (1.36-3.02)	-	**2.03** (1.35-3.05)	**1.97** (1.33-2.93)	-	**1.97** (1.32-2.94)	**2.02** (1.36-3.02)	-	**2.02** (1.35-3.04)
NNR Score	-	0.91 (0.62-1.34)	0.93 (0.63-1.37)	-	-	-	-	-	-
HD Score	-	-	-	-	1.27 (0.84-1.94)	1.26 (0.83-1.93)	-	-	-
PCA Score	-	-	-	-	-	-	-	0.97 (0.64-1.47)	0.98 (0.65-1.48)

Genetic model: age, age^2, sex, follow up years and ob-GRS. Lifestyle model: age, age^2, sex, FFQ type, follow up years, education, smoking status, alcohol intake, physical activity and diet scores. Combined model: Lifestyle model + ob-GRS. Statistically significant values (P<0.05) are marked in bold. Smoking status = non-smokers vs current smokers; education = school vs university education; physical activity= inactive vs active; alcohol intake= 1st vs 4th quartiles; ob-GRS=1st vs 4th quartiles; NNR Score= 1st vs 4th quartiles; HD Score= 1st vs 4th quartiles; PCA Score= 1st vs 4th quartiles.

The ability of the *genetic* and *lifestyle* models to predict an incidence of ≥10% weight gain was AUC = 65% (for both models). The predictive ability of the *combined* model (lifestyle + genetic factors) was significantly better than the *lifestyle* model alone (AUC 68% $P_{\text{difference}}$ =0.0004). The net reclassification improvement for the *combined* model compared to the *lifestyle* model alone was 26% (P <0.0001).

Viewing the results in the context of other published work

We observed no improvement in reclassification of IGT incidence when adding the genetic factors to the *lifestyle* model. This may be partly explained by the low number of SNPs (9 SNPs) associated with 2hr glucose. For other outcomes where more loci are available, we observed greater improvements from adding genetic factors; in obesity 61% (97 SNPs), T2D 52% (65 SNPs), and IFG incidence 33% (36 SNPs associated with fasting glucose). The limited addition to predictive ability of genetic factors in screening has been noted by others. A review by Lyssenko and Laakso, which included older studies often with fewer SNPs, concluded that genetic markers add little predictive ability to clinical factors[92]. Adding other clinical factors to the models, however, may improve the ROC AUC. Lyssenko and Laakso[92] reported on several studies where BMI (and, in most, fasting glucose) are included in predictive models for T2D, and have a ROC

AUC above 90% for the incidence of T2D[93-96]. Here, we did not include intermediate clinical phenotypes in the prediction models, as the objective was to compare the predictive ability of modifiable lifestyle factors against a set of easily measured and salient biological markers (genotypes). Including intermediate clinical phenotypes, such as BMI and fasting glucose, in the comparison model is counter-intuitive in this context, as this would lead to the underestimation of the predictive ability of lifestyle and genotypes due to mediation[97]. We did, however, adjust for age and sex, as they are not on the causal pathway between either lifestyle or genetic exposures, and the index outcomes (i.e., lifestyle factors cannot affect the sex or age of a participant). Moreover, excluding age and sex from the models could cause bias and confounding. We did not include gene-environment interaction terms in our models, as the study is unlikely to be powered, and multiplicative interaction effects are unlikely to have a major impact on predictive ability[98].

Our findings agree with existing literature, but results from analyses carried out in North American[95], Nordic[99] and British[100] cohorts demonstrated that the inclusion of 11-20 T2D SNPs did not improve the predictive ability of clinical risk factor scores for T2D incidence. Nevertheless, all of these analyses included family history of T2D in their clinical variable model, which, to some extent, reflects genetic influences on T2D risk. Thus, it is not surprising that adding T2D-associated variants to a model already including family history of the disease does not improve the model's predictive ability substantially. However, an update to the Whitehall II study[101] including 65 T2D associated SNPs, showed that including genetic factors in a clinical risk-factor model did improve predictive power, mirroring our own findings.

Limitations

The self-report methods used here, though clinically practicable, can lead to self-report bias. For instance, participants may systematically underreport unhealthy behaviours, which would make them less reliable predictors.

The SNPs used in this analysis have been discovered in a cross-sectional analysis, which may not reflect the same biological mechanisms that govern changes over time.

Paper 1 Conclusions

In a head-to-head comparison, this study illustrates comparable predictive ability of selected lifestyle and genetic factors for IFG, IGT, T2D and obesity incidence. We also show that when genetic factors are added to the *lifestyle* model, a significant net reclassification improvement was achieved for IFG (36%), T2D (58%) and obesity (64%). In past studies, this improvement has not been reported, possibly because of a low number of SNPs available, and clinical factors (including phenotypic mediators) explaining some of the genetic effects.

The findings of this study add to this thesis by illustrating that lifestyle factors and genetic factors complement each other in a model to predict incidence of obesity and T2D, and that neither surpasses the other. This strengthens the notion that genetic and lifestyle factors provide informative context for the biological effects each have. They are predictors which can be used in parallel, as well as in addition to, other clinical variables. Moreover, the study illustrates how lifestyle factors representing an origin risk factor (i.e. in a causal pathway, they would begin the cascade of effects that cause the outcome) and not an intermediate phenotype (i.e. caused by a combination of lifestyle and genetic effects that are upstream in the causal pathway) can add predictive value for glycaemic control and adiposity traits.

Baseline visit results of the IMI DIRECT Cohorts (Paper 2 and Paper 3)

Here I describe the baseline characteristics of the two prospective cohorts in DIRECT focusing on glycaemic deterioration. I also discuss some potential mechanisms behind the differences in glycaemic control between ADA glycaemic strata[85] in Cohort 1 and treatment strata in the Cohort 2, with an emphasis on the role of PA.

Recruitment

Cohort 1

Participants for the 'prediabetes' cohort in DIRECT (Study 1 in Paper 2), were recruited from existing population cohort studies in Europe. We did this by applying the DIRECT-DETECT glycaemic deterioration risk prediction algorithm (described in the Methods) to identify those at higher risk of rapid glycaemic

deterioration. We recruited 2,335 participants (10%) from a combined sampling frame of 24,196 participants across four cohorts as follows:

In Kuopio, we recruited 1,340 participants from the Metabolic Syndrome in Men (METSIM) study (total N=6,414)[102].

In Hoorn/Amsterdam, we recruited 500 from a total of 2,607 participants enrolled in studies around Amsterdam. Of these, 18 (out of 76) participants were recruited from the Relationship between Insulin Sensitivity and Cardiovascular disease (RISC) cohort[103]; 48 participants (out of 201) were recruited from the Hoorn Meal Study (HMS) and; 434 participants (out of 2,330) were recruited from the New Hoorn Study (NHS)[104].

In Copenhagen, we recruited 326 participants from a total sample frame of 11,441. Of these, 56 (out of 1,522) were recruited from the Health2010 study[105]; 172 (out of 2,308) were recruited from the Health2006 study[106]; 87 (out of 7439) were recruited from the Danish Study of Functional Disorders (DanFunD)[107]; 11 (out of 118) were recruited from the Gut, Grain and Greens(GGG) study[108].

In Malmö, we recruited 169 participants out of a sampling-frame of 3,734 participants from the Malmö Diet and Cancer (MDC) study[109].

Thus, the 2335 participants in Cohort 1 constitutes 58%, 21%, 14% and 7% participants from Kuopio, Amsterdam/Hoorn, Copenhagen and Malmö, respectively. Of the 2,335 participants recruited: 2,247 had the minimum measurements (standard clinical and eligibility criteria variables, and a fasting blood sample); 490 participants had apparently NGR; 545 participants had iIA1c; 369 participants had iIFG; 38 participants had iIGT; and, 805 participants had cIGR.

The DIRECT-DETECT algorithm was used to help maximize the rate of glycaemic deterioration within Cohort 1; thus, participants were purposefully recruited with the highest DIRECT-DETECT scores first. However, the sampling frame of around 24000 was insufficient to ensure only participants with dysregulated glucose were recruited; around 500 participants who had apparently normal glycaemic regulation were also included in the cohort, as a matter of necessity.

Cohort 2

Participants for the 'type 2 diabetes' cohort, in IMI DIRECT (Study 2 in Paper 2) were recruited from General Practice clinics and other clinical registries from several northern European countries. We recruited a total of 830 participants with new-onset T2D (diagnosed between 6 and 24 months before the baseline examination): 184 (22%) from Dundee, UK; 170 (20%) in Exeter, UK; 169 (20%) participants in Hoorn/Amsterdam, Netherlands; 146 (18%) participants in

Newcastle upon Tyne, UK; 107 (13%) participants in Lund, Sweden; and, 54 (7%) participants in Copenhagen, Denmark. Of the 830 participants recruited to the T2D cohort, 804 participants had minimum measurements (standard clinical and eligibility criteria variables, and a fasting blood sample) available.

Baseline Characteristics

Tables 5-14 describe the baseline characteristics of the main phenotypic variables in the two IMI DIRECT glycaemic deterioration cohorts. Values are presented as mean (sem) adjusted for age, sex and recruitment centre, unless indicated otherwise. It is important to note that observed differences in characteristics between the strata are meant to be descriptive rather than to make any causal inferences. This is of particular importance for data from Cohort 2, where observed differences between participants treated with lifestyle only vs. lifestyle plus metformin therapy are not necessarily true (because the latter group also receive metformin). As treatment guidelines are based on glycaemic control, participants with higher glucose concentrations are those most likely to receive metformin treatment (e.g. in this case, adding metformin to lifestyle changes). Accordingly, patients who respond well to lifestyle therapy may not progress to metformin during the course of the DIRECT study.

Clinical characteristics

Table 5 shows the baseline clinical characteristics of Cohort 1 stratified by ADA glycaemic strata[85]. The majority (76%) of participants in Cohort 1 were male, largely because around 57% of the participants were recruited from METSIM, which is a male-only study. Of note, 95% of the 368 participants who had iIFG were male. Mean age, BMI, waist circumference, systolic blood pressure, and diastolic blood pressure all differed between glycaemic strata ($P_{\text{difference}} < 0.0001$).

Table 5.
Baseline clinical characteristics by ADA glycaemic control strata for Cohort 1.

	Cohort 1 ("Prediabetes")						
	NGR	**iIA1c**	**iIFG**	**iIGT**	**cIGR**	**All**	$P_{\text{difference}}$
N	485	537	368	38	799	**2227**	
Male (%)*	81	58	95	76	76	**76**	
Age (yrs)*	61.2 (0.3)	62.3 (0.3)	60.4 (0.3)	64 (1.1)	61.8 (0.2)	**61.6 (0.1)**	<0.001
BMI ($kg \cdot m^{-2}$)	27.9 (0.2)	27.7 (0.2)	28.9 (0.2)	29.1 (0.6)	29.7 (0.2)	**28.7 (0.2)**	<0.001
Waist circumference (cm)	98.1 (0.5)	97.3 (0.5)	101.2 (0.6)	100.3 (1.7)	102.9 (0.4)	**99.9 (0.4)**	<0.001
Systolic blood pressure (mmHg)	127.9 (0.8)	127.5 (0.7)	130.2 (0.9)	132.1 (2.4)	131.6 (0.6)	**129.9 (0.6)**	<0.001
Diastolic blood pressure (mmHg)	77.3 (0.4)	77.4 (0.4)	78.8 (0.5)	79.6 (1.4)	79.5 (0.3)	**78.5 (0.4)**	<0.001
Smoking (% current / ex- / never)*	15/50/35	18/42/40	12/47/41	11/58/32	13/51/36	**14/48/38**	
Alcohol Usage (% regularly / occasionally /never)*	67/20/13	63/24/14	74/14/12	14/12/71	71/21/8	**69/19/12**	

Values are mean (sem) adjusted for age, sex and centre unless indicated otherwise, * not adjusted for age, sex and centre.

Table 6 shows the baseline clinical characteristics of Cohort 2 stratified by treatment. Participants on lifestyle treatment only, compared with participants on lifestyle and metformin treatment, were on average 1 year older and had a systolic blood pressure 3 mmHg greater ($P_{\text{difference}}$ = 0.025 and $P_{\text{difference}}$ = 0.029, respectively).

Table 6.
Baseline clinical characteristics by treatment for Cohort 2.

	Cohort 2 ("Diabetes")			
	LS	**Met+LS**	**All**	$P_{\text{difference}}$
N	527	277	**804**	
Male (%)*	58	56	**57**	
Age (yrs)*	62 (0.3)	61 (0.5)	**62 (0.3)**	0.025
BMI (kg·m^{-2})	30.3 (0.2)	30.6 (0.3)	**30.4 (0.2)**	0.4
Waist circumference (cm)	103 (0.6)	104 (0.8)	**103 (0.5)**	0.2
Systolic blood pressure (mmHg)	133 (0.9)	130 (1.1)	**131 (0.7)**	0.029
Diastolic blood pressure (mmHg)	76 (0.5)	75 (0.6)	**76 (0.4)**	0.3
Smoking (% current / ex- / never)*	13/50/37	15/49/36	**14/50/37**	
Alcohol Usage (% regularly / occasionally /never)*	58/24/18	58/27/15	**58/25/17**	

Values are mean (sem) adjusted for age, sex and centre unless indicated otherwise, * not adjusted for age, sex and centre.

Glycaemic Measures

The glycaemic measures in Cohort 1 are derived from 2-hr fsOGTT data (see materials and methods section for more details). Glycaemic measures in Cohort 2 are measured and modelled from a 2-hr MMTT (see materials and methods section for more details).

Table 7 shows the baseline glycaemic control characteristics of Cohort 1 stratified according to ADA criteria[85]. All glycaemic measures differ by glycaemic control group ($P_{\text{difference}} < 0.0001$).

Table 7.
Baseline glycaemic characteristics by ADA glycaemic control strata for Cohort 1.

	Cohort 1 ("Prediabetes")						
	NGR	**iIA1c**	**iIFG**	**iIGT**	**cIGR**	**All**	$P_{\text{difference}}$
N	490	545	369	38	805	**2247**	
HbA1c (mmol·mol^{-1})	36.8 (0.1)	41.7 (0.1)	37.4 (0.2)	37 (0.4)	42.6 (0.1)	**39.1 (0.1)**	<0.001
Fasting glucose (mmol·L^{-1})	5.2 (0)	5.2 (0)	5.9 (0)	5.3 (0.1)	6 (0)	**5.5 (0)**	<0.001
Mean 2-hr glucose (mmol·L^{-1})	6.7 (0.1)	7 (0.1)	7.6 (0.1)	9 (0.2)	8.8 (0.1)	**7.8 (0.1)**	<0.001
2-hr glucose (mmol·L^{-1})	5.4 (0.1)	5.6 (0.1)	5.8 (0.1)	8.9 (0.3)	7.4 (0.1)	**6.6 (0.1)**	<0.001
Fasting insulin (pmol·l^{-1})	52.6 (2)	53.3 (1.9)	66.6 (2.4)	65.7 (6.5)	75.8 (1.6)	**62.8 (1.7)**	<0.001
Mean 2-hr insulin (pmol·l^{-1})	315 (13)	330 (12)	392 (16)	468 (42)	448 (11)	**391 (11)**	<0.001
2-hr insulin (pmol·l^{-1})	37 (2)	42 (2)	46 (3)	97 (8)	74 (2)	**59 (2)**	<0.001
Fasting insulin secretion (pmol·min·m^{-2})	94 (2)	97 (2)	113 (2)	114 (6)	124 (2)	**108 (2)**	<0.001
Integral of total insulin secretion (nmol·m^{-2})	47 (1)	48 (1)	52 (1)	62 (3)	57 (1)	**53 (1)**	<0.001
Glucose sensitivity (pmol·min^{-1}·m^{-2}·mM^{-1})	124 (3)	122 (3)	114 (3)	97 (9)	101 (2)	**112 (2)**	<0.001
Potentiation factor ratio	1.7 (0)	1.7 (0)	1.9 (0)	1.2 (0.1)	1.6 (0)	**1.6 (0)**	<0.001
Rate sensitivity (pmol·m^{-2}·mM^{-1})	1102 (35)	933 (32)	989 (42)	1041 (112)	859 (28)	**985 (29)**	<0.001
Insulin sensitivity 2-h OGIS (ml·min^{-1}·m^{-2})	419 (3)	413 (2)	366 (3)	365 (8)	344 (2)	**381 (2)**	<0.001

Values are mean (sem) adjusted for age, sex and centre unless indicated otherwise, * not adjusted for age, sex and centre.

As expected, HbA1c, fasting glucose, and 2-hr glucose differ between the ADA glycaemic control strata[85]. However, given that poor glycaemic control is often an indicator of poor insulin sensitivity and incapacity to compensate by increasing insulin secretion, we can surmise that participants with either iIGT or cIGR are likely to show signs of reduced insulin sensitivity and beta-cell function. Indeed, we see that insulin sensitivity (by 2-hr OGIS) is substantially lower in the cIGR strata and higher in the NGR and iIA1c strata. Similarly, for beta-cell function, we see that glucose sensitivity (dose response slope of insulin secretion in response to glucose) and rate sensitivity (early secretion parameter) are lower in cIGR compared with NGR, and even iIA1c and iIFG. However, the opposite is apparent for iIGT. There are two possible explanations for these observations. The first is that loss of beta-cell function, associated with progression to T2D[110], may also affect participants in IGR[111]. A study tested the association between insulin sensitivity and insulin secretion rate (both assessed by euglycaemic hyperinsulinaemic clamp, unlike fsOGTT/MMTT, as we have in DIRECT) in 156 and 1,123 participants with IGR and NGR, respectively[111]. They found that insulin sensitivity was associated with insulin secretion rate in IGR, though not with beta-cell glucose sensitivity. As we measure insulin sensitivity from fsOGTT/MMTTs rather that euglycaemic hyperinsulinaemic clamps, we may see a relationship which they did not. A second explanation could be that a drink with 75g of glucose (the fsOGTT) may not elicit a maximum insulin secretion response. Instead these participants experience an adequate compensatory rise in insulin secretion based on their glucose and insulin sensitivity.

Table 8 shows the baseline glycaemic control characteristics of Cohort 2 stratified by treatment. HbA1c, fasting glucose, mean-2-hr glucose (mean glucose during 2-hr OGTT), 2-hr glucose and fasting insulin secretion were all higher in the group receiving both lifestyle and metformin (Met+LS) treatment ($P_{\text{difference}} \leq 0.012$). Glucose sensitivity and potentiation fraction ratio (indication of beta-cell function at late stage insulin secretion) were reduced in the Met+LS strata ($P_{\text{difference}} \leq 0.004$).

Table 8.
Baseline glycaemic characteristics by treatment for Cohort 2.

	Cohort 2 ("Diabetes")			
	LS	**Met+LS**	**All**	$P_{difference}$
N	527	277	**804**	
HbA1c (mmol·mol^{-1})	46 (0.3)	47 (0.4)	**47 (0.2)**	0.005
Fasting glucose (mmol·L^{-1})	7 (0.07)	7.3 (0.09)	**7.1 (0.06)**	0.012
Mean 2-hr glucose (mmol·L^{-1})	9 (0.1)	9.9 (0.1)	**9.5 (0.1)**	<0.001
2-hr glucose (mmol·L^{-1})	8.2 (0.1)	9.4 (0.2)	**8.8 (0.1)**	<0.001
Fasting insulin (pmol·l^{-1})	101 (3.4)	110 (4.3)	**105 (2.7)**	0.1
Mean 2-hr insulin (pmol·l^{-1})	448 (14)	440 (17)	**444 (11)**	0.7
2-hr insulin (pmol·l^{-1})	422 (17)	443 (21)	**432 (14)**	0.5
Fasting insulin secretion (pmol·min·m^{-2})	130 (2)	140 (3)	**135 (2)**	0.006
Integral of total insulin secretion (nmol·m^{-2})	43 (1)	44 (1)	**43 (1)**	0.4
Glucose sensitivity (pmol·min^{-1}·m^{-2}·mM^{-1})	88 (3)	73 (4)	**81 (3)**	0.004
Potentiation factor ratio	1.5 (0.03)	1.3 (0.04)	**1.4 (0.03)**	<0.001
Rate sensitivity (pmol·m^{-2}·mM^{-1})	1078 (55)	1135 (69)	**1107 (44)**	0.5
Insulin sensitivity 2-h OGIS (ml·min^{-1}·m^{-2})	305 (4)	295 (5)	**300 (3)**	0.1

Values are mean (sem) adjusted for age, sex and centre unless indicated otherwise, * not adjusted for age, sex and centre.

The apparently poor glycaemic control in the Met+LS strata is unlikely to be caused by metformin. To the contrary, it is likely to be the reverse, where worse glycaemic control prompts treatment with metformin. As indicated in the Methods, participants who were taking metformin were asked not to take metformin for 48 hours before the visit, in part removing its effect from these results.

Comparisons between cohorts 1 and 2

Because there is a difference between the glycaemic control assessment protocol in the two cohorts (fsOGTT in Cohort 1; MMTT in Cohort 2), comparison between the two cohorts should be done with caution when relating to glycaemic control. The modelled parameters presented here are based on the measured levels of glucose, insulin and c-peptide in the blood, as well as some phenotypic variables and the nutritional content of the test drink. Thus, the main differences in the protocol are accounted for in the modelling. For instance, we can see markedly lower glucose sensitivity in Cohort 2. This could be indicative of the reduced beta cell function in participants with T2D, who cannot adequately compensate for their insulin resistance with increased insulin secretion. We also see similar expected differences in other important glycaemic control parameters (e.g. fasting glucose, 2-hr glucose, insulin sensitivity, rate sensitivity and potentiation fraction ratio). However, making any inferences based on these analyses (that the condition of T2D is a causal factor) is not possible due to the protocol difference, as mentioned previously.

Physical Activity

All PA parameters are objectively measured using an ActiGraph GT3X+ device worn continuously on the non-dominant wrist for 10 days (see materials and methods section for more details).

Table 9 shows the baseline PA characteristics of Cohort 1 stratified by ADA glycaemic strata[85]. The mean daily PA intensity in Cohort 1 was 36.1 mGs (hpfVM) or 24.7 mGs (ENMO). On average, participants were: i) sedentary 83% of the time; ii) in light-intensity PA 11%; iii) in moderate-intensity PA 5%; and, iv) in vigorous-intensity PA 1% of the time. We observe that mean values of all hpfVM parameters except vigorous intensity PA were significantly different between glycaemic strata ($P_{\text{difference}} \leq 0.016$). Vigorous PA, which only made up a mean of 1.5% of the time, showed a nominally statistically significant difference ($P_{\text{difference}} = 0.054$).

Table 9.

Baseline physical activity characteristics by glycaemic control strata for Cohort 1.

	Cohort 1 ("Prediabetes")						
	NGR	**iIA1c**	**iIFG**	**iIGT**	**cIGR**	**All**	$P_{\text{difference}}$
N	391	444	284	31	666	**1816**	
Average physical activity intensity - ENMO (mGs)	24.7 (0.8)	24.1 (0.7)	25.1 (0.9)	25.8 (2.5)	24.2 (0.6)	**24.8 (0.6)**	0.8
Average physical activity intensity - hpfVM (mGs)	37.1 (0.5)	37.5 (0.5)	37.1 (0.7)	33.2 (1.8)	35.8 (0.4)	**36.1 (0.5)**	0.009
Percent sedentary intensity (<48 mGs hpfVM)	82.2 (0.2)	81.9 (0.2)	82.2 (0.3)	84.1 (0.8)	82.7 (0.2)	**82.6 (0.2)**	0.003
Percent light intensity (48-154 mGs hpfVM)	10.9 (0.1)	11 (0.1)	10.8 (0.2)	9.8 (0.4)	10.6 (0.1)	**10.6 (0.1)**	0.003
Percent moderate intensity (154-389 mGs hpfVM)	5.29 (0.08)	5.34 (0.08)	5.28 (0.1)	4.75 (0.27)	5.09 (0.07)	**5.15 (0.07)**	0.016
Percent vigorous intensity (>389 mGs hpfVM)	1.51 (0.04)	1.52 (0.04)	1.51 (0.05)	1.29 (0.13)	1.42 (0.03)	**1.45 (0.03)**	0.054

Values are mean (sem) adjusted for age, sex and centre unless indicated otherwise, * not adjusted for age, sex and centre.

Though the results indicate that there is heterogeneity in PA behaviour between the glycaemic strata, the differences do not appear to be substantial, with the exception of the small group of participants with iIGT. Participants in this strata did about 10% less light-intensity PA than the participants in the NGR, iIA1c and iIFG strata. This result is also established in existing literature [112] in a free-living setting with a robust objective measurement method.

Table 10 shows the baseline PA characteristics of Cohort 2 stratified by treatment. The mean daily PA intensity in Cohort 2 was 34mGs (hpfVM) or 22 mGs (ENMO). On average, participants were sedentary 83% of the time, in light-intensity PA 10%, in moderate-intensity PA 5%, and the remaining 2% in vigorous PA intensity. We observe that all PA parameters were statistically different between glycaemic strata ($P_{\text{difference}} < 0.0001$).

Table 10.
Baseline physical activity characteristics by treatment for Cohort 2.

	Cohort 2 ("Diabetes")			
	LS	Met+LS	All	$P_{difference}$
N	483	253	**736**	
Average physical activity intensity - ENMO (mGs)	23 (0.4)	20 (0.5)	**22 (0.3)**	<0.001
Average physical activity intensity - hpfVM (mGs)	36 (0.5)	32 (0.6)	**34 (0.4)**	<0.001
Percent sedentary intensity (<48 mGs hpfVM)	82.4 (0.22)	84 (0.27)	**83.2 (0.18)**	<0.001
Percent light intensity (48-154 mGs hpfVM)	10.8 (0.12)	10 (0.15)	**10.4 (0.1)**	<0.001
Percent moderate intensity (154-389 mGs hpfVM)	5.2 (0.08)	4.6 (0.1)	**4.9 (0.06)**	<0.001
Percent vigorous intensity (>389 mGs hpfVM)	1.4 (0.03)	1.2 (0.04)	**1.3 (0.03)**	<0.001

Values are mean (sem) adjusted for age, sex and centre unless indicated otherwise, * not adjusted for age, sex and centre.

The differences in PA between the treatment strata in Cohort 2 appear to be more substantial than between the glycaemic strata in Cohort 1. Average PA intensity appears to be about 11-13% less in the metformin group. Light, moderate and vigorous PA was 8%, 11%, and 14% less in the metformin group. Though PA is measured identically in Cohort 1 and 2, allowing a comparison between the cohorts, such analyses were not carried out as the two cohorts have not been designed with such comparisons in mind. Notwithstanding, a visual comparison suggests that Cohort 2 is less active with potentially around 10% less time spent in vigorous PA. Such comparisons should be considered with care, as the age, sex and centre adjustments do not apply when comparing the adjusted means between the two cohorts. Within Cohort 2, a plausible explanation for the metformin treated strata being less active than the lifestyle + metformin treated group is not that they failed to respond to the lifestyle treatment, but rather that they failed to adhere to it, and thus did not benefit.

Diet

Diet parameters in both cohorts are derived from self-reported multi-pass food habit questionnaires, in combination with a 24-hr diet record (see materials and methods section for more details).

Table 11 shows the baseline diet characteristics of Cohort 1 stratified by ADA glycaemic strata[85]. Total intake of energy, carbohydrate, protein, and sugar differed between glycaemic strata ($P_{\text{difference}} \leq 0.014$).

Table 11.
Baseline diet characteristics by ADA glycaemic control strata for Cohort 1.

	Cohort 1 ("Prediabetes")						
	NGR	iIA1c	iIFG	iIGT	cIGR	All	$P_{difference}$
Total Energy Intake (kCal)	1842 (38)	1930 (35)	1852 (46)	1580 (121)	1812 (32)	**1803 (32)**	0.014
Total Carbohydrate (g)	201.8 (4.9)	218.2 (4.5)	199.5 (5.8)	184.8 (15.5)	197.1 (4)	**200.3 (4.1)**	0.002
Total Fat (g)	76.4 (2)	79.4 (1.9)	76.3 (2.4)	63.6 (6.4)	75.7 (1.7)	**74.3 (1.7)**	0.1
Total Protein (g)	92 (2.2)	92.1 (2)	97.1 (2.7)	71.4 (7)	90.4 (1.8)	**88.6 (1.9)**	0.005
Total Sugar (g)	89.3 (2.7)	96.9 (2.5)	89.6 (3.3)	85.5 (8.7)	85.7 (2.3)	**89.4 (2.3)**	0.009
Total Fibre (g)	18.8 (0.5)	18.9 (0.5)	17.9 (0.6)	15.9 (1.6)	17.6 (0.4)	**17.8 (0.4)**	0.054
Total Saturated Fat (g)	28.7 (0.9)	29.8 (0.8)	28 (1)	23.8 (2.7)	28.5 (0.7)	**27.8 (0.7)**	0.2
Total Monounsaturated Fat (g)	25.6 (0.9)	26.6 (0.8)	24.6 (1)	19.8 (2.8)	25.1 (0.7)	**24.3 (0.7)**	0.1
Total Polyunsaturated Fat (g)	11.7 (0.4)	12.7 (0.4)	12.4 (0.5)	9.4 (1.3)	11.7 (0.4)	**11.6 (0.4)**	0.056

Values are mean (sem) adjusted for age, sex and centre unless indicated otherwise, * not adjusted for age, sex and centre.

The heterogeneity in dietary intake we see in Cohort 1 pertains to participants in the iIGT and iIA1c strata. Participants in the iIGT strata had lower total energy intake than in iIA1c and lower total protein intake than in the NGR, iIA1c and iIFG strata ($P_{\text{difference}} < 0.05$) in pairwise difference tests. The total energy intake in the iIA1c group being 18% higher than in the iIGT group is of note ($P_{pairwise\ difference} <0.05$, analyses described in more detail in paper). The participants in the iIA1c strata also reported a higher total carbohydrate and sugar intake than participants in the cIGR strata ($P_{\text{difference}} < 0.05$).

Table 12 shows the baseline diet characteristics of Cohort 2 stratified by treatment. We observe no statistically significant differences between participants in the Met+LS and lifestyle treatment strata.

Table 12.
Baseline diet characteristics by treatment for Cohort 2.

	Cohort 2 ("Diabetes")			
	LS	Met+LS	All	$P_{difference}$
N	493	229	**722**	
Total Energy Intake (kCal)	1826 (28)	1806 (40)	**1816 (24)**	0.7
Total Carbohydrate (g)	206 (3.6)	209 (5.1)	**208 (3.1)**	0.7
Total Fat (g)	73 (1.6)	71 (2.2)	**72 (1.3)**	0.4
Total Protein (g)	88 (1.4)	85 (2)	**86 (1.2)**	0.2
Total Sugar (g)	84 (2)	84 (2.8)	**84 (1.7)**	0.9
Total Fibre (g)	19 (0.4)	18 (0.6)	**19 (0.3)**	0.1
Total Saturated Fat (g)	26 (0.7)	27 (0.9)	**27 (0.6)**	0.5
Total Monounsaturated Fat (g)	25 (0.6)	24 (0.9)	**24 (0.5)**	0.4
Total Polyunsaturated Fat (g)	12 (0.4)	11 (0.5)	**12 (0.3)**	0.1

Values are mean (sem) adjusted for age, sex and centre unless indicated otherwise, * not adjusted for age, sex and centre.

MRI

MRI-derived adiposity estimates were carried out in roughly half of the participants in Cohort 1 and 60% in Cohort 2 using a multiecho technique (see

materials and methods section for details). No MRIs were carried out in the Lund/Malmö study sites for either cohort.

Table 13 shows the baseline MRI-derived adiposity characteristics of Cohort 1 stratified by ADA glycaemic strata[85]. Total abdominal adipose tissue (TAAT), visceral fat, liver fat and pancreatic fat differ between glycaemic strata ($P_{\text{difference}} \leq 0.024$).

Table 13.
Baseline MRI characteristics by ADA glycaemic control strata for Cohort 1.

	Cohort 1 ("Prediabetes")						
	NGR	**iIA1c**	**iIFG**	**iIGT**	**cIGR**	**All**	$P_{\text{difference}}$
N	217	200	194	11	399	**1021**	
Total Abdominal Adipose Tissue (Litres)	11.5 (0.4)	12.7 (0.4)	12.5 (0.7)	11.9 (0.8)	12.1 (0.5)	**12.1 (0.4)**	<0.001
Visceral Fat (Litres)	4.6 (0.2)	4.9 (0.2)	5 (0.4)	4.9 (0.4)	4.5 (0.3)	**4.8 (0.2)**	<0.001
Liver Fat (%)	5.1 (0.6)	5.6 (0.6)	5.6 (1)	5.6 (1.1)	6 (0.7)	**5.6 (0.5)**	<0.001
Pancreatic Fat (%)	11.3 (0.8)	12.2 (0.8)	13.4 (1.4)	13.1 (1.5)	9.9 (0.9)	**12 (0.7)**	0.024
Liver Iron Content (S^{-1})	55.3 (1.7)	55.4 (1.7)	56.3 (3)	56.8 (3.2)	55.6 (2)	**55.9 (1.5)**	0.8
Pancreatic Iron Content (S^{-1})	47.3 (1.6)	47.4 (1.6)	51.1 (2.8)	51.9 (3)	43 (1.8)	**48.2 (1.4)**	0.067

Values are mean (sem) adjusted for age, sex and centre unless indicated otherwise, * not adjusted for age, sex and centre.

Despite the heterogeneity across all strata, in pairwise comparisons we only observe a statistically significant difference in TAAT between participants in the NGR and iIA1c strata, where NGR had roughly 10% smaller mean TAAT. For the key measures of visceral, liver and pancreatic fat, we do not observe statistically significant pairwise difference between glycaemic strata, despite findings reported by others, suggesting they are related to glycaemic regulation[71,113-115]. There are some plausible explanations, one is that the participants in Cohort 1 with impaired glycaemic control are still too close to the NGR end of the spectrum of glycaemic regulation. Another explanation could be that the type of lipids making up the detected fat are not those (or in that sub fraction) that are causing the dysregulation detected by others. Regarding pancreatic fat, a further explanation could be that our measurement methods may not be able to distinguish between intra-organ fat, or fat surrounding the organ, especially under circumstances where pancreatic shape is irregular[113,116].

Table 14 shows the baseline MRI derived adiposity characteristics of Cohort 2 stratified by treatment. In this cohort, we observe that participants in the Met+LS treated strata have nearly 25% greater liver fat percentage than those receiving lifestyle treatment only ($P_{\text{difference}} < 0.0001$).

Table 14.
Baseline MRI characteristics by treatment for Cohort 2.

	Cohort 2 ("Diabetes")			
	LS	Met+LS	All	$P_{difference}$
N	357	157	**514**	
Total Abdominal Adipose Tissue (Litres)	13 (0.3)	13.7 (0.4)	**13.4 (0.2)**	0.2
Visceral Fat (Litres)	5.1 (0.1)	5.4 (0.2)	**5.2 (0.1)**	0.1
Liver Fat (%)	7.3 (0.4)	9.7 (0.5)	**8.5 (0.3)**	<0.001
Pancreatic Fat (%)	11 (0.4)	12 (0.6)	**12 (0.3)**	0.2
Liver Iron Content (S^{-1})	54 (1)	54 (1.3)	**54 (0.8)**	0.7
Pancreatic Iron Content (S^{-1})	46 (1)	46 (1.3)	**46 (0.8)**	0.9

Values are mean (sem) adjusted for age, sex and centre unless indicated otherwise, * not adjusted for age, sex and centre.

The marked increase in liver fat in the Met+LS strata in Cohort 2 may be due to these participants having progressed further in T2D than those in the LS strata. This would lend weight to the hypothesis that liver fat is a central intermediate phenotype in the pathogenesis of T2D, as reported by others[30,31] (see introduction for more on this hypothesis).

Genotyping

Figure 8 illustrates the genetic population substructure of the two cohorts in DIRECT.

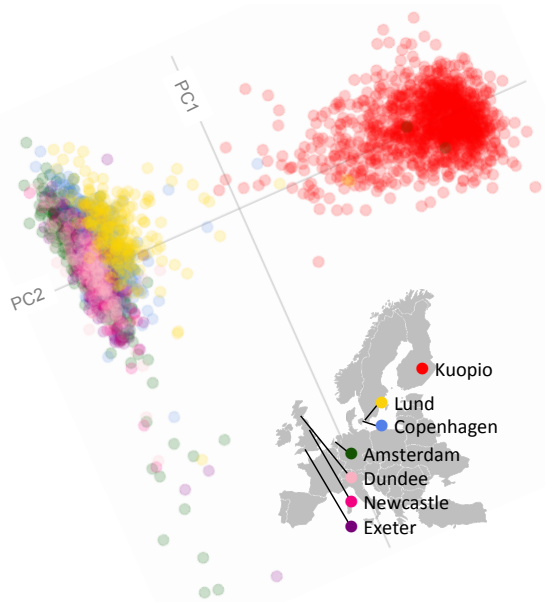


Figure 8
Population structure within WP2 baseline study samples. A statistical summary of genetic data from Cohorts 1 and 2 based on principal component axis one (PC1) and axis two (PC2). Circles represent "Prediabetes" subjects and stars represent "New-onset Diabetes" subjects. Circles and stars are coloured as per the recruitment centres. Red, Kuopio; Yellow, Lund; Blue, Copenhagen; Green, Amsterdam; Pink, Newcastle; Salmon, Dundee; Purple, Exeter.

In agreement with findings by others, we can see in Figure 7 that the genetic substructure maps to the geographic location of the study centres[117]. This illustrates ethnic homogeneity within study centres and heterogeneity between study centres. The main heterogeneity in these cohorts is driven by inclusion of Finnish participants, a population known to be an ancestral genetic isolate[118].

Paper 2 and Paper 3 Conclusions

The ability to generate data from physiological measurements and cost-effectively store biosamples long-term has advanced over recent years. At the same time, increasing prevalence of T2D and the development of methods to analyse complex multivariable prospective datasets continues to drive the demand for well phenotyped prospective datasets, coupled with biobank material. Papers 2 and 3 illustrate recent progress in one of the largest international efforts to date trying to address this demand for the study of glycaemic deterioration. As data availability from these cohorts grows, and analysts worldwide begin to use them, understanding the context of these data and coupled biobank material becomes pivotal to informative analyses. The purpose of papers 2 and 3 have been to facilitate this process and highlight potential future research questions.

The results described in Paper 3 highlight heterogeneity in T2D-related metabolic and lifestyle factors, between established glycaemic strata in prediabetes, and treatment strata in T2D. Among the differentiating factors were PA, diet, intra-abdominal fat and glycaemic control parameters.

The role of physical activity in metabolic homeostasis before and after the onset of type 2 diabetes (Paper 4)

Paper 4 describes an analysis that focuses on the complex regulation of beta-cell function and glucose control. Of all papers in this thesis, Paper 4 is the most relevant to my interests, the role of PA in energy metabolism and T2D; as this was done later in my PhD, it also represents my most independent work. Using baseline data from the two prospective DIRECT cohorts described in Paper 2 and 3, I used a SEM to test a network of hypothesised relationships nested within a single multivariate model.

The paper has three aims: first, to test the 'twin cycle' hypothesis, a model that has been set forth to explain the pathogenesis of T2D (for more information see Introduction); second, to determine if the association of PA with glycaemic control is mediated by parameters in this model; and third, to explore if these relationships differ before and after the onset of T2D.

Pairwise associations between physical activity and other parameters in the model.

Initially, I sought to determine pairwise correlations (Figures 9 and 10) between the parameters that are relevant to the twin cycle hypothesis: PA, fasting glucose, 2-hr glucose, oral glucose insulin sensitivity, liver fat, pancreatic fat, fasting insulin secretion rate, glucose sensitivity (beta-cell function estimated as insulin secretion per unit glucose), rate sensitivity (beta-cell function estimated as early insulin secretion enhancement), and potentiation fraction ratio (beta-cell function estimated as late insulin secretion enhancement). See the Materials and Methods section for more details on how these parameters have been measured.

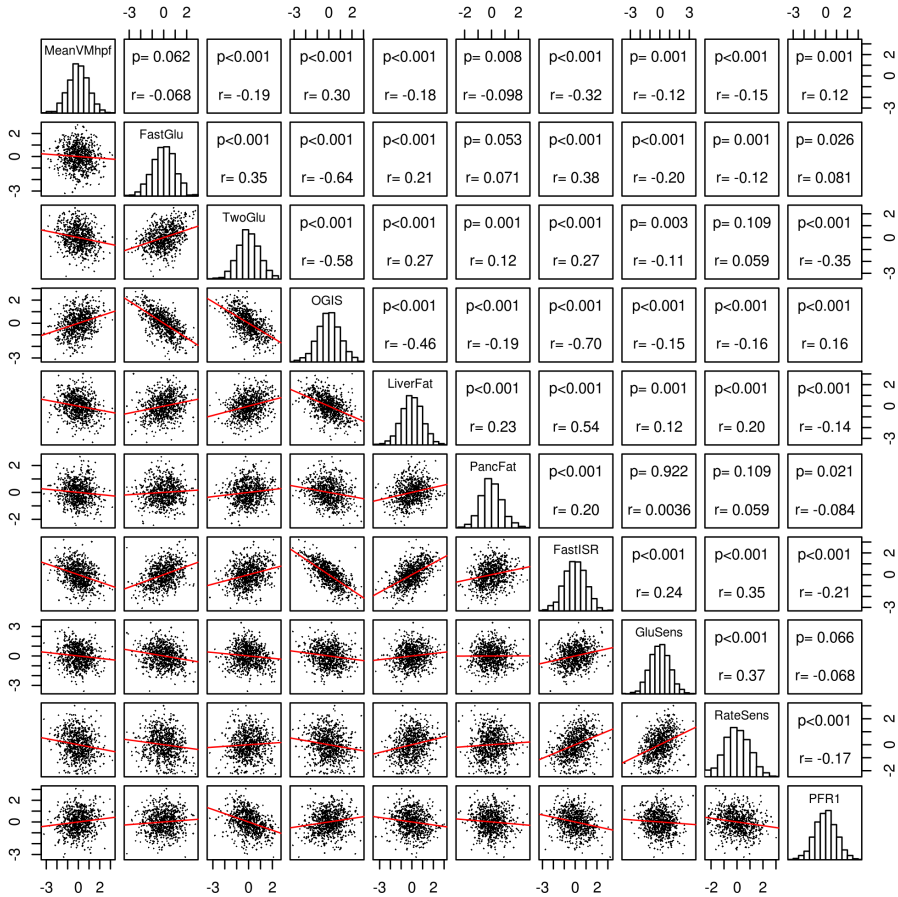


Figure 9
Cohort 1 (Prediabetes): Pairwise Pearson correlations, scatter plots, and histograms. Plot matrix shows pairwise correlations in upper panels, histograms in diagonal panels and pairwise scatterplots in the lower panels. All continuous variables normally transformed and adjusted for age, sex, metformin treatment, study centre, total energy-, carbohydrate-, fat-, and protein-intake. MeanVMhpf: Physical Activity, FastGlu: Fasting Glucose, TwoGlu: 2-hr Glucose, OGIS: Oral Glucose Insulin Sensitivity, LiverFat: Liver Fat, PancFat: Pancreatic Fat, FastISR: Fasting insulin secretion rate, GluSens: Glucose Sensitivity (insulin secretion per glucose), RateSens: Rate Sensitivity (early insulin secretion enhancement), PFR1: Potentiation Fraction Ratio (late insulin secretion enhancement).

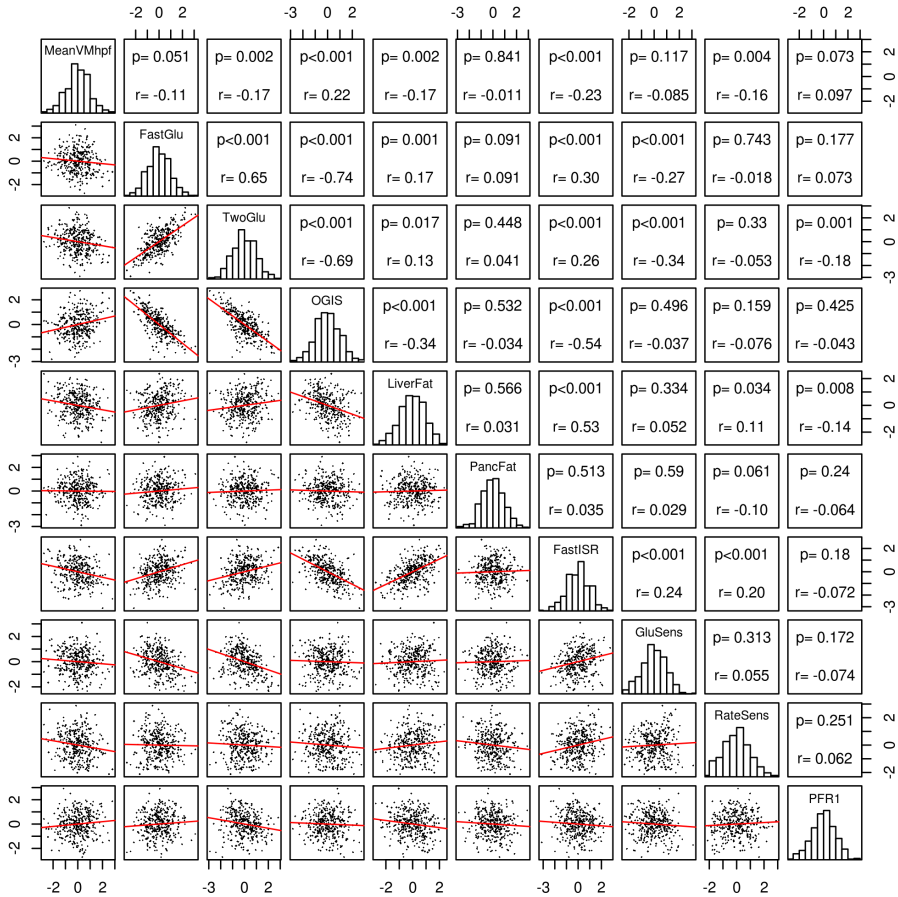


Figure 10
Cohort 2 (T2D): Pairwise Pearson correlations, scatter plots, and histograms. Plot matrix shows pairwise correlations in upper panels, histograms in diagonal panels and pairwise scatterplots in the lower panels. All continuous variables normally transformed and adjusted for age, sex, metformin treatment, study centre, total energy-, carbohydrate-, fat-, and protein-intake. MeanVMhpf: Physical Activity, FastGlu: Fasting Glucose, TwoGlu: 2-hr Glucose, OGIS: Oral Glucose Insulin Sensitivity, LiverFat: Liver Fat, PancFat: Pancreatic Fat, FastISR: Fasting insulin secretion rate, GluSens: Glucose Sensitivity (insulin secretion per glucose), RateSens: Rate Sensitivity (early insulin secretion enhancement), PFR1: Potentiation Fraction Ratio (late insulin secretion enhancement).

Below I describe the key pairwise correlations between PA and all the model parameters, between parameters in the liver and pancreas cycles (see figure 2 in the Introduction), and between insulin sensitivity and beta cell function parameters. Where results are presented in the text, they are as Pearson correlation coefficients (*r*) and *P*-values.

Physical activity (see Figures 9 and 10)

PA was significantly correlated with all liver and pancreatic fat cycle parameters apart from fasting glucose (P =0.062, and P =0.051 in Cohort 1 and Cohort 2, respectively), pancreatic fat in Cohort 2 (P =0.84), and the potentiation fraction ratio (P =0.073).

Liver cycle (see Figures 9 and 10)

The following correlations were observed in both cohorts: liver fat content is correlated with fasting glucose, basal insulin secretion, and 2-hr glucose. Fasting glucose level is correlated with insulin sensitivity and basal insulin secretion, and 2-hr glucose is correlated with insulin sensitivity. Insulin sensitivity is associated with basal insulin secretion.

Pancreas cycle (see Figures 9 and 10)

Pancreatic fat content is correlated with liver fat content, 2-hr glucose levels, and the potentiation fraction ratio in Cohort 1 only. In both cohorts, 2-hr glucose is correlated with glucose sensitivity and the potentiation fraction ratio.

Insulin sensitivity – beta cell function (see Figures 9 and 10)

Insulin sensitivity is correlated with all three beta-cell function parameters (glucose sensitivity, rate sensitivity and potentiation fraction ratio) in Cohort 1 only.

Structural equation model to test the established hypothesis for mediation pathways

Baron and Kenny illustrated how two variables can be associated with an outcome because one of them is a mediator for the other[97]. This may be the case for the associations observed above. In order to test relationships within such a network, I fitted a SEM and tested the effect of direct paths between two variables (known as edges) and pathways through several edges. I describe the results from these analyses below.

Overall model fit (see Figure 11)

Figure 11 illustrates the effect estimates and fit indices of the SEM fitted to Cohort 1 and Cohort 2.

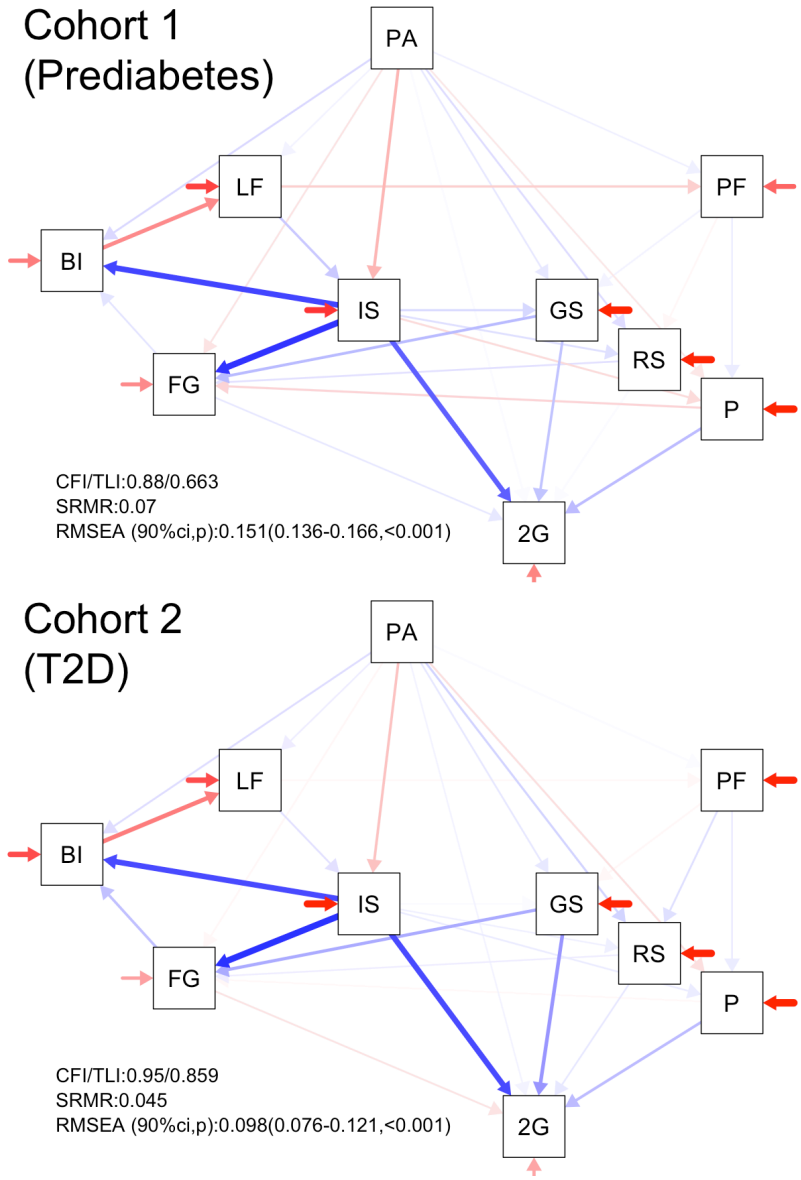


Figure 11
Structural equation model effect estimate diagram from a hypothesised model for the role of physical activity and liver fat in glycaemic control. The diagram illustrates effect size of the defined model applied on Cohort 1 (upper panel) and Cohort 2 (lower panel) where the arrow thickness is weighted by effect estimate magnitude and colours red and blue indicate +ve and –ve estimates respectively. Arrows not originating from a node represent residual variance. All continuous variables normally transformed and adjusted for age, sex, metformin treatment (Cohort 2), study centre, total energy-, carbohydrate-, fat-, and protein-intake. PA: MeanVMhpf (physical activity), FG: Fasting Glucose, 2G: 2-hr Glucose, IS: Oral Glucose Insulin Sensitivity, LF: Liver Fat, PF: Pancreatic Fat, BI: Fasting insulin secretion rate, GS: Glucose Sensitivity (insulin secretion per glucose), RS: Rate Sensitivity (early insulin secretion enhancement), P: Potentiation Fraction Ratio (late insulin secretion enhancement). CFI: Comparative Fit Index, TLI: Tucker-Lewis Index, SRMR: Standardised Root Mean Square Residual.

I report the main fit statistics in the context of established cut-offs for adequacy of fit, as is common practice. However, because there is controversy about the use of fit index cut-offs to decide adequate model fit[89], I do not use them to accept or reject a hypothesis.

The comparative fit index (CFI) for the model is 0.88 and 0.95 in Cohort 1 and 2, respectively. The Tucker-Lewis index (TLI) is 0.66 and 0.86 in Cohort 1 and 2, respectively. CFI and TLI are the main relative fit indexes indicating 'goodness' of fit, they compare fit relative to a specified model or in this case a null model. Their minimum is 0 and maximum 1, with 1 indicating a perfect fit[89]. The accepted cut-off for both indices is >0.95 for a good fit[119,120].

The standardised root mean square residual (SRMR) is 0.07 and 0.045 in Cohort 1 and 2, respectively. The root mean square error of approximation (RMSEA) is 0.151 and 0.098 in Cohort 1 and 2, respectively. SRMR and RMSEA are the main absolute fit indices indicating 'badness' of model fit. The minimum estimate for both is 0 and the maximum 1, with 0 indicating a perfect fit[89]. The accepted cut off for bad fit is >0.08 for SRMR and >0.06 for RMSEA[121,122].

For the current analysis, these indices indicate that the model fit is good in Cohort 2 when assessed by CFI and SRMR. The indices indicate that model fit is good in Cohort 1 only when assessed by SRMR. All indices had values indicating that the models fitted better in Cohort 2 than Cohort 1, but formal comparisons were not done as these estimates are based on different datasets. The apparent better fit in Cohort 2 likely reflects that the twin-cycle model is based on data primarily including T2D participants.

Direct effects (see Figure 11, Table 15)

Nested within the defined model (see figure 6 in materials and methods) are multiple regressions within which direct effects (edges) between variables (nodes) are estimated. These are listed in Table 15.

Table 15
Individual edge effect estimates for the effect of physical activity and abdominal ectopic fat in glycaemic control within a mechanistic model (see Figure 8).

		Cohort 1 (Prediabetes)				Cohort 2 (Diabetes)		
Outcome Node	**Parent Node (edge)**	**β**	**βSE**	**P**		**β**	**βSE**	**P**
OGIS								
	Liver Fat (lf1)	-0.20	0.04	**<0.001**		-0.10	0.07	0.127
	Mean hpfVM (pa1)	0.27	0.03	**<0.001**		0.21	0.05	**<0.001**
Liver Fat								
	Basal Insulin Secretion (bi1)	0.43	0.04	**<0.001**		0.47	0.05	**<0.001**
	Mean hpfVM (pa2)	-0.04	0.03	0.225		-0.06	0.04	0.21
Basal Insulin Secretion								
	Fasting Glucose (fg1)	-0.09	0.03	**0.006**		-0.22	0.06	**0.001**
	OGIS (is1)	-0.68	0.04	**<0.001**		-0.64	0.07	**<0.001**
	Mean hpfVM (pa3)	-0.13	0.03	**<0.001**		-0.11	0.04	**0.01**
Fasting Glucose								
	Mean hpfVM (pa4)	0.10	0.03	**<0.001**		0.03	0.03	0.382
	OGIS (is2)	-0.76	0.03	**<0.001**		-0.74	0.03	**<0.001**
	Glucose Sensitivity (gs1)	-0.25	0.02	**<0.001**		-0.30	0.03	**<0.001**
	Rate Sensitivity (rs1)	-0.10	0.02	**<0.001**		-0.06	0.03	0.068
	Potentiation Fraction Ratio (p1)	0.15	0.02	**<0.001**		0.02	0.03	0.53
Pancreas Fat								
	Liver Fat (lf2)	0.17	0.03	**<0.001**		0.03	0.05	0.567
	Mean hpfVM (pa5)	-0.05	0.03	0.1		-0.01	0.05	0.91
Glucose Sensitivity								
	Pancreas Fat (pf1)	-0.04	0.05	0.36		0.03	0.05	0.599
	Mean hpfVM (pa6)	-0.08	0.04	**0.032**		-0.08	0.05	0.12
	OGIS (is3)	-0.14	0.04	**<0.001**		-0.02	0.05	0.773
Rate Sensitivity								
	Pancreas Fat (pf2)	0.03	0.04	0.489		-0.11	0.05	**0.038**
	Mean hpfVM (pa7)	-0.12	0.04	**0.002**		-0.15	0.05	**0.004**
	OGIS (is4)	-0.12	0.04	**0.002**		-0.05	0.05	0.365
Potentiation Fraction Ratio								
	Pancreas Fat (pf3)	-0.07	0.05	0.148		-0.07	0.05	0.203
	Mean hpfVM (pa8)	0.08	0.04	**0.032**		0.11	0.05	**0.032**
	OGIS (is5)	0.13	0.04	**0.001**		-0.07	0.05	0.177
2-hr Glucose								
	OGIS (is6)	-0.59	0.04	**<0.001**		-0.65	0.05	**<0.001**
	Fasting Glucose (fg2)	-0.08	0.04	**0.035**		0.09	0.05	0.065
	Glucose Sensitivity (gs2)	-0.22	0.03	**<0.001**		-0.37	0.03	**<0.001**
	Rate Sensitivity (rs2)	-0.01	0.03	0.832		-0.07	0.03	**0.015**
	Potentiation Fraction Ratio (p2)	-0.24	0.03	**<0.001**		-0.23	0.03	**<0.001**
	Mean hpfVM (pa9)	-0.01	0.03	0.724		-0.04	0.03	0.254

β units are standard deviations. βSE: Standard error. All continuous variables normally transformed and adjusted for age, sex, metformin treatment (Cohort 2), study centre, total energy-, carbohydrate-, fat-, and protein-intake.

Physical activity direct effects (see Figure 11, Table 15)

PA is associated with insulin sensitivity after adjustment for liver fat in both cohorts. PA is inversely associated with basal insulin secretion after adjustment for fasting glucose and insulin sensitivity in both cohorts. PA is associated with fasting glucose after adjustment for insulin sensitivity, glucose sensitivity, rate sensitivity and potentiation fraction ratio in Cohort 1.

PA is inversely associated with glucose sensitivity and rate sensitivity after adjustment for pancreatic fat and insulin sensitivity in Cohort 1. PA is associated with potentiation fraction ratio after adjustment for insulin sensitivity and pancreatic fat in both cohorts.

There were two key associations postulated by the twin-cycle hypothesis that I do not observed here. First, PA is not associated with liver fat after adjustment for basal insulin secretion in either cohort. However, PA was associated with liver fat in pairwise associations. Second, PA is not associated with 2-hr glucose after adjustment for insulin sensitivity, fasting glucose, glucose sensitivity, rate sensitivity and potentiation fraction ratio in either cohort. PA was associated with 2-hr glucose in pairwise associations. These two differences may be due to indirect pathway effects, i.e. mediation pathways.

Liver cycle direct effects (see Figure 11, Table 15)

Insulin sensitivity is inversely associated with liver fat after adjustment for PA in Cohort 1. Liver fat is associated with basal insulin secretion after adjustment for PA in both cohorts. Basal insulin secretion is inversely associated with fasting glucose after adjustment for PA and insulin sensitivity in both cohorts. Basal insulin secretion is also inversely associated with insulin sensitivity after adjustment for PA and fasting glucose in both cohorts.

Fasting glucose is inversely associated with insulin sensitivity after adjustment for PA, glucose sensitivity, rate sensitivity and potentiation fraction ratio in both cohorts.

2-hr glucose is inversely associated with insulin sensitivity after adjustment for PA, fasting glucose, glucose sensitivity, rate sensitivity, and potentiation fraction ratio in both cohorts.

Pancreas cycle direct effects (see Figure 11, Table 15)

Pancreatic fat is associated with liver fat after adjustment for PA in Cohort 1. Rate sensitivity is inversely associated with pancreatic fat after adjustment for PA and insulin sensitivity in Cohort 2.

Fasting glucose is inversely associated with glucose sensitivity after adjustment for PA, insulin sensitivity, rate sensitivity, and potentiation fraction ratio in both cohorts. Fasting glucose is inversely associated with rate sensitivity after adjustment for PA, insulin sensitivity, glucose sensitivity, and potentiation fraction ratio in Cohort 1. Fasting glucose is associated with potentiation fraction ratio after adjustment for PA, insulin sensitivity, glucose sensitivity, and rate sensitivity in Cohort 1.

2-hr glucose is inversely associated with glucose sensitivity after adjustment for PA, insulin sensitivity, rate sensitivity, and potentiation fraction ratio in both cohorts. 2-hr glucose is inversely associated with rate sensitivity after adjustment for PA, insulin sensitivity, glucose sensitivity, and potentiation fraction ratio in Cohort 2 only. 2-hr glucose is inversely associated with potentiation fraction ratio after adjustment for PA, insulin sensitivity, glucose sensitivity, and rate sensitivity in both cohorts.

Potentiation fraction ratio is not associated with pancreatic fat after adjustment for PA and insulin sensitivity in either cohort. Note that potentiation fraction ratio is, however, associated with pancreatic fat in pairwise associations in Cohort 1. This may reflect the compensatory capacity of beta-cells in the prediabetic state, depending on the degree of insulin sensitivity, which was not considered in the pairwise associations.

Pathways (see Figure 11, Table 16)

Mediation analyses were carried out to estimate indirect effects of PA on fasting glucose, 2-hr glucose, liver fat and insulin sensitivity. This was only carried out on pathways composed of edges where we observe statistically significant direct effects, as indicated in Table 16.

Table 16

Pathway (mediation) effect estimates for the association of physical activity with glycaemic control within a mechanistic model (see Figure 7).

		Cohort 1 (Prediabetes)				Cohort 2 (Diabetes)		
Outcome Node	**Edge Path**	**β**	**βSE**	**P**		**β**	**βSE**	**P**
Fasting Glucose	PA→IS→FG	-0.204	0.026	**<0.001**		-0.158	0.039	**<0.001**
2-hr Glucose	PA→IS→2G	-0.157	0.022	**<0.001**		-0.137	0.035	**<0.001**
Liver Fat	PA→IS→BI→LF	-0.079	0.014	**<0.001**		-0.063	0.019	**0.001**
Liver Fat	PA→BI→LF	-0.057	0.013	**<0.001**		-0.053	0.021	**0.012**
OGIS	PA→BI→LF→IS	0.011	0.003	**0.001**				
Fasting Glucose	PA→BI→LF→IS→FG	-0.008	0.002	**0.001**				
2-hr Glucose	PA→BI→LF→IS→2G	-0.007	0.002	**0.001**				
Fasting Glucose	PA→GS→FG	0.020	0.010	**0.036**				
Fasting Glucose	PA→RS→FG	0.012	0.005	**0.011**				
Fasting Glucose	PA→P→FG	0.012	0.006	**0.043**		0.002	0.004	0.546
2-hr Glucose	PA→GS→2G					0.029	0.019	0.124
2-hr Glucose	PA→RS→2G	0.001	0.003	0.833		0.011	0.006	**0.064**
2-hr Glucose	PA→P→2G	-0.020	0.009	**0.037**		-0.026	0.013	**0.038**

β units are standard deviations. βSE: Standard error. All continuous variables normally transformed and adjusted for age, sex, metformin treatment, study centre, total energy-, carbohydrate-, fat-, and protein-intake.

Physical activity and the liver fat cycle (see Figure 11, Table 16)

PA is inversely associated with liver fat through pathway PA→BI→LF in both cohorts. PA is inversely associated with liver fat through pathway PA→IS→BI→LF in both cohorts. PA is associated with insulin sensitivity through pathway PA→BI→LF→IS in Cohort 1.

Physical activity, the liver fat cycle and fasting glucose (see Figure 11, Table 16)

The associations of PA with estimates of glycaemic regulation are mostly mediated by parameters in the liver fat cycle. PA is inversely associated with fasting glucose through pathway PA→IS→FG in both cohorts. PA is inversely associated with fasting glucose through pathway PA→BI→LF→IS→FG in Cohort 1.

Physical activity, the liver fat cycle and 2-hr glucose (see Figure 11, Table 16)

PA is inversely associated with 2-hr glucose through pathway PA→IS→2G in both cohorts. PA is inversely associated with 2-hr glucose through pathway PA→BI→LF→IS→2G in Cohort 1.

Physical activity, beta-cell function and fasting glucose (see Figure 11, Table 16)

The association of PA with glycaemic regulation is also mediated by beta-cell function parameters in the pancreatic fat cycle. PA is associated with fasting glucose through pathway PA→GS→FG in Cohort 1. PA is associated with fasting glucose through pathway PA→RS→FG in Cohort 1. PA is associated with fasting glucose through pathway PA→P→FG in Cohort 1.

Physical activity, beta-cell function and 2-hr glucose (see Figure 11, Table 16)

PA is associated with 2-hr glucose through pathway PA→GS→2G in Cohort 1 only. PA is inversely associated with 2-hr glucose through pathway PA→P→2G in both cohorts.

Discussion of combined results in the context of current literature

The association between PA and glycaemia being mediated by insulin sensitivity is an established mechanism, as the American College of Sports Medicine (ACSM) indicate in their position statement[112,123-126].

While the relationship we observe between liver fat and glycaemic control has been seen in several studies[115,127,128], there are some noteworthy novel insights provided by our analyses. We observe a strong relationship between PA and insulin sensitivity in both cohorts, in terms of both pairwise and direct effects nested within the SEM. In the pairwise association analyses, we observe strong relationships between liver fat, PA and insulin sensitivity. However, when the edge between liver fat and insulin sensitivity in the SEM was adjusted for PA, the association between liver fat and insulin sensitivity is no longer statistically significant. Similarly, the association between PA and liver fat was not statistically significant, which would be the case if the association between PA and liver fat is mediated by another parameter in the model. The pathway analysis indicates that insulin sensitivity and basal insulin secretion mediate the effects of PA in liver fat. This is in agreement with previous commentaries[49]. In the RAED2 study by

Bacchi et al.[129], 31 middle-aged adults with either non-alcoholic fatty liver disease or T2D were followed-up during two separate four-month exercise interventions of aerobic and resistance exercise. The authors found that the aerobic exercise intervention reduced their liver fat content and increased their insulin sensitivity.

In another investigation based on the RAED2 study, Bacchi et al.[130] report an improvement in oxidative capacity following the aerobic exercise intervention, which may reflect a greater capacity for fat oxidation. This may be a result of aerobic exercise training, which would fit with the notion of the *athlete paradox*[39,131,132] and, as Brouwers et al.[49] suggest, through increased FFA uptake by muscle. A study by Shaw et al. show that an endurance exercise intervention increased intramuscular triglyceride content despite no change in glycaemic control parameters in participants with T2D (i.e. fasting glucose, fasting insulin, 2-hr glucose, HbA1c and HOMA index). In a randomised controlled trial, 100 participants with a mean age of 71 years from the Hertfordshire Cohort Study took part in 36 × 1h cycling sessions over 12 weeks[133]. The researchers found that the intervention arm decreased liver fat, risk for glycaemic dysregulation and increased aerobic capacity compared with the control arm. The study by Shaw et al.[132], the RAED2 study[129], and the RCT using Hertfordshire Cohort Study participants[133] were composed of older participants, similar to the two cohorts in IMI DIRECT. In these studies, exercise was associated with increased insulin sensitivity, as would be expected if the effect was mediated by insulin sensitivity. This is similar to what we observe in the IMI DIRECT cohorts. Unfortunately, neither study tested to see if the magnitude of reduction in liver fat was reduced following adjustment for insulin sensitivity, which could be indicative of mediation.

We find an inverse relationship between PA and glucose sensitivity, as well as rate sensitivity in both the prediabetic and T2D cohorts, which was evident in both the pairwise and the multivariate models. We also observed that PA paths mediated by glucose sensitivity, rate sensitivity and PFR have a small but significant positive association with fasting glucose in Cohort 1. However, we see that PA is inversely associated with 2-hr glucose through PFR in both cohorts. Importantly, the direct effect estimates of PA on the beta-cell function parameters modelled in the SEM are adjusted for insulin sensitivity, indicating that there may be other factors mediating this relationship. There are potential explanations for this in the literature. The relationship between PA and beta-cell function is less commonly reported, but has been shown before. It may be less commonly reported because, counter intuitively, exercise decreases insulin secretion both during exercise[134] and in an exercise-trained state[135-138]. This would result in an inverse association similar to what we observe in our analyses. In the European RISC study[139], 549 adult men and women with three year follow-up data showed similar findings. This study found that the 3-year increase in insulin secretion was of a lesser extent

in participants who spent more time doing moderate or vigorous PA. In the exercising state, lower insulin secretion is not a detrimental attribute, as insulin inhibits hepatic gluconeogenesis and adipose tissue from releasing FFA, both of which are vital sources of energy during exercise. The apparent chronic nature of the PA-insulin secretion association may be due, in part, to repeated exposure to adrenaline, which increases during and after exercise compared with the basal state[140]. As previously discussed, exercise-trained individuals have a greater capacity to use FFA as an energy source, and do not develop insulin resistance despite increasing uptake of FFA into muscle[39,131,132].

In this analysis, we observe relatively little effect of pancreatic fat on beta-cell function parameters (In SEM only rate sensitivity in Cohort 2, and in pair-wise correlations only PFR in Cohort 1), which to some extent contradicts prior reports, including those describing the twin-cycle model of T2D pathogenesis. This might be because it is difficult to accurately estimate pancreatic fat content. In DIRECT, we use a multiecho MRI method to measure both liver fat and pancreatic fat, as described in the Materials and Methods section of this thesis and elsewhere[71,141]. Measuring pancreatic fat is more complex than liver fat and, if not done correctly, can lead to spurious results[114]. In trials carried out by the group who formulated the twin-cycle model[47,48], researchers used a 'three-point-Dixon' method[142]. This method differs from the multiecho method, in that the Dixon method focuses on a signal phase for fat and a signal phase for water within voxels[71]. By contrast, multiecho uses a series of echoes from which oscillations in signal decay give an indication of water and fat content on the whole image rather than a voxel. This means you get an estimate from the whole pancreas rather than defined voxels. The pancreas is not regularly shaped, which may make measurement of ectopic fat difficult to estimate near the boundary of the organ. Macauley et al.[113] report in a study of 41 T2Ds and 14 controls that the pancreas in the T2D group was typically serrated and involuted. An abstract and poster, presented this year at the European Association for the Study of Diabetes[116], quantitatively assessed the irregularity of the pancreas and its relationship with recovery from T2D after a very low calorie diet (in the same population as Steven et al.[48]). It was found that shape irregularity was decreased in participants who recovered their early phase insulin secretion following the very low calorie diet, compared to those who did not recover. If our method has increasing measurement error for pancreatic fat when the pancreas has a more irregular shape, and the irregular shape is related to beta-cell function, it is possible that organ shape irregularity causes confounding in the DIRECT measurements. On the other hand, in the second very low calorie diet trial[48], responders to the intervention arm show a reduction in pancreatic fat and an increase in early phase insulin response during the very low calorie diet intervention phase (0.12±0.04 to 0.26±0.04, $P = 0.03$). This corresponds to the

direct effect estimate of pancreatic fat association with rate sensitivity (our beta cell parameter for early phase insulin secretion enhancement[143]) in Cohort 2.

Limitations

One weakness of this analysis is that it is cross-sectional. This makes an association between any two variables open to reverse causality. However, in the definition of a SEM, we assign directionality to the relationships we test based on a hypothesis built on evidence, which follows a biologically plausible pathway[144], thus minimizing the extent to which reverse causality is likely to occur.

In the absence of longitudinal data (which is not yet available in DIRECT), we compare effects obtained in two populations (DIRECT cohorts 1 and 2) at different stages (before and after the onset of T2D). However, despite the cross-sectional nature of the studies, the high similarity of the protocols used in the two cohorts facilitates a degree of direct comparisons between the disease stages. This is rarely possible in other settings, as protocols usually differ between cohorts, which renders comparisons of the nature undertaken here prone to bias.

Because we did not assess all variables in the original twin-cycle model, it was not possible to determine whether insulin resistance inhibits hepatic gluconeogenesis. It was also not possible to determine if hyperinsulinaemia increased *de-novo* lipogenesis, very low density lipoprotein circulation and pancreatic fat. These differences in the model we defined and the original twin-cycle model may also explain why some of the relationships were not seen.

It is also possible that the methods used here to correct for non-normally distributed data (blanket rank normal transformation of continuous variables) and to control for confounding are overly conservative. These methods make effect estimates of associations between variables easier to compare and more robust, but makes the magnitude difficult to relate to results from other studies. In this study however, the main aim is to establish robust effect estimates for the relationships within the twin-cycle model which favour the methods adopted.

Paper 4 Conclusions

This analysis tests the previously hypothesised twin-cycle model of T2D pathogenesis[30,31] in a free-living epidemiological setting in participants before and after the onset of T2D. Moreover, this analysis integrates the assessment of overall model fit, direct effects and pathway effects of relationships between parameters in the model, by testing them as nested within the same multivariate model.

The results from these analyses support the veracity of the twin-cycle hypothesis as a whole. We observe strong relationships in the liver cycle but only weaker ones in the pancreatic cycle, possibly due to the difficulty in assessing pancreatic fat. We find that the relationship between PA and glucose regulation is mainly mediated by insulin sensitivity, but also to a lesser extent by beta-cell function. The relationship between PA and liver fat is shown to be mediated by whole body insulin sensitivity and basal insulin secretion. Notably, we also observe a relationship between insulin sensitivity and beta-cell function in the prediabetic cohort but not in the T2D cohort, which likely reflects the insulin secretory capacity reserve in the former cohort.

Summary and overall conclusions

My major interest has been the role of PA in glycaemic control, before and after the onset of T2D. During my PhD, I undertook analyses using existing cohort data from Sweden (from the GLACIER Study) and in two newly established European prospective cohorts within the IMI DIRECT consortium; for the DIRECT Study, I have been centrally involved in project coordination, planning and data collection, as well as data QC and processing. I have authored and co-authored 10 papers during my PhD, but the four papers included in my thesis are representative of my primary area of interest and work I have been centrally involved in for the past few years.

In Paper 1, I carried out analysis in an existing population-based prospective study called the GLACIER Study. The cohort consists of around 20,000 participants nested within the ongoing Västerbotten Health Survey in the county of Västerbotten in Sweden. Participants are invited to attend a primary healthcare centre in the year of their 30^{th}, 40^{th}, 50^{th} and 60^{th} birthdays for a thorough health examination. Of the 20,000 GLACIER participants, about 6,000 currently have follow-up data and of these, around 3,500 participants were eligible for the current analysis. The aim was to compare the predictive ability of established genetic risk loci and lifestyle risk factors for the incidence of impaired glycaemic control, T2D and obesity. We found that:

- Lifestyle and genetic factors have comparable predictive ability for the incidence of IFG (ROC AUC 63% and 66%, respectively), IGT (ROC AUC 64% and 61%, respectively), T2D (ROC AUC 75% and 74%, respectively), and obesity (ROC AUC 68% and 72%, respectively).
- Adding information on genetic risk factors to the lifestyle model improved predictive ability for IFG, T2D and obesity (ROC AUC 68%, 80% and

79%, respectively). Genetic factors did not improve the predictive ability for IGT incidence, possibly because only nine SNPs were available for this trait.

- The addition of genetic risk factors to the lifestyle prediction model improved net reclassification by 64% for obesity, 58% for T2D and 36% for IFG. This illustrates that combining information on lifestyle and genetic factors improves the accuracy of models focused on predicting IFG, T2D and obesity.
- No previous studies had shown to what extent genetic factors can add predictive value beyond well-established lifestyle factors. We demonstrate that they can. The reason this has not been more clearly demonstrated is most likely because previous clinical factor models have included intermediate phenotypes, which have encompassed some of the genetic risk.

Papers 2 and 3 describe the output of the data collection aspect of my PhD, through my central involvement in the coordination of two new epidemiological cohorts within the IMI DIRECT consortium; in Paper 2, I overview the design and rationale of the two new glycaemic deterioration prospective cohorts; I describe the baseline characteristics of the cohorts in Paper 3. Below, I summarise a few key points arising from this work:

- The DIRECT Consortium was formed under the banner of the Innovative Medicines Initiative, a joint undertaking between the EU, European academic institutions, and pharmaceutical companies that forms part of the EU's Seventh Framework Programme.
- The aim of DIRECT is to identify novel biomarkers for glycaemic deterioration before and after the onset of T2D. To facilitate this work, we are undertaking two new prospective cohort studies: Cohort 1 'Prediabetes' and Cohort 2 'T2D'.
- Both cohorts are prospective with main visits at baseline, 18 months, 36 months (Cohort 2), and 48 months (Cohort 1). Both cohorts are comprehensively phenotyped with assessments for standard clinical characteristics, glycaemic control and beta-cell function (using fsOGTT/MMTT), regional adiposity (by MRI), self-reported diet intake, PA (by tri-axial accelerometry), and multiple omics, including genomic, transcriptomic, metabolomic, proteomic and faecal microbiome.

Cohort 1 'Prediabetes'.

- Around 2,300 participants were identified using a glycaemic deterioration risk prediction algorithm (DIRECT-DETECT) and recruited from a sampling frame of around 24,000 adults living in or near Kuopio (Finland), Malmö (Sweden), Amsterdam (Netherlands) and Copenhagen (Denmark).
- Using ADA-2011 glycaemic categories for prediabetes, 22% were NGR, 24% had iIA1c, 16% iIFG, 2% iIGT, and 36% cIGR.
- Statistics are presented as mean±SEM; P-values are shown to illustrate differences by glycaemic strata within the cohort. 76% of participants were male, age 62±0.1 years. Age, sex and centre adjusted means were: BMI = 28.7±0.2 kg/m^2 (P<1.6×10^{-21}); fasting glucose was 5.5±0.02 mmol/l (P<2.5×10^{-295}); glucose sensitivity was 112±2 pmol/min/m^2/mM (P=9.8×10^{-15}); oral glucose insulin sensitivity was 381±2 ml/min/m^2 (P=6.6×10^{-165}); and liver fat was 5.6±0.6 percent (P =1.9×10^{-6}).

Cohort 2 'T2D'

- Around 850 participants identified from primary care centres and associated registries were recruited to study centres in Dundee (UK), Exeter (UK), Newcastle (UK), Malmö (Sweden), Amsterdam (Netherlands) and Copenhagen (Denmark).
- 66% of the recruited participants were lifestyle treated (LS) and 34% were metformin + lifestyle treated (Met+LS).
- Statistics are presented as mean±SEM; P-values are shown to illustrate differences by treatment strata within the cohort. Age, sex and centre adjusted means for: BMI was 30.4 ±0.2 kg/m^2 (P=0.4); fasting glucose was 7.1±0.06 mmol/l (P=0.012); glucose sensitivity was 81±3 pmol/min/m^2/mM (P=0.004); oral glucose insulin sensitivity was 300±3 ml/min/m^2 (P =0.09); and liver fat was 8.5±0.3 percent (P =9.8×10^{-4}).

Towards the end of my PhD, I used the newly collected IMI DIRECT cohorts described in Papers 2 and 3 to perform the analysis presented in Paper 4. Here, I tested the hypothesised twin-cycle model for the pathogenesis of T2D, proposed by professor Roy Taylor. I also estimated whether the effect of PA in glycaemic control is likely to operate through factors in this pathway. In this study, I found that:

- The majority of the relationships postulated in the twin-cycle model were replicated in pairwise correlation analyses, with the exception of the associations relating to pancreatic fat in general, and in particular its correlations with parameters of beta-cell function. This was the case in

both the prediabetic and T2D cohort. PA was associated with almost all parameters in the twin-cycle model in both cohorts.

- A twin-cycle model, defined as closely as possible to the original model and using only measured variables, was fitted as an SEM using data from Cohort 1 and Cohort 2. Fit statistics indicated better fit of the model in Cohort 2 than Cohort 1.
- The majority of the main effect associations between variables nested within the SEM (as postulated by the hypothesised twin-cycle relationships) were also replicated, with the exception of edges to and from pancreatic fat. A review of the literature and recent presentations at international conferences suggest this may be due to the irregular shape of the pancreas in diabetes, which has been associated with poor beta-cell function, which in turn might cause measurement error.
- Pathway analyses indicated that the relationship between PA and parameters of glycaemic control are mainly mediated by the association between PA and insulin sensitivity. The analyses also show that the association between PA and liver fat is mediated by insulin sensitivity and basal insulin secretion. We also observed that PA has a small inverse association with glucose sensitivity.
- Most observed associations are directionally consistent and of similar magnitude in both Cohort 1 and Cohort 2. There is one main difference, however: in Cohort 1 (but not in Cohort 2), we observe associations between insulin sensitivity and beta-cell function parameters. This likely reflects the insulin secretion capacity reserve present in people without diabetes, which enables adequate compensatory insulin secretion in the face of insulin resistance, which people with manifest diabetes cannot achieve.

Future Perspective

The challenge

The mechanisms of metabolic dysregulation are composed of a network of factors and their causal effects on each other. Lifestyle risk factors, such as physical inactivity and diet, are considered primordial risk factors and origins of such pathways. They stimulate a cascade of metabolic responses. These responses depend on genetic predisposition, as well as acute and chronic environmental

stimuli. Thus, the factors that regulate metabolism and affect T2D risk are not only multifactorial, but also dynamic, making adequate characterization of these processes extremely challenging. Typically, the data we use to study T2D are selective snapshots of this complex system, and its dynamic nature is rarely captured. Failing to measure everything is not necessarily a problem if all the unmeasured stimuli remained constant, but this is usual not the case in a population study. An overwhelming majority of the genetic diversity of the human race can be attributed to approximately one million genetic variants (a seemingly large number, yet still only about 0.03% of the whole human genome); of these 65, 36 and 9 SNPs have been robustly associated with T2D susceptibility, fasting glucose, and 2-hr glucose levels, respectively. Varying degrees of physical inactivity and healthfulness of our diet also exposes us to differentiating lifestyle risk factors. Furthermore, measurements are not made simultaneously under the same conditions. As these factors affect potential outcomes of interest, the unmeasured and unaccounted for factors add error to our models. This challenge necessitates a more complex and comprehensive approach to address the dynamic system leading to the onset of T2D.

Thankfully, scientific advances are allowing us to overcome these challenges more efficiently all the time. Cheaper, more precise measures are allowing us to make more comprehensive assessments in greater sample sizes. The advances from the GWAS era are a prime example of this. Development of arrays and sequencing technologies mean that increasingly detailed assessments can be made in multiple omic levels such as transcriptome, proteome, epigenome, metabolome and microbiome. Phenotypic measures, too, have advanced with more accurate and cheaper assessments of body composition from MRIs, beta-cell function modelling from frequently sampled oral glucose tolerance tests, and objectively measured physical activity from continuously worn triaxial accelerometry.

Technological advances make it possible to obtain more frequent and higher resolution snapshots of disease biology than ever before. But such data are extremely complex, to an extent where conventional modelling approaches are insufficient. Thus, the new technologies are only valuable if advances in data analysis methods and computing power are commensurate. Indeed, high performance computers are about a thousand times more powerful today than in 2003 when the human genome was first sequenced. Moreover, even the most advanced technologies are only as good as the materials to which they are applied and the conditions under which the materials were obtained. Until recently, the small subfields of systems biology and bioinformatics have emerged as central in modern epidemiology and with this, new analytical tools have become more accessible. The application of interdisciplinary methods such as structural equation modelling, traditionally used in social sciences, and methods for quantifying

pancreatic morphology developed by geographers, is also aiding data processing and analysis.

My strategy for addressing the challenge

I intend to continue on the topic surrounding the role of lifestyle factors, such as physical activity, on metabolic regulation in the pathogenesis of T2D. Building on the findings and methods I have used during my PhD, I will apply a systems biology-oriented approach to existing and new datasets, and continue to build comprehensive datasets for future use.

Background

Based on the published evidence and the results presented in this thesis, there is a mounting body of evidence for the following relationships:

- Peripheral insulin resistance raises insulin levels
- Intramyocellular lipid content increases muscle insulin resistance
- Intrahepatic lipid decreases insulin mediated suppression of gluconeogenesis
- Insulin signals de-novo lipogenesis
- de-novo lipogenesis increases liver fat
- High fat oxidation attenuates intramyocellular lipid content meditaed insulin resistance
- PA increases peripheral insulin sensitivity
- PA reduces liver fat
- PA increases fat oxidation in muscle
- PA reduces insulin secretion

While the evidence for these relationships increase, as to how they relate to each other within a system is poorly understood; unravelling these relationships are crucial to understanding the aetiology of disease. In Paper 4 of this thesis, we begin to unravel these effects in a cross-sectional setting.

Potential Project Outlines

Below are some potential research projects which I intend to pursue in the future:

1. **Do the relationships identified in Paper 4 hold in a longitudinal setting?** For example, if habitual PA increased, does IS increase as the cross-sectional results suggest? As longitudinal data becomes available in DIRECT, it may be possible to test this using similar structural equation modelling methods. In the DIRECT cohorts, the strategy has been to prospectively make comprehensive assessments to form as complete a snapshot of the pathogenic system as possible. Moreover, this has been done in two cohorts where glycaemic deterioration is expected.
2. **Do the directions of the relationships hypothesised in Paper 4 hold in causal inference analyses?** Using genetic instrumental variables in a Mendelian randomisation framework it would be possible to assess the causal link between insulin sensitivity and liver fat. This would potentially require a very large sample size, which may be possible in the UK BioBank public dataset: a very large dataset at around 500,000 participants, also a comprehensively phenotyped dataset with MRI's and objectively measured PA (both in a subset).
3. **Does PA modify the effect of genetic markers indicating a high degree of heterogeneity in the variance between genotypes?** The study and detection of gene-environment interactions can require a very large sample size depending on the frequency and effect size of the genotype in question. However, prioritizing SNPs which might be prone to be genetic components in gene-environment interactions by calculating variance heterogeneity by genotypes could be a feasible strategy to decrease the sample size needed to detect gene-environment interactions. Using this approach to first identify SNPs, and a genotype-based recall trial framwork to identify a sufficient number of participants with and without the risk allele, a gene-environment interaction trial could be performed. The Oxford BioBank would be an ideal candidate for this type of analysis. The Oxford BioBank is a well phenotyped dataset, currently of about 7500 participants with detailed body composition and genetic assessments. Importantly, the Oxford BioBank participants have given consent to be re-invited for further studies based on their genetic results allowing genotype-based-recall studies.

Popular science summary

Like any living organism, we need energy to survive. We get energy primarily from the food we eat. Once digested, the macronutrients (carbohydrates, protein and fats) are used to fuel movement and repair damaged tissues, or stored for future use throughout the body. Carbohydrates are comprised of carbon, hydrogen and oxygen structures in simple or complex forms. The complexity of the carbohydrate influences the rate at which it can be digested and metabolized. Glucose (sugar), for example, has a simple structure, providing energy that can be rapidly metabolised. When carbohydrates are digested, glucose is released and enters the blood stream where it is transported around the body. In healthy people, the concentration of glucose in the blood is tightly regulated (termed *glycaemic regulation*). Unlike other organs, the brain relies almost exclusively on glucose as its fuel. Thus, when blood glucose falls too low (hypoglycaemia), this can cause a person to feel faint and dizzy; however, too much bloody glucose (hyperglycaemia), when sustained for a long period, can lead to tissue damage such as kidney and liver problems, blindness, and foot ulcers (termed *diabetes-related complications*).

Type 2 diabetes (T2D) is a disease where the body loses its ability to regulate blood glucose levels. Although many people with T2D can control their blood glucose levels by increasing their levels of physical activity, improving their diets, and/or taking special medicines, most people with the disease will never fully recover their ability to control blood glucose, and many will go on to develop diabetes-related complications. Before the 1980s it is estimated that less than 1% of the adult population worldwide had T2D, today about 10% of the global population has the disease, and one in five people aged 65 years or older is affected. Today, around 11% of healthcare costs are attributable to dealing with diabetes.

Unhealthy lifestyles are a major cause of T2D. Specifically, T2D is primarily caused by too little physical activity, excessive consumption of energy dense foods with poor nutritional content, genetic predisposition, and adverse early life exposures. Many other potential causes have also been proposed. Whilst the prevalence of T2D has increased dramatically during the past half century, people's genetic characteristics have not changed markedly within this timeframe. This suggests that the genetic predisposition to T2D seen in many populations is

triggered by lifestyle related factors, specifically those that have emerged in recent decades and that are obesogenic.

Around the world, large clinical intervention trials have demonstrated that a healthy lifestyle change in people at high risk of T2D can reduce risk of disease by about 60% compared with routine care. This substantial risk reduction for T2D is achieved by interventions aimed at weight loss through increasing physical activity, reduced caloric intake and improved nutritional content in the diet. The role lifestyle plays in the physiological and molecular changes that occur as T2D develops are still not well understood. But because healthy lifestyle can be so effective at reducing the risk for T2D it's important that we try to understand the mechanisms underlying this, as this may help develop new strategies to prevent or treat T2D more effectively.

Successfully identifying populations at high risk of T2D, and predicting the onset and progression of T2D is important because it might allow more effective targeting of prevention and treatment efforts. In Paper 1, we show in 3,444 GLACIER Study participants that the ability of self-reported lifestyle information compared with genetic data to predict the development of obesity and loss of glycaemic control is similar during 10-years follow-up. We also show that adding the genetic risk factors to a prediction model based on lifestyle risk factors further increases the predictive ability.

Scientists have discovered many risk factors that are associated with worsening of glycaemic regulation. However, it is often difficult to determine whether these risk factors are the cause or consequence of the T2D, confounding our ability to pinpoint the specific factors upon which to intervene and precisely how this should be done.

Within a large European consortium called DIRECT (www.direct-diabetes.org), I have been centrally involved in the coordination of two new prospective cohort studies. In these studies we made comprehensive measurements of risk factors and biological markers related to T2D over a three to four year follow-up period. We did this in more than 2,000 people at high risk of developing T2D, and another 850 who had been recently diagnosed with the disease. Our objective is to find biomarkers that can be used in the clinical setting as tools to better predict onset, progression or response to treatment in T2D. We have placed specific emphasis on measuring physical activity accurately and precisely, using wrist worn accelerometers. During my PhD I spent half a year at the University of Cambridge, where I learned state-of-the-art methods for processing and analysing the physical activity data that I used in Papers 3 and 4 of this thesis. In Paper 2, the rationale and design behind the two DIRECT cohorts are described, and in Paper 3 the physiological characteristics from the first main visit in each cohort is presented. I anticipate that these papers will help guide scientists to better

understand results from other analyses using the DIRECT data, and to design subsequent analyses and studies.

One of the driving factors in the development of T2D is that the cells in the body become resistant to the effects of insulin (the main hormone responsible for moving glucose from the blood into the body's cells, where it can be metabolised). This happens in combination with a gradual decline in the body's ability to produce insulin. Research by Professor Roy Taylor (the opponent at the defence of this thesis) indicates that a greater intake than expenditure of calories from food leads to increased levels of fat stored in the liver. This, in combination with decreased whole body insulin sensitivity leads to insulin insensitivity in the liver too. This is a problem, because the liver acts much like an energy redistribution depot, releasing glucose in the fasted state to ensure that blood glucose levels do not drop too low. The glucose production by the liver is in part regulated by insulin. Insensitivity to insulin in the liver stops it from functioning correctly in the fed state, causing it to continue releasing glucose and fats into the blood when it shouldn't. Importantly, whole body insulin sensitivity is increased and fat in the liver has been shown to be reduced by physical activity. In Paper 4 we use data from the new DIRECT cohorts to investigate the role of physical activity in the model of T2D development proposed by Professor Taylor. Here, we show that physical activity likely affects glycaemic control and liver fat by improving insulin sensitivity.

Populärvetenskaplig sammanfattning

Som allt levande behöver vi energi för att överleva och den får vi främst från maten vi äter. Makronäringsämnena i maten (kolhydrater, protein och fett) bryts ner i mag-tarmkanalen och dess mindre beståndsdelar används runt om i kroppen som bränsle till kroppens rörelser, för att reparera vävnadsskador eller lagras för senare användning. Kolhydrater är ett samlingsnamn för stärkelse, kostfibrer och olika sockerarter och delas upp i enkla och sammansatta, vilket i sin tur påverkar hur snabbt de kan brytas ner av kroppen. Det mesta av kolhydraterna bryts ner till glukos (socker) som har en enkel struktur som snabbt kan brytas ner och ge kroppens celler energi. När kolhydrater bryts ner frisätts glukos till blodet där det transporteras runt i kroppen. I friska människor är mängden glukos i blodet starkt reglerat (kallas glykemisk reglering). En anledning är att till skillnad från andra organ i kroppen är hjärnan nästan helt beroende av glukos som energikälla. Om blodsockernivån är för låg (hypoglykemi) kan man känna yrsel, darrighet och bli lättretlig; vid höga blodsockernivåer (hyperglykemi) under en längre tid uppstår vävnadsskador som kan leda till lever- och njursvikt, blindhet och fotsår (dessa kallas gemensamt diabetesrelaterade komplikationer).

Typ 2-diabetes (T2D) är en sjukdom där kroppen förlorar möjligheten att reglera nivån av glukos i blodet. Även om många människor som utvecklar T2D kan kontrollera sin blodsockernivå genom ökad fysisk aktivitet, förbättrad kosthållning och/eller med hjälp av läkemedel kommer de flesta aldrig att återfå en fungerande blodsockerreglering och utvecklar med tiden diabetesrelaterade komplikationer. Man har uppskattat att före 1980 så hade mindre än 1% av världens vuxna T2D, idag är siffran runt 10% för vuxna generellt och en av fem över 65 års ålder är drabbad. Idag beräknas 11% av sjukvårdskostnaderna gå till diabetesrelaterad vård.

Ohälsosam livsstil är en av huvudorsakerna till T2D. Mer specifikt är T2D resultatet av för lite fysisk aktivitet, överkonsumtion av energirik men näringsfattig mat, ärftlig riskbenägenhet och ogynnsamma förhållanden under foster- och spädbarnstiden. Många andra potentiella orsaker har också föreslagits. Medan förekomsten av T2D har ökat dramatiskt under de senaste 50 åren har människors genetiska uppsättning under samma tidsperiod inte genomgått några dramatiska förändringar. Detta tyder på att den genetiska riskbenägenhet för T2D som ses i flera populationer har triggats av livsstilsrelaterade faktorer, speciellt de

som blivit mer vanliga de senaste årtiondena och som är obesogena (kan framkalla fetma).

Stora kliniska studier världen över har demonstrerat att människor i riskgruppen för T2D som genomför hälsosamma livsstilsförändringar kan minska risken för sjukdom med ca 60% jämfört med rutinmässig vård. De positiva resultaten uppnås med interventioner som har som målsättning att minska vikten genom ökad fysisk aktivitet och reducerat energiintag med förbättrad kosthållning. Hur livsstilsfaktorer påverkar de fysiologiska och molekylära förändringar som sker under utvecklingen av T2D är fortfarande inte helt kända. Men eftersom en hälsosam livsstil så effektivt kan minska risken för sjukdom är det viktigt att vi försöker förstå de underliggande mekanismerna då det kan hjälpa oss att utveckla nya strategier för att mer effektivt förebygga och behandla T2D.

Att framgångsrikt identifiera grupper med hög risk att utveckla T2D och kunna förutsäga sjukdom och sjukdomsutveckling är viktigt eftersom det kanske skulle möjliggöra att förebyggande åtgärder och behandling mer effektivt kunde riktas mot de som bäst behöver dem. I det första arbetet i denna avhandling visar vi att livsstilsfaktorer mätt med enkät och information kring genetisk riskbenägenhet var för sig har likvärdig förmåga att förutsäga försämrad blodsockerkontroll och fetma efter 10-års uppföljning hos 3,444 deltagare in GLACIER-studien. Vi visar också att möjligheten att kunna förutsäga försämrad blodsockerkontroll och fetma ökar om både genetisk och livsstilsrelaterad information tillsammans tas i beaktande.

Forskare har upptäckt många riskfaktorer kopplade till försämrad glykemisk reglering. Men det är ofta svårt att avgöra om dessa riskfaktorer är orsak till eller en konsekvens av T2D, vilket gör det svårt att avgöra vilka specifika riskfaktorer som vi borde förebygga och hur det skulle gå till.

Inom ett stort europeiskt konsortium som heter DIRECT (www.direct-diabetes.org) har jag varit centralt involverad i att samordna två nya prospektiva studier. I dessa två studier har vi under tre till fyra års tid samlat in omfattande data kring riskfaktorer och biologiska markörer relaterade till T2D hos studiedeltagarna. Detta genomfördes i mer än 2000 människor med hög risk men som ännu inte utvecklat T2D, och ytterligare 850 människor som nyligen blivit diagnostiserad med sjukdomen. Målsättningen med DIRECT är att upptäcka biomarkörer som redan i ett tidigt stadium kan visa på en framtida försämring av blodsockerkontroll respektive sjukdomsutveckling och som kan användas inom sjukvården för att bättra kunna förebygga och behandla sjukdomen. Inom DIRECT har vi speciellt fokuserat på att försöka mäta fysisk aktivitet på ett korrekt och objektivt sätt genom att använda accelerometrar (monitorer) som bärs på handleden. Under min doktorandtid har jag tillbringat sex månader vid universitetet i Cambridge där jag lärt mig spjutspetsmetoder för att bearbetat, förädlat och analysera de fysisk aktivitetsdata som jag använt i arbete 3 och 4 i

min avhandling. I arbete 2 beskrivs bakgrunden till och designen av de två kohortstudierna inom DIRECT och i artikel 3 presenteras studiedeltagarnas fysiologiska och kliniska data från det första besöket. Min målsättning är att dessa artiklar kommer att hjälpa till att orientera forskare så att de lättare förstår forskningsresultat som baseras på DIRECT-data och att designa uppföljande analyser och studier.

En av huvudorsakerna till att T2D utvecklas är att kroppens celler förlorar sin känslighet för insulins effekt (insulin är ett av de huvudsakliga hormon som ser till att glukos transporteras från blodet in i cellen där det kan brytas ner och tillföra cellen energi). Detta sker samtidigt som kroppen gradvis förlorar förmågan att producera insulin. Forskning av professor Roy Taylor (opponent vid disputationen av denna avhandling) visar att ett större intag än förbrukning av kalorier (energi) från mat leder till att en ökad mängd fett lagras in i levern. Detta, i kombination med en minskad insulinkänslighet i kroppen, leder till insulinkänslighet också i levern. Detta är ett problem eftersom levern fungerar som en central för omfördelning av energi och utsöndrar glukos till blodet under fasta för att säkerställa att blodsockernivån inte blir för låg. Glukosproduktionen i levern är delvis reglerad av insulin. En okänslighet för insulin i levern gör att den inte längre fungerar normal efter måltid och leder till att den fortsätter att frisätta glukos och fetter till blodet när den inte borde. Fysisk aktivitet ökar kroppens insulinkänslighet och minskar fettinlagring i levern. I arbete 4 använder vi data från DIRECT-studierna för att undersöka vilken roll fysisk aktivitet spelar i den modell över sjukdomsutveckling som professor Taylor lagt fram. Vi visar att fysisk aktivitet troligtvis påverkar glykemisk kontroll och fettinlagring i levern genom att förbättra insulinkänsligheten.

Acknowledgements

I thank **the participants in the GLACIER and DIRECT studies** for their contributions to these studies and my thesis.

I am grateful for the remarkable mentorship I have had from **Paul Franks**. During my PhD I have had a broad range of incredible opportunities to learn new skills, gain experience and be exposed to tremendously inspiring scientific environments. Throughout my PhD, Paul found a great balance of being mentor, boss and friend.

I am thankful to **Frida Renström**, for her inspiring capacity to keep work a fun and exciting place no matter what. *Smile and the world smiles with you*, the veracity of this statement is undeniable when you have worked with Frida.

Tibor V. Varga and I started our doctoral studies on the same day. We soon developed a *bromantic* camaraderie as we shared countless laughs, sweat, a fair few beers, and overcame many challenges together. I am grateful for all the memories and his support through my PhD.

Past and present members of **The Genetic and Molecular Epidemiology (GAME) unit**: **Paul, Frida, Azra, Nick, Alaitz, Maria, Shafqat, Tibor, Angela, Naeimeh, Yan, Simon, Pernilla, Jerneja, Britta, Adem, Ashfaq, Dmitry, Sonja, Johanna, Jeremy, Åsa, Louise and Bayo.** The words to describe how unbelievably lucky and honoured I feel to have done my PhD with these guys and receive their support are far beyond my literary capabilities.

The Sauna - 'Bastu klubben med Koivula': Karl Bacos, Tibor V. Varga, Dmitry Shungin and Ashfaq Ali (and dad on occasion). It may seem like an odd thing to acknowledge, but every single part of my PhD has been pondered and discussed at length in the sauna. This thesis was 'spikad'/announced in the CRC sauna.

I thank **Anders Rosengren** for always finding time to give advice and a fresh perspective on any of my research. I am thankful to **Inger Lund Pedersen** and **Regine Bergholdt** at Novo Nordisk for their generosity and support through my studies.

The IMI DIRECT Consortium has been ubiquitous with my PhD. Over the last half-decade I have had the extraordinary privilege to have 'direct' access to leading scientists in the field of diabetes all over Europe. This has without a doubt

been the most rewarding experience in my PhD and I am thankful to every nurse, analyst, assistant, technician, investigator and principal investigator for making it so. In particular I am grateful for the advice and support from: **Ewan Pearson, Hartmut Ruetten, Ian Forgie, Imre Pavo, Veikko Koivisto, Anubha Mahajan, Mark McCarthy, Inger Lund Pedersen, Regine Bergholdt, Bernd Jablonka, Soren Brunak, Ramneek Gupta, Jurek Adamski, Jayne Kaye, Harriet Teare, Manolis Dermitzakis, Mark Walker, Markku Laakso, Oluf Pedersen, Torben Hansen, Andrew Hattersley, Jacqueline Dekker, Tim McDonald, Mandy Perry, Andrea Mari, Gary Frost, Jimmy Bell, Kristoffer Rapacki, Hans-Henrik Stærfeldt, Peter Sackett and Fred DeMasi.**

A big thanks to **Maria Klintenberg, Margit Bergström, Ann-Marie Räihä, Anna Pålsson, Sophie Örjansdotter, Ylva Wessman, Ingrid Wetterholtz, Magnus Löndahl and Martin Ridderstråle** for their hard work in the Lund University study site for DIRECT.

I thank **Soren Brage, Tom White, Nick Wareham and the MRC Epidemiology unit** at the University of Cambridge for a fantastically inspiring research visit.

I am very grateful for the support I have received from **Ulrika Blom-Nilsson, Jacqueline Postma, Johan Lind, and Viktoria Petri** during my PhD.

Past and present DPLU PhD steering committee: **Anna, Cajsa, Caroline, Christina, Eliana, Elin, Hanna, Hedvig, Jennifer, Johanna, Linda, Lovisa, Mia, Nina and Yan**. Looking back at the events organised by this group, I am very grateful to having been part of it.

I am grateful for the all too rare, but all-important work-hard-play-hard support of my friends **Cash, Emily, Bex, Lol, Nick, Beck, John, Gran, Amy, Teri, Gary, Kristina, Johan, Neal, Malin, Aiden, Jess and Jacob**.

I am thankful to my family: **Linda, Jonas, Laleh, Sarah, Anthony, Max, Moritz, Nathalie, Kevin**, **'Fammo' Ebba Koivula, Duncan, Laura, Valerie, Jean and Fred** for their loving encouragement and not giving up on questioning me despite risking a science lecture.

I thank my parents **Bettina and Ulf**, for their unconditional love, boundless generosity and resolute belief that I can achieve anything I set my mind to. Thanks to their support I have had an international education and upbringing giving me the confidence to chase my dreams wherever they may take me.

I have known **Fi** since we both enrolled at the University of Glasgow for the same undergraduate degree. Through every *practical*, *dissertation*, *exam*, *thesis* and *viva*, I have progressed in my scientific career with Fi by my side, supporting me, loving me, and encouraging me to do better. For this I am forever grateful. Oh, and "Fi is just a little bit better than Robert at everything".

Funding

My PhD position was jointly funded by: i) IMI DIRECT, an Innovative Medicines Initiative Joint Undertaking under grant agreement n°115317, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution; ii) a NovoNordisk STAR PhD fellowship; iii) grant support from the Swedish Research Council.

I received funding for a half-year research visit to the University of Cambridge from BLUEScY (EXODIAB and EpiHealth), Swedish Society for Medical Research (SSMF), Lund University Medical Faculty, Swedish Heart Lung Foundation (HLF), Regementsläkaren dr Hartelii stipendiestiftelse, and a Diabetes Programme at Lund University (DPLU) PhD Student Prize 2015.

I have received additional travel funding from the European Association for the Study of Diabetes (EASD), HLF, and Lund University Medical Faculty.

References

1. Frerichs R. John Snow. Encyclopaedia Britannica. Online; 2009.
2. Snow J. On the mode of communication of Cholera. 2nd ed. London: John Churchill; 1854.
3. Reaven GM. Role of Insulin Resistance in Human Disease. *Diabetes* 1988; **37**: 1595-607.
4. IDF. Treatment Algorithm for People with Type 2 Diabetes. 2016. http://www.idf.org/treatment-algorithm-people-type-2-diabetes (accessed 10/10/2016 2016).
5. Ringborg A, Lindgren P, Yin DD, Martinell M, Stalhammar J. Time to insulin treatment and factors associated with insulin prescription in Swedish patients with type 2 diabetes. *Diabetes & metabolism* 2010; **36**(3): 198-203.
6. Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiology Biomarkers & Prevention* 2005; **14**(8): 1847-50.
7. Franks PW, Pare G. Putting the Genome in Context: Gene-Environment Interactions in Type 2 Diabetes. *Current diabetes reports* 2016; **16**(7): 57.
8. Franks PW, Pearson E, Florez JC. Gene-environment and gene-treatment interactions in type 2 diabetes: progress, pitfalls, and prospects. *Diabetes Care* 2013; **36**(5): 1413-21.
9. International-Diabetes-Federation. IDF Diabetes Atlas. 7th ed. Brussels, Belgium; 2015.
10. Rates of Diagnosed Diabetes per 100 Civilian, Non-Institutionalized Population, by Age, United States, 1980–2014. December 1 2015. http://www.cdc.gov/diabetes/statistics/prev/national/figbyage.htm (accessed September 27 2016).
11. NCD-Risk-Factor-Collaboration. Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants. *Lancet* 2016; **387**(10026): 1377-96.
12. Jeffery RW, French SA. Epidemic obesity in the United States: Are fast foods and television viewing contributing? *American journal of public health* 1998; **88**(2): 277-80.
13. Tucker LA. Objectively measured physical activity predicts subsequent energy intake in 300 women. *Public health nutrition* 2016: 1-9.

14. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet* 2012; **380**(9838): 219-29.
15. Ding D, Lawson KD, Kolbe-Alexander TL, et al. The economic burden of physical inactivity: a global analysis of major non-communicable diseases. *The Lancet* 2016.
16. Willemsen G, Ward KJ, Bell CG, et al. The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium. *Twin research and human genetics : the official journal of the International Society for Twin Studies* 2015; **18**(6): 762-71.
17. Munoz M, Pong-Wong R, Canela-Xandri O, Rawlik K, Haley CS, Tenesa A. Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nature genetics* 2016; **48**(9): 980-3.
18. Almgren P, Lehtovirta M, Isomaa B, et al. Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia* 2011; **54**(11): 2811-9.
19. Voight BF, Scott LJ, Steinthorsdottir V, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature genetics* 2010; **42**(7): 579-89.
20. Morris AP, Voight BF, Teslovich TM, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics* 2012; **44**(9): 981-90.
21. Scott RA, Lagou V, Welch RP, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature genetics* 2012; **44**(9): 991-1005.
22. Falconer DS. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics* 1965; **29**(1): 51-76.
23. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009; **461**(7265): 747-53.
24. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. *Nature* 2016; **536**(7614): 41-7.
25. Eaton SB, Konner M. Paleolithic nutrition. A consideration of its nature and current implications. *The New England journal of medicine* 1985; **312**(5): 283-9.
26. Booth FW, Gordon SE, Carlson CJ, Hamilton MT. Waging war on modern chronic diseases: primary prevention through exercise biology. *J Appl Physiol (1985)* 2000; **88**(2): 774-87.
27. Hu FB, Manson JE, Stampfer MJ, et al. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *The New England journal of medicine* 2001; **345**(11): 790-7.
28. Helmrich SP, Ragland DR, Leung RW, Paffenbarger RS, Jr. Physical activity and reduced occurrence of non-insulin-dependent diabetes mellitus. *The New England journal of medicine* 1991; **325**(3): 147-52.
29. Lee DC, Park I, Jun TW, et al. Physical activity and body mass index and their associations with the development of type 2 diabetes in korean men. *Am J Epidemiol* 2012; **176**(1): 43-51.

30. Taylor R. Pathogenesis of type 2 diabetes: tracing the reverse route from cure to cause. *Diabetologia* 2008; **51**(10): 1781-9.
31. Taylor R. Type 2 diabetes: etiology and reversibility. *Diabetes Care* 2013; **36**(4): 1047-55.
32. Marinou K, Hodson L, Vasan SK, et al. Structural and functional properties of deep abdominal subcutaneous adipose tissue explain its association with insulin resistance and cardiovascular risk in men. *Diabetes Care* 2014; **37**(3): 821-9.
33. Pinnick KE, Nicholson G, Manolopoulos KN, et al. Distinct developmental profile of lower-body adipose tissue defines resistance against obesity-associated metabolic complications. *Diabetes* 2014; **63**(11): 3785-97.
34. Philipsen A, Hansen AL, Jorgensen ME, et al. Associations of Objectively Measured Physical Activity and Abdominal Fat Distribution. *Medicine and science in sports and exercise* 2014.
35. Demerath EW, Reed D, Rogers N, et al. Visceral adiposity and its anatomical distribution as predictors of the metabolic syndrome and cardiometabolic risk factor levels. *The American journal of clinical nutrition* 2008; **88**(5): 1263-71.
36. Eriksson KF, Lindgarde F. Prevention of type 2 (non-insulin-dependent) diabetes mellitus by diet and physical exercise. The 6-year Malmo feasibility study. *Diabetologia* 1991; **34**(12): 891-8.
37. Tuomilehto J, Lindstrom J, Eriksson JG, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *The New England journal of medicine* 2001; **344**(18): 1343-50.
38. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *The New England journal of medicine* 2002; **346**(6): 393-403.
39. Goodpaster BH, He J, Watkins S, Kelley DE. Skeletal muscle lipid content and insulin resistance: evidence for a paradox in endurance-trained athletes. *The Journal of clinical endocrinology and metabolism* 2001; **86**(12): 5755-61.
40. Kelley DE, Goodpaster B, Wing RR, Simoneau J-A. Skeletal muscle fatty acid metabolism in association with insulin resistance, obesity, and weight loss. *American journal of physiology Endocrinology and metabolism* 1999; **277**(6): E1130-41.
41. Thomas EL, Brynes AE, McCarthy J, et al. Preferential loss of visceral fat following aerobic exercise, measured by magnetic resonance imaging. *Lipids* 2000; **35**(7): 769-76.
42. Trenell MI, Hollingsworth KG, Lim EL, Taylor R. Increased daily walking improves lipid oxidation without changes in mitochondrial function in type 2 diabetes. *Diabetes Care* 2008; **31**(8): 1644-9.
43. Herzig KH, Ahola R, Leppaluoto J, Jokelainen J, Jamsa T, Keinanen-Kiukaanniemi S. Light physical activity determined by a motion sensor decreases insulin resistance, improves lipid homeostasis and reduces visceral fat in high-risk subjects: PreDiabEx study RCT. *International journal of obesity (2005)* 2014; **38**(8): 1089-96.
44. Slentz CA, Bateman LA, Willis LH, et al. Effects of exercise training alone vs a combined exercise and nutritional lifestyle intervention on glucose homeostasis in

prediabetic individuals: a randomised controlled trial. *Diabetologia* 2016; **59**(10): 2088-98.

45. Hawley JA, Gibala MJ. What's new since Hippocrates? Preventing type 2 diabetes by physical exercise and diet. *Diabetologia* 2012; **55**(3): 535-9.
46. Whyte LJ, Ferguson C, Wilson J, Scott RA, Gill JM. Effects of single bout of very high-intensity exercise on metabolic health biomarkers in overweight/obese sedentary men. *Metabolism: clinical and experimental* 2013; **62**(2): 212-9.
47. Lim EL, Hollingsworth KG, Aribisala BS, Chen MJ, Mathers JC, Taylor R. Reversal of type 2 diabetes: normalisation of beta cell function in association with decreased pancreas and liver triacylglycerol. *Diabetologia* 2011; **54**(10): 2506-14.
48. Steven S, Hollingsworth KG, Al-Mrabeh A, et al. Very-Low-Calorie Diet and 6 Months of Weight Stability in Type 2 Diabetes: Pathophysiologic Changes in Responders and Nonresponders. *Diabetes Care* 2016.
49. Brouwers B, Hesselink MKC, Schrauwen P, Schrauwen-Hinderling VB. Effects of exercise training on intrahepatic lipid content in humans. *Diabetologia* 2016; **59**(10): 2068-79.
50. Kurbasic A, Poveda A, Chen Y, et al. Gene-Lifestyle Interactions in Complex Diseases: Design and Description of the GLACIER and VIKING Studies. *Current nutrition reports* 2014; **3**(4): 400-11.
51. Norberg M, Wall S, Boman K, Weinehall L. The Vasterbotten Intervention Programme: background, design and implications. *Global health action* 2010; **3**.
52. Alssema M, Vistisen D, Heymans MW, et al. The Evaluation of Screening and Early Detection Strategies for Type 2 Diabetes and Impaired Glucose Tolerance (DETECT-2) update of the Finnish diabetes risk score for prediction of incident type 2 diabetes. *Diabetologia* 2011; **54**(5): 1004-12.
53. Mari A, Tura A, Gastaldelli A, Ferrannini E. Assessing insulin secretion by modeling in multiple-meal tests: role of potentiation. *Diabetes* 2002; **51 Suppl 1**: S221-6.
54. Mari A, Ferrannini E. Beta-cell function assessment from modelling of oral tests: an effective approach. *Diabetes, obesity & metabolism* 2008; **10 Suppl 4**: 77-87.
55. Hallal PC, Victoria CG. Reliability and validity of the international physical activity questionnaire (IPAQ). *Medicine and science in sports and exercise* 2004; **36**(3): 556-.
56. Craig C, Marshall A, Sjöström M, et al. and the IPAQ Consensus Group and the IPAQ Reliability and Validity Study Group. International Physical Activity Questionnaire (IPAQ): 12-country reliability and validity. *Medicine and science in sports and exercise* 2003; **35**(13): 81-95.
57. van Hees VT, Fang Z, Langford J, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *J Appl Physiol (1985)* 2014; **117**(7): 738-44.
58. Brage S, Westgate K, Wijndaele K, Godinho J, Griffin S, Wareham N. Evaluation of a method for minimising diurnal information bias in objective sensor data. *Int Conf Amb Mon Phys Act Mov* 2013.

59. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. *British journal of sports medicine* 2014; **48**(13): 1019-23.
60. Johansson I, Hallmans G, Wikman A, Biessy C, Riboli E, Kaaks R. Validation and calibration of food-frequency questionnaire measurements in the Northern Sweden Health and Disease cohort. *Public health nutrition* 2002; **5**(3): 487-96.
61. Johansson I, Van Guelpen B, Hultdin J, Johansson M, Hallmans G, Stattin P. Validity of food frequency questionnaire estimated intakes of folate and other B vitamins in a region without folic acid fortification. *European journal of clinical nutrition* 2010; **64**(8): 905-13.
62. Wennberg M, Vessby B, Johansson I. Evaluation of relative intake of fatty acids according to the Northern Sweden FFQ with fatty acid levels in erythrocyte membranes as biomarkers. *Public health nutrition* 2009; **12**(9): 1477-84.
63. Nordic nutrition recommendations 2012. Copenhagen, 2012.
64. Nettleton JA, Hivert MF, Lemaitre RN, et al. Meta-analysis investigating associations between healthy diet and fasting glucose and insulin levels and modification by loci associated with glucose homeostasis in data from 15 cohorts. *Am J Epidemiol* 2013; **177**(2): 103-15.
65. Moshfegh AJ, Rhodes DG, Baer DJ, et al. The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. *The American journal of clinical nutrition* 2008; **88**(2): 324-32.
66. Subar AF, Kipnis V, Troiano RP, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. *Am J Epidemiol* 2003; **158**(1): 1-13.
67. Neuhouser ML, Tinker L, Shaw PA, et al. Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative. *Am J Epidemiol* 2008; **167**(10): 1247-59.
68. Wood DA, Kotseva K, Connolly S, et al. Nurse-coordinated multidisciplinary, family-based cardiovascular disease prevention programme (EUROACTION) for patients with coronary heart disease and asymptomatic individuals at high risk of cardiovascular disease: a paired, cluster-randomised controlled trial. *Lancet* 2008; **371**(9629): 1999-2012.
69. Black AE. Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity* 2000; **24**(9): 1119-30.
70. Black AE. The sensitivity and specificity of the Goldberg cut-off for EI:BMR for identifying diet reports of poor validity. *European journal of clinical nutrition* 2000; **54**(5): 395-404.
71. Thomas EL, Fitzpatrick JA, Malik SJ, Taylor-Robinson SD, Bell JD. Whole body fat: content and distribution. *Progress in nuclear magnetic resonance spectroscopy* 2013; **73**: 56-80.

72. Franks PW, Rolandsson O, Debenham SL, et al. Replication of the association between variants in WFS1 and risk of type 2 diabetes in European populations. *Diabetologia* 2008; **51**(3): 458-63.
73. Renstrom F, Payne F, Nordstrom A, et al. Replication and extension of genome-wide association study results for obesity in 4923 adults from northern Sweden. *Human molecular genetics* 2009; **18**(8): 1489-96.
74. Voight BF, Kang HM, Ding J, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS genetics* 2012; **8**(8): e1002793.
75. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 2003; **26 Suppl 1**: S5-20.
76. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518**(7538): 197-206.
77. Fontaine-Bisson B, Renstrom F, Rolandsson O, et al. Evaluating the discriminative power of multi-trait genetic risk scores for type 2 diabetes in a northern Swedish population. *Diabetologia* 2010; **53**(10): 2155-62.
78. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44**(3): 837-45.
79. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in medicine* 2011; **30**(1): 11-21.
80. Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley; 2000.
81. Siddiqi N. Credit risk scorecards: developing and implementing intelligent credit scoring. New Jersey: Wiley; 2006.
82. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 2007; **81**(3): 559-75.
83. R-Core-Team. R: A language and environment for statistical computing. . Vienna, Austria: R Foundation for Statistical Computing; 2014.
84. The SAS system for Windows. Cary, NC, USA: SAS Institute; 2011.
85. American-Diabetes-Association. Standards of medical care in diabetes--2011. *Diabetes Care* 2011; **34 Suppl 1**: S11-61.
86. Sobel ME. Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology* 1982; **13**: 290-312.
87. Hooper D, Coughlan J, Mullen M. Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods* 2008; **6**(1): 53-60.
88. Kayan Fadlelmula K. Assessing Power of Structural Equation Modeling Studies: A Meta-Analysis. *Education Research Journal* 2011; **1**(3): 37-42.
89. West SG, Taylor AB, Wu W. Model Fit and Model Selection in Structural Equation Modeling. New York, USA: The Guilford Press; 2012.

90. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software* 2012; **48**(2).
91. Epskamp S. Package 'semPlot'. 2015.
92. Lyssenko V, Laakso M. Genetic screening for the risk of type 2 diabetes: worthless or valuable? *Diabetes Care* 2013; **36 Suppl 2**: S120-6.
93. Sparsø T, Grarup N, Andreasen C, et al. Combined analysis of 19 common validated type 2 diabetes susceptibility gene variants shows moderate discriminative value and no evidence of gene–gene interaction. *Diabetologia* 2009; **52**(7): 1308-14.
94. Janipalli C, Kumar M, Vinay D, et al. Analysis of 32 common susceptibility genetic variants and their combined effect in predicting risk of Type 2 diabetes and related traits in Indians. *Diabetic medicine* 2012; **29**(1): 121-7.
95. Meigs JB, Shrader P, Sullivan LM, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *The New England journal of medicine* 2008; **359**(21): 2208-19.
96. de Miguel-Yanes JM, Shrader P, Pencina MJ, et al. Genetic risk reclassification for type 2 diabetes by age below or above 50 years using 40 type 2 diabetes risk single nucleotide polymorphisms. *Diabetes care* 2011; **34**(1): 121-5.
97. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986; **51**(6): 1173-82.
98. Aschard H, Chen J, Cornelis MC, Chibnik LB, Karlson EW, Kraft P. Inclusion of gene-gene and gene-environment interactions unlikely to dramatically improve risk prediction for complex diseases. *American journal of human genetics* 2012; **90**(6): 962-72.
99. Lyssenko V, Jonsson A, Almgren P, et al. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *The New England journal of medicine* 2008; **359**(21): 2220-32.
100. 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 (Clinical research ed)* 2010; **340**: b4838.
101. Talmud PJ, Cooper JA, Morris RW, et al. Sixty-five common genetic variants and prediction of type 2 diabetes. *Diabetes* 2015; **64**(5): 1830-40.
102. Stancakova A, Javorsky M, Kuulasmaa T, Haffner SM, Kuusisto J, Laakso M. Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. *Diabetes* 2009; **58**(5): 1212-21.
103. Hills SA, Balkau B, Coppack SW, et al. The EGIR-RISC STUDY (The European group for the study of insulin resistance: relationship between insulin sensitivity and cardiovascular disease risk): I. Methodology and objectives. *Diabetologia* 2004; **47**(3): 566-70.
104. Mooy JM, Grootenhuis PA, de Vries H, et al. Prevalence and determinants of glucose intolerance in a Dutch caucasian population. The Hoorn Study. *Diabetes Care* 1995; **18**(9): 1270-3.

105. Christensen AI, Ekholm O, Glumer C, et al. The Danish National Health Survey 2010. Study design and respondent characteristics. *Scand J Public Health* 2012; **40**(4): 391-7.
106. Thuesen BH, Cerqueira C, Aadahl M, et al. Cohort Profile: the Health2006 cohort, research centre for prevention and health. *International journal of epidemiology* 2014; **43**(2): 568-75.
107. DanFunD: Dansk undersøgelse af funktionelle lidelser 2016. https://www.regionh.dk/fcfs/befolkningsbaseret-epidemiologi/befolkningsunders%C3%B8gelser/Sider/DanFund.aspx (accessed 07/10/2016 2016).
108. Ibrügger S, Gøbel RJ, Vestergaard H, et al. Two randomized cross-over trials assessing the impact of dietary gluten or wholegrain on the gut microbiome and host metabolic health. *Journal of Clinical Trials* 2014; **2014**.
109. Berglund G, Elmstahl S, Janzon L, Larsson SA. The Malmo Diet and Cancer Study. Design and feasibility. *Journal of internal medicine* 1993; **233**(1): 45-51.
110. 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. *The Journal of clinical investigation* 1999; **104**(6): 787-94.
111. Mari A, Tura A, Natali A, et al. Impaired beta cell glucose sensitivity rather than inadequate compensation for insulin resistance is the dominant defect in glucose intolerance. *Diabetologia* 2010; **53**(4): 749-56.
112. Colberg SR, Sigal RJ, Fernhall B, et al. Exercise and type 2 diabetes: the American College of Sports Medicine and the American Diabetes Association: joint position statement. *Diabetes Care* 2010; **33**(12): e147-67.
113. Macauley M, Percival K, Thelwall PE, Hollingsworth KG, Taylor R. Altered volume, morphology and composition of the pancreas in type 2 diabetes. *PloS one* 2015; **10**(5): e0126825.
114. Hollingsworth KG, Al-Mrabeh A, Steven S, Taylor R. Pancreatic triacylglycerol distribution in type 2 diabetes. *Diabetologia* 2015; **58**(11): 2676-8.
115. Gastaldelli A, Kozakova M, Hojlund K, et al. Fatty liver is associated with insulin resistance, risk of coronary heart disease, and early atherosclerosis in a large European population. *Hepatology (Baltimore, Md)* 2009; **49**(5): 1537-44.
116. Al-Mrabeh A, Hollingsworth KG, Steven S, Taylor R. Morphology of the pancreas in type 2 diabetes: effect of return of normal insulin secretion over 6 months. European Association for the Study of Diabetes. Munich, Germany; 2016.
117. Novembre J, Johnson T, Bryc K, et al. Genes mirror geography within Europe. *Nature* 2008; **456**(7218): 98-101.
118. Palo JU, Ulmanen I, Lukka M, Ellonen P, Sajantila A. Genetic markers and population history: Finland revisited. *European journal of human genetics : EJHG* 2009; **17**(10): 1336-46.
119. Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika* 1973; **38**(1): 1-10.

120. Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin* 1990; **107**(2): 238-46.
121. Steiger JH, Lind JC. Statistically based tests for the number of common factors. annual meeting of the Psychometric Society, Iowa City, IA; 1980; 1980. p. 424-53.
122. Bentler PM. EQS structural equations program manual: Multivariate Software; 1995.
123. Duncan GE, Perri MG, Theriaque DW, Hutson AD, Eckel RH, Stacpoole PW. Exercise training, without weight loss, increases insulin sensitivity and postheparin plasma lipase activity in previously sedentary adults. *Diabetes Care* 2003; **26**(3): 557-62.
124. Evans EM, Racette SB, Peterson LR, Villareal DT, Greiwe JS, Holloszy JO. Aerobic power and insulin action improve in response to endurance exercise training in healthy 77–87 yr olds. *Journal of Applied Physiology* 2005; **98**(1): 40-5.
125. Bajpeyi S, Tanner CJ, Slentz CA, et al. Effect of exercise intensity and volume on persistence of insulin sensitivity during training cessation. *Journal of applied physiology* 2009; **106**(4): 1079-85.
126. Houmard JA, Tanner CJ, Slentz CA, Duscha BD, McCartney JS, Kraus WE. Effect of the volume and intensity of exercise training on insulin sensitivity. *Journal of Applied Physiology* 2004; **96**(1): 101-6.
127. Kotronen A, Seppala-Lindroos A, Bergholm R, Yki-Jarvinen H. Tissue specificity of insulin resistance in humans: fat in the liver rather than muscle is associated with features of the metabolic syndrome. *Diabetologia* 2008; **51**(1): 130-8.
128. Kotronen A, Juurinen L, Tiikkainen M, Vehkavaara S, Yki-Jarvinen H. Increased liver fat, impaired insulin clearance, and hepatic and adipose tissue insulin resistance in type 2 diabetes. *Gastroenterology* 2008; **135**(1): 122-30.
129. Bacchi E, Negri C, Targher G, et al. Both resistance training and aerobic training reduce hepatic fat content in type 2 diabetic subjects with nonalcoholic fatty liver disease (the RAED2 Randomized Trial). *Hepatology (Baltimore, Md)* 2013; **58**(4): 1287-95.
130. Bacchi E, Negri C, Zanolin ME, et al. Metabolic effects of aerobic training and resistance training in type 2 diabetic subjects: a randomized controlled trial (the RAED2 study). *Diabetes Care* 2012; **35**(4): 676-82.
131. Goodpaster BH, Katsiaras A, Kelley DE. Enhanced Fat Oxidation Through Physical Activity Is Associated With Improvements in Insulin Sensitivity in Obesity. *Diabetes* 2003; **52**(9): 2191-7.
132. Shaw CS, Shepherd SO, Wagenmakers AJ, Hansen D, Dendale P, van Loon LJ. Prolonged exercise training increases intramuscular lipid content and perilipin 2 expression in type I muscle fibers of patients with type 2 diabetes. *American journal of physiology Endocrinology and metabolism* 2012; **303**(9): E1158-65.
133. Finucane F, Sharp S, Purslow L, et al. The effects of aerobic exercise on metabolic risk, insulin sensitivity and intrahepatic lipid in healthy older people from the Hertfordshire Cohort Study: a randomised controlled trial. *Diabetologia* 2010; **53**(4): 624-31.
134. Galbo H. Hormonal and metabolic adaptation to exercise: Georg Thieme Verlag; 1983.

135. King DS, Dalsky GP, Clutter WE, et al. Effects of lack of exercise on insulin secretion and action in trained subjects. *American Journal of Physiology-Endocrinology And Metabolism* 1988; **254**(5): E537-E42.
136. King DS, Staten MA, Kohrt WM, Dalsky G, Elahi D, Holloszy J. Insulin secretory capacity in endurance-trained and untrained young men. *American Journal of Physiology-Endocrinology And Metabolism* 1990; **259**(2): E155-E61.
137. Mikines KJ, Sonne B, Tronier B, Galbo H. Effects of training and detraining on dose-response relationship between glucose and insulin secretion. *American Journal of Physiology-Endocrinology And Metabolism* 1989; **256**(5): E588-E96.
138. Dela F, Mikines K, Tronier B, Galbo H. Diminished arginine-stimulated insulin secretion in trained men. *Journal of Applied Physiology* 1990; **69**(1): 261-7.
139. Lahjibi E, Heude B, Dekker JM, et al. Impact of objectively measured sedentary behaviour on changes in insulin resistance and secretion over 3 years in the RISC study: interaction with weight gain. *Diabetes & metabolism* 2013; **39**(3): 217-25.
140. Dela F. Functional adaptation of the human β-cells after frequent exposure to noradrenaline. *The Journal of physiology* 2015; **593**(Pt 14): 3199-206.
141. O'Regan DP, Callaghan MF, Wylezinska-Arridge M, et al. Liver fat content and T2*: simultaneous measurement by using breath-hold multiecho MR imaging at 3.0 T--feasibility. *Radiology* 2008; **247**(2): 550-7.
142. Glover G, Schneider E. Three‐point dixon technique for true water/fat decomposition with B0 inhomogeneity correction. *Magnetic resonance in medicine* 1991; **18**(2): 371-83.
143. Ferrannini E, Mari A. beta-Cell function in type 2 diabetes. *Metabolism: clinical and experimental* 2014; **63**(10): 1217-27.
144. Kline RB. Assumptions in Structural Equation Modeling. In: Hoyle RH, ed. Handbook of Structural Equation Modeling. New York, USA: The Guilford Press; 2012: 111-25.

Printed by Media-Tryck, Lund University 2016
Nordic Ecolabel 3041 0903

Robert W. Koivula completed his BSc(Hons) in Physiology and Sports Science at the University of Glasgow and his MSc in Physical Activity and Health at Loughborough University. Robert has completed his PhD at the Genetic and Molecular Epidemiology unit at the Lund University Diabetes Centre. Robert´s doctoral thesis focuses on the role of lifestyle on glycaemic control before and after the onset of type 2 diabetes.

LUND UNIVERSITY
Faculty of Medicine

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2016:138
ISBN 978-91-7619-365-5
ISSN 1652-8220