Gene x lifestyle interactions in type 2 diabetes mellitus and related traits

Ema C. Brito

Umeå 2010
To my heroes, my son Luis Afonso and my daughter Inês
ABSTRACT

Gene x lifestyle interactions in type 2 diabetes mellitus and related traits

Background: Type 2 diabetes is thought to result from interactions between genetic and lifestyle factors, but few robust examples exist. The overarching aim of this thesis was to discover such interactions by studying cohorts of white youth and adults from northern Europe in which physical activity, genotypes, and diabetes-related traits or diabetes incidence had been ascertained.

Methods: The thesis includes four papers. In Paper I, we investigated associations and interactions between 35 common PPARGC1A polymorphisms and cardiovascular and metabolic disease traits in 2,101 Danish and Estonian children from the European Youth Heart Study (EYHS). Paper II used the same cohort to test associations and interactions on cardiometabolic traits for the diabetes-predisposing TCF7L2 polymorphism. In Paper III, we assessed associations for 17 type 2 diabetes gene polymorphisms on impaired glucose regulation (IGR) or incident type 2 diabetes, and tested whether these effects are modified by physical activity in a prospective cohort study of ~16,000 initially non-diabetic Swedish adults – the Malmö Preventive Project (MPP). Paper IV aimed to replicate main genetic effects and gene x physical activity interactions for an FTO polymorphism on obesity in 18,435 primarily non-diabetic Swedish (MPP) and Finnish (Prevalence, Prediction and Prevention of Diabetes in Botnia) adults.

Results: In Paper I, nominally significant associations were observed for BMI (rs10018239, \(P=0.039\)), waist circumference (rs7656250, \(P=0.012\); rs8192678 [Gly482Ser], \(P=0.015\); rs3755863, \(P=0.02\); rs10018239, \(P=0.043\)), and fasting glucose concentrations (rs2970869, \(P=0.018\)). Stronger associations were observed for aerobic fitness (rs7656250, \(P=0.005\); rs13117172, \(P=0.008\)) and fasting glucose concentrations (rs7657071, \(P=0.002\)). None remained significant after correcting for multiple statistical comparisons. We proceeded by testing for gene x physical activity interactions for the polymorphisms that showed statistical evidence of association (\(P<0.05\)) in the main effect models, but none was statistically significant. In Paper II, the minor T allele at the rs7903146 variant was associated with higher glucose levels in older (beta=–0.098 mmol/l per minor allele copy, \(P=0.029\)) but not in younger children (beta=–0.001 mmol/l per minor allele copy, \(P=0.972\)). A significant inverse association between the minor allele at rs7903146 and height was evident in boys (beta=–1.073 cm per minor allele copy, \(P=0.001\)), but not in girls. The test of interaction between the TCF7L2 rs7903146 variant and physical activity on HOMA-B was nominally statistically significant (beta=0.022, \(P_{interaction}=0.015\)), whereby physical activity reduced the effect of the risk allele on estimated beta-cell function. In Paper III, tests of gene x physical activity interactions on IGR-risk for three polymorphisms were nominally statistically significant: CDKN2A/B rs10811661 (\(P_{interaction}=0.015\)); HNF1B rs4430796 (\(P_{interaction}=0.026\)); PPARG rs1801282 (\(P_{interaction}=0.04\)). Consistent interactions were observed for the CDKN2A/B (\(P_{interaction}=0.013\)) and HNF1B (\(P_{interaction}=0.0009\)) variants on 2 hr glucose concentrations. Where type 2 diabetes was the outcome, only one statistically significant interaction effect was observed and this was for the HNF1B rs4430796 variant (\(P_{interaction}=0.0004\)). The interaction effects for HNF1B on 2 hr glucose and incident diabetes remained significant after correction for multiple testing (\(P_{interaction}=0.015\) and 0.0068, respectively). In Paper IV, the minor A allele at rs9939609 was associated with higher BMI (\(P<0.0001\)). The tests of gene x physical activity interaction on BMI were not statistically significant in either cohort (Sweden: \(P=0.71\), Finland: \(P=0.18\)).

Conclusions: Variation at PPARGC1A is unlikely to have a major impact on cardiometabolic health in European children, but physical activity may modify the effect of the TCF7L2 variants on beta-cell function in this cohort. In Swedish adults, physical activity modifies the effects of common HNF1B and CDKN2A/B variants on risk of IGR and also modifies the effect of the HNF1B on type 2 diabetes risk. In Swedish and Finnish adults, we were unable
confirm previous reports of an interaction between *FTO* gene variation and physical activity on obesity predisposition.

**Keywords:** Gene x environment interaction · gene x lifestyle interaction · physical activity · type 2 diabetes · European Youth Heart Study · Malmö Preventive Project · Prevalence, Prediction and Prevention of Diabetes in Botnia · *PPARGC1A* · *CDKN2A/B* · *HNF1B* · *TCF7L2* · *FTO*
LIST OF PAPERS

This thesis is based on the following papers which will hereafter be referred to by their Roman numerals:


RELATED PUBLICATIONS


The original papers have been reprinted with permission of the publishers.
ABBREVIATIONS

ADA American Diabetes Association
ADAMTS9 ADAM metallopeptidase with thrombospondin type 1 motif, 9
BMI body mass index
BP blood pressure
CAMK1D calcium/calmodulin-dependent protein kinase ID
CDKAL1 CDK5 regulatory subunit associated protein 1-like 1
CDKN2A/B cyclin-dependent kinase inhibitor 2A/B
CI confidence interval
CVD cardiovascular disease
DNA deoxyribonucleic acid
DBP diastolic blood pressure
EYHS European Youth Heart Study
FDR false discovery rate
FTO fat mass and obesity associated
GWAS genome-wide association study
HDL-C high density lipoprotein cholesterol
LDL-C low density lipoprotein cholesterol
HHEX hematopoietically expressed homeobox
HOMA homeostasis model assessment
HNF1B HNF1 homeobox B
HWE Hardy-Weinberg Equilibrium
IFG impaired fasting glucose
IGF2BP2 insulin-like growth factor 2 mRNA binding protein 2
IGR impaired glucose regulation
IGT impaired glucose tolerance
JAZF1 JAZF zinc finger 1
KCNJ11 potassium inwardly-rectifying channel, subfamily J, member 11
LDL-C low-density lipoprotein-cholesterol
MAF minor allele frequency
MET metabolic equivalent of task
MTNR1B melatonin receptor 1B
MPP Malmö Preventive Project
NEFA non-esterified fatty acid
NOTCH2 notch homolog 2 (drosophila)
OGTT oral glucose tolerance test
OR odds ratio
PPARG peroxisome proliferator-activated receptor gamma
PPARGC1A peroxisome proliferator-activated receptor gamma, coactivator 1 alpha
PPP-Botnia Prevalence, Prediction and Prevention of Diabetes in Botnia
RD risk difference
RNA ribonucleic acid
RR relative risk
SBP systolic blood pressure
SLC30A8 solute carrier family 30 (zinc transporter), member 8
SNP single nucleotide polymorphism
TCF7L2 transcription factor 7-like 2 (T-cell specific, HMG-box)
THADA thyroid adenoma associated
TSPAN8 TSPAN8 tetraspanin 8
TCF7L2 transcription factor 7-like 2 (T-cell specific, HMG-box)
T2D type 2 diabetes
WFS1 wolfram syndrome 1 (wolframin)
WHO World Health Organization
WHR waist-to-hip ratio
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>I</td>
</tr>
<tr>
<td>LIST OF PAPERS</td>
<td>III</td>
</tr>
<tr>
<td>ABBREVIATIONS</td>
<td>IV</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>V</td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 What is diabetes?</td>
<td>2</td>
</tr>
<tr>
<td>1.2 The global burden of T2D</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Lifestyle and T2D risk</td>
<td>4</td>
</tr>
<tr>
<td>1.3.1 Observational epidemiology</td>
<td>4</td>
</tr>
<tr>
<td>1.3.2 Clinical trials and other experimental studies</td>
<td>7</td>
</tr>
<tr>
<td>1.4 Features of the metabolic syndrome and mechanisms through which they influence diabetes risk</td>
<td>9</td>
</tr>
<tr>
<td>1.5 Inherited factors in T2D</td>
<td>11</td>
</tr>
<tr>
<td>1.5.1 Ethnicity</td>
<td>11</td>
</tr>
<tr>
<td>1.5.2 Genetics</td>
<td>12</td>
</tr>
<tr>
<td>1.6 Genetic association studies of T2D</td>
<td>13</td>
</tr>
<tr>
<td>1.7 Overview of literature of genes of interest</td>
<td>15</td>
</tr>
<tr>
<td>1.7.1 NOTCH2</td>
<td>15</td>
</tr>
<tr>
<td>1.7.2 THADA</td>
<td>16</td>
</tr>
<tr>
<td>1.7.3 ADAMTS9</td>
<td>16</td>
</tr>
<tr>
<td>1.7.4 PPARG</td>
<td>16</td>
</tr>
<tr>
<td>1.7.5 IGF2BP2</td>
<td>16</td>
</tr>
<tr>
<td>1.7.6 WFS1</td>
<td>16</td>
</tr>
<tr>
<td>1.7.7 CDKAL1</td>
<td>17</td>
</tr>
<tr>
<td>1.7.8 JAZF1</td>
<td>17</td>
</tr>
<tr>
<td>1.7.9 SLC30A8</td>
<td>17</td>
</tr>
<tr>
<td>1.7.10 CDKN2A/B</td>
<td>17</td>
</tr>
<tr>
<td>1.7.11 CAMK1D</td>
<td>17</td>
</tr>
<tr>
<td>1.7.12 HHEX</td>
<td>18</td>
</tr>
<tr>
<td>1.7.13 TCF7L2</td>
<td>18</td>
</tr>
<tr>
<td>1.7.14 KCNJ11</td>
<td>18</td>
</tr>
<tr>
<td>1.7.15 MTNR1B</td>
<td>18</td>
</tr>
<tr>
<td>1.7.16 TSPAN8</td>
<td>19</td>
</tr>
<tr>
<td>1.7.17 FTO</td>
<td>19</td>
</tr>
<tr>
<td>1.7.18 HNF1B</td>
<td>19</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

Type 2 diabetes mellitus (T2D) is a heterogeneous disease that results from the interplay between adverse environmental and genetic risk factors. In the past four years, major advances relating to the genetic basis of T2D have been made. This has cumulated in the discovery and confirmation of around 20 common predisposing loci (1), but the variance in disease risk explained by these variants is much lower than predicted based on heritability studies. It follows that the genetic associations discovered to date represent the tip of the iceberg with respect to the genetic landscape of T2D risk.

In contrast to the genetic basis to T2D, a great deal of robust evidence exists which documents the impact of lifestyle behaviours on the development of T2D. Epidemiological studies have identified strong T2D risk relationships for obesity, sedentary behaviours (2-4) and diets rich in energy (5), processed carbohydrates (6) and animal fats (7). The strongest evidence comes from clinical trials, which show that intensive lifestyle interventions targeting weight-loss through diet modification and physical activity have a major beneficial impact on diabetes incidence in high-risk individuals (8, 9).

The pattern of disease occurrence within and between populations that differ in their genetic and environmental underpinnings suggests that T2D is caused by the interaction between adverse lifestyle behaviours and in part by the genetic profile of an individual. For many, this seems a reasonable assumption, but there is little empirical evidence that defines the specific nature of these interactions. The availability of detailed information on gene x lifestyle interactions may enhance our understanding of the molecular basis of T2D, elucidate the mechanisms through which lifestyle exposures influence diabetes risk, and possibly help refine strategies for diabetes prevention or treatment.

The overarching objective of the work described in this thesis was to discover and describe examples of gene x lifestyle interactions on T2D and related quantitative traits. I specifically focused on physical activity defined as any “bodily movement produced by skeletal muscles that requires energy expenditure” (10) as the “lifestyle” exposure for three reasons: i) it is well established that physical inactivity is a major modifiable risk factor for T2D (11); ii) physical activity can be quantified within the setting of an epidemiological study; iii) experimental studies clearly show that physical activity can modify the way in which genes involved in energy homeostasis (a key factor in diabetes pathogenesis) act (12).

“The problem of understanding the genetic nature of man is both a philosophical and, in these days of rapidly changing environment, a practical challenge. Progress demands both a broad approach on the theoretical level and a very specific approach geared to particular traits presenting favorable analytic opportunities.”

James V Neel, 1962 (13)
1.1 What is diabetes?

Diabetes mellitus is a group of metabolic diseases characterized by hyperglycaemia (elevated levels of glucose in the blood) resulting from defects in insulin secretion, insulin action, or both (14). There are two major types of diabetes mellitus: type 1 and type 2 diabetes (14). Type 1 diabetes is an autoimmune disease that usually occurs in childhood but the onset may occur at any age; this type of diabetes results from a cellular-mediated autoimmune destruction of the beta-cells in the pancreatic islets which usually leads to absolute insulin secretion deficiency. T2D on the other hand is a metabolic disorder that generally appears later in life but may occur in childhood and is characterized by the combination of insulin resistance and relative insulin secretion deficiency (15). This type of diabetes usually begins predominantly with insulin resistance, which is a condition characterized by the inability of cells to respond to the action of insulin in transporting glucose from the bloodstream into muscle, fat, and liver cells (16). This condition causes a compensatory increase in the secretion of insulin from the pancreatic beta cells (hyperinsulinemia) in order to overcome the state of insulin resistance and thus help glucose enter the cells. However, in the long term, beta-cell mass and function progressively declines (17). The natural history of T2D in many individuals involves years of insulin resistance balanced by elevated insulin secretion. The pivotal point is when the beta-cells begin to fail, and insulin production declines. Thus T2D is characterized by both defects in insulin secretion and by cellular insulin resistance.

Beside type 1 and type 2 diabetes there are several other classes of diabetes which are characterized by genetic defects of beta-cell function (Maturity Onset Diabetes of the Young: MODY1-6), transient neonatal diabetes, genetic defects in insulin action, disorders of the exocrine pancreas, endocrinopathies, drug- or chemical-induced diabetes, infections induced diabetes, uncommon forms of immune-mediated diabetes, and gestational diabetes mellitus (15). However, approximately 90–95% of all diabetes cases are T2D (18).

T2D is diagnosed using either repeat fasting or two hour plasma glucose concentrations follow oral glucose challenge (i.e. fasting blood glucose levels ≥126 mg/dl [≥7.0 mmol/l] without symptoms, 2-hour glucose levels ≥200 mg/dl [≥11.1 mmol/l] after an oral glucose tolerance test (OGTT) without symptoms, or random blood glucose levels ≥200 mg/dl [≥11.1 mmol/l] with symptoms); such tests should be repeated on a separate day in order to confirm the diagnosis of T2D (19).

Because the progression from normoglycaemia to hyperglycaemia is slow and gradual, there are intermediate stages. These are defined as impaired fasting glucose (IFG) and impaired glucose tolerance (IGT) where glucose values are considered to be above “normal” glucose tolerance but below those used to diagnose diabetes. As a result, many individuals have ‘intermediate hyperglycaemia’ (14) (sometimes referred as ‘impaired glucose regulation’, ‘pre-diabetes’ (20) or, ‘non-diabetic hyperglycaemia’ (21)). According to the criteria of World Health Organization (WHO) (14) IFG and IGT are diagnosed when a person presents with fasting venous plasma glucose levels between 100–125 mg/dl (6.1 to 6.9 mmol/l) and 2-hour blood glucose level between 140–199 mg/dl (7.8 –11.1 mmol/l) during a 75-g OGTT. In 2003, the American Diabetes Association (ADA) recommended that the IFG threshold should be lowered to
100 mg/dl (5.6 mmol/l) (22). However, not all agencies, including WHO, have accepted this recommendation.

The majority of those diagnosed with IFG and IGT (around 60%) do subsequently develop T2D (23). It is for this reason that IFG and IGT are commonly used to identify high-risk groups. For example, all-cause mortality rates in individuals with IFG or IGT are almost twice those of persons with normal glucose levels (24), justifying early intervention.

Although the aetiology of T2D has not been established in full, a number of risk factors are well defined. According to the ADA (11), the risk of developing T2D is associated with age (increased risk at ≥45 years), overweight/obesity, and lack of physical activity (PA). T2D is more common in individuals with a family history of the disease, in certain ethnic groups (e.g. African-Americans, Hispanic-Americans, Native Americans, Asian-Americans, and Pacific Islanders), and in individuals with hypertension (≥140/90 mmHg in adults), dyslipidaemia (HDL cholesterol ≤35 mg/dl [0.90 mmol/l] and/or a triglyceride level ≥250 mg/dl [2.82 mmol/l]), IFG, IGT, a history of vascular disease or gestational diabetes, or polycystic ovary syndrome. In addition, a range of common genetic variants are also known to raise the risk of T2D (25-27), of which some may interact with lifestyle factors to modify the risk of the disease (12).

1.2 The global burden of T2D

Over recent decades, the progressively increasing global prevalence of T2D (28) has created a major public health challenge. This is because T2D is a major cause of premature morbidity and mortality, and as such it imposes a heavy burden on affected individuals and society as a whole. Furthermore, the disease is associated with long-term microvascular and macrovascular injury, such as retinopathy (eye disease), nephropathy (kidney disease), neuropathy (damaged nerves), peripheral vascular disease, cerebrovascular disease (including hemorrhagic stroke), and atherosclerotic disease (often leading to myocardial infarcts) (14, 19). Mortality rates in adults with T2D are 2- to 4-fold higher than those observed in non-diabetic individuals, with many premature deaths in people with diabetes being attributable to cardiovascular disease (CVD) (29, 30).

According to the WHO, the number of people with diabetes of all ages worldwide increased from 30 million to 171 million between 1985 and 2000 (31). These numbers are expected to increase to 366 million in 2030. The estimated prevalences of diabetes approximated 2.8% in 2000 and are predicted to be around 5.8% in 2030 (32). In Sweden, it is estimated that diabetes affects ~350,000 people (2.2–4.5% of the population) (33-35). Costs incurred from diabetes complications make up 1.6–6.6% of total health care spending in eight European countries, including Sweden (~5%) (36).

Although T2D has traditionally been considered a disease of adult onset, in the past decade T2D incidence has increased rapidly in the young; in some aboriginal groups such as Pima Indians, T2D is as common in children as it is middle-aged adults of lower risk ethnic groups (37). The explanations for the rising trends in paediatric T2D are likely to be attributable to changing lifestyles and the high prevalence of obesity in contemporary children (38, 39). Data on T2D incidence in European children are scarce. Nevertheless, the proportion of children of European descent diagnosed with T2D appears to remain low. A
French study (40) indicated a relatively low, but increasing, number of children with glucose levels exceeding the thresholds for T2D and an Austrian population-based study (41) reported an incidence of 0.25/100,000/year. In the U.K., the incidence of T2D was substantially higher in children from ethnic minority groups: 3.9 and 1.25/100,000/year for children of African and South Asian origin, respectively, compared to 0.35/100,000/year in ethnically European children (42).

Despite the increasing number of children with T2D, the WHO estimates that between 2000 and 2030 the most striking increase in T2D prevalence will be among persons aged 65 years and older. By 2030, it is estimated that more than 48 and 82 million older adults (>65 yrs) in developed and developing countries, respectively, will be afflicted with T2D (32). The DECODE Study Group, which is comprised of nine European countries (including Sweden) estimates that the prevalence of T2D will be <10% in persons younger than 60 years of age and 10–20% in persons aged 60–79 years (43). The reason for this shift in the demographic distribution of affected individuals is threefold: firstly, global populations are aging; secondly the complications of T2D can be treated more efficiently than ever before, which means that people are living longer with diabetes; and thirdly, lifestyle behaviours that increase diabetes risk are becoming more common in all age groups (44, 45).

1.3 Lifestyle and T2D risk

There are three major lifestyle components that are established risk factors for T2D: physical inactivity, poor diet (i.e. high animal fat/sugar and low fresh fruit/vegetable content), and obesity (2, 6, 46, 47). Although studies of T2D frequently focus on specific lifestyle behaviours, adverse lifestyle behaviours tend to coalesce and the effect of a specific behaviour on diabetes risk may be confounded, mediated, or modified by other behaviours. For example, it is likely that very sedentary persons also eat less healthy diets, smoke more frequently, and engage in other unhealthy behaviours more frequently than physically active persons. In contrast, people who maintain a healthy diet may be more likely to exercise regularly, smoke less, and so forth. The extent to which studies are able to distinguish the effects of lifestyle factors on diabetes risk varies, but is generally greater in randomized clinical trials than in observational epidemiological studies. Thus, the purpose of the following section is to give an overview of epidemiological and experimental studies that have focused on the role of lifestyle behaviours in T2D risk, with emphasis on physical activity and obesity.

1.3.1 Observational epidemiology

Obesity is associated with at least 45 co-morbidities (48) including T2D, some types of cancer, and cardiovascular disease, all of which are major causes of death in modern society (49).

Obesity results from chronic positive energy balance, caused mainly by overeating and physical inactivity (50, 51). Globally, more than 1.6 billion adults are estimated to be overweight and 400 million are estimated to be clinically obese according to the WHO (52). The cost of obesity on individual wellbeing and in financial terms has been estimated to account for up to 16% of the global burden of disease (53). The WHO defines overweight and obesity as a body
mass index (BMI: calculated as the ratio between weight [kg] and height squared [m²]) of ≥25.0–<30 kg/m² and ≥30 kg/m², respectively (52). The International Obesity Task Force has estimated that more than 155 million children worldwide are overweight or obese (53). This information was based on growth curves which estimate cut-points related to different age groups as a function of adult categories for overweight and obesity (54). Numerous studies worldwide have documented the progressively rising prevalence of obesity in paediatric cohorts during the past two decades. This increase appears to be followed by an increasing incidence of T2D in these age-groups (55).

Many epidemiological studies have documented the strong association between obesity and T2D. For example, the 14-year follow-up Nurses’ Health Study of 114,281 women aged 30–55 years at baseline observed that BMI was the most important risk factor for T2D (56). In that study, age-adjusted BMI was positively related with T2D; for example, women with a BMI of 27.0–28.9 kg/m² are at 15.8-fold (95% CI: 12.7–19.8) greater risk of T2D compared with women whose BMI was below 22 kg/m². The same study also showed that weight change influences the risk of T2D. For example, women who after the age of 18 years gained between 5.0–7.9 kg and 8.0–10.9 kg were at 1.9-fold (95% CI, 1.5–2.3) and 2.7-fold (95% CI, 2.1–3.3) greater risk of developing T2D, respectively, compared with women who gained less than 5 kg. Thus, even fairly small increments in weight gain can substantial increase diabetes risk. Convervely, women who lost around 5 kg in weight reduced their risk of diabetes by approximately 50%.

Further evidence for an association between obesity and T2D comes from the 12.8-year follow-up British Regional Heart Study, which included 7,577 men aged 40–59 years at baseline. The study identified a strong graded association between BMI and risk of diabetes (57). The age-adjusted relative risk (RR) was 11.6 (95% CI: 5.4–16.8) for men in the highest quintile of BMI (BMI >27.9 kg/m²) compared with those in the lowest quintile (BMI ≤22.9 kg/m²). Adjustment for potential confounders or mediators reduced the magnitude of the relationship, but even in these more conservative analyses the most obese men were at roughly seven times greater risk of T2D compared with the least obese (BMI quintiles 5 vs. 1).

The 16-year follow-up of the Nurses’ Health Study (N=84,941 women) reached similar conclusions to the British Regional Heart Study, confirming that overweight and obesity are major predictors of diabetes (2). However, in this report the authors also illustrated that adjustment for physical inactivity, poor diet, and smoking reduced the magnitude of the association between obesity and T2D, supporting the view that these modifiable behaviours may act on obesity to reduce T2D risk. Other epidemiological studies such as the Womens’ Health Study (N=37,878 women aged ≥45), have reached similar conclusions (58).

Although many studies focus on body mass per se, body fat distribution is also an important independent risk factor for diabetes. A cross-sectional study of 5,080 individuals from different ethnic populations (Hindu and Muslim Asian Indians, African-origin Creoles, and Chinese Mauritians) showed that waist-to-hip ratio (WHR) is positively associated with T2D independently of BMI (59). This study also illustrated that physical inactivity is an important independent risk factor for T2D. Elsewhere, a 13.5-year follow-up of 792 Swedish men aged 54 years confirmed the previous findings, suggesting that
WHR is significantly associated with T2D risk even after adjustment for BMI (60).

In the 5-year follow-up of the Health Professionals Study (N=51,529 men aged 40–75 years), abdominal obesity assessed by both waist circumference and by WHR were positively associated with the risk of T2D; waist circumference was a stronger predictor of risk than WHR (61).

Besides obesity, other adverse lifestyle behaviours involving multiple mechanisms of action contribute to the development of T2D (62-65). Studies including children and young adolescents from several countries have shown that the main risk factors for obesity include the consumption of energy-dense foods, physical inactivity, TV viewing, and parental obesity. Studies of TV viewing illustrate that children who watch TV for >2 h per day are more likely to consume high-energy drinks, snacks, sweets, and white bread (66), and less fruit, water, milk, and brown bread, than children who watch <2 hrs per day (67). In addition, the odds of being overweight or obese increases with duration of TV viewing (68). By contrast, physical activity is inversely related with TV viewing and snacking (66, 69). In European children, those who accumulated <1 hr of moderate physical activity per day were more obese than those who accumulated >2 hrs of activity per day (70). Parental obesity, especially in mothers (71), more than doubles the risk of adult obesity in both obese and non-obese children <10 years of age (72, 73). Furthermore, parents of overweight or obese children tend to have a weaker understanding of their child’s level of overweight and the associated health risks (74).

As with paediatric cohorts, epidemiological studies of adults consistently demonstrate that physically active individuals are less likely to develop T2D compared to sedentary individuals. For example, the 14-year follow-up University of Pennsylvania Alumni Health Study (N=5,990 men aged 39–68 years) showed that physical activity (leisure-time physical activity, expressed in kcal expended per week through walking, stair climbing, and sports) was inversely associated with the incidence of T2D (75). Incidence rates declined as energy expenditure rose from 500 to 3,500 kcal/week. The age-adjusted RR of T2D was reduced by about 6% for each 500 kcal increment increase in physical activity energy expenditure. Similarly, in the 8-year follow-up Nurses’ Health Study (n=87,253 women aged 34–59 years), an inverse graded association between physical activity and incidence of T2D was observed (76). The age-adjusted risk of T2D was 0.67 (95% CI: 0.60–0.75) in women who engaged in vigorous exercise at least once a week compared with women who exercised less than once weekly. The Nurses’ Health Study also showed that both walking and vigorous activity are protective of T2D (77). After adjustment for potential confounders and mediators, including BMI, the risk of developing T2D across increasing quintiles of physical activity (defined as MET hrs per week, with the lowest MET quintile as the referent) yielded a 0.84– (95% CI: 0.72–0.97), 0.87– (95% CI: 0.75–1.02), 0.77– (95% CI: 0.65–0.91), and 0.74– (95% CI: 0.62–0.89) fold reductions in risk. Even in women who abstained from vigorous exercise, the reduction in risk across quintiles was substantial (RR in Q2=0.95 [95% CI: 0.79–1.15], RR in Q3=0.80 [95% CI: 0.65–0.99], RR in Q4=0.81 [95% CI: 0.66–1.01], and RR in Q5=0.74 [95% CI: 0.59–0.93]).

In the 10-year follow-up Health Professional’s Follow-up Study (n=37,918 men aged 40–75 years), sedentary lifestyle behaviours such as TV viewing were positively related with increased risk of T2D (47). After adjustment for potential confounders and mediators, the risk of developing T2D
increased in a dose-dependent manner across quintiles of physical activity (MET-hours per week), with the most active men having roughly half the risk of T2D compared with the least active 0.51 (95% CI: 0.41–0.63). In analyses adjusted for multiple covariates, including physical activity, weekly TV viewing was associated in a dose-dependent manner with T2D risk. For example, diabetes risk increased in men who viewed more than 40 hrs of TV each week by 2.87-fold (95% CI: 1.46–5.65) compared with those who viewed <1 hr/week. These data illustrate the complex and independent relationships of physical activity and sedentary behaviour with T2D risk.

Because obesity and central fat distribution are major risk factors for T2D, and physical activity is both inversely correlated with the risk of developing diabetes and obesity, the conclusions of previous studies on the association between physical activity and insulin resistance have depended somewhat on whether the level of obesity was adjusted for in analyses. Although epidemiological studies such as those described above suggest that physical activity is protective of obesity and T2D, such studies are limited in their ability to control for factors that might confound these relationships. Because of this limitation, it is often difficult or impossible to deduce the causal relationships between physical activity and diabetes risk based on epidemiological evidence alone. The derivation of causal evidence requires randomized clinical trials of lifestyle intervention, which is the focus of the following section.

1.3.2 Clinical trials and other experimental studies

Exercise training studies have found that physical activity is positively related with insulin action (78) and glucose metabolism (79) in healthy individuals and those at high risk of T2D. Exercise often normalizes plasma glucose levels by improving insulin sensitivity and glucose transportation (62). Exercise can also improve endothelial function, reduce inflammation, and beneficially affect the autonomic nervous system (80). Even in the absence of weight loss, exercise can enhance insulin sensitivity (81) and glycemic control (82). These findings are particularly relevant as they show that regular exercise can be used effectively as a treatment for preventing T2D from developing in individuals with IFG/IGT and for improving insulin action in people with manifest diabetes.

In one of the first nonrandomized intervention study based in the southern Swedish city of Malmö, 415 men (aged 47 to 49 years) were recruited from a cohort of 6,956 men whose glucose levels had been assessed. The trial was one of the first to show that weight reduction and increased aerobic fitness brought about by exercise and diet intervention improve glucose tolerance (83). The intervention groups comprised 41 and 181 men with T2D and IGT respectively, and the comparison groups comprised 79 and 114 men with IGT and normal glucose tolerance respectively. Participants in the intervention groups received 6 months of dietary treatment, and 6 months of supervised exercise training. After 6 years, body weight was reduced by 2.0–3.3 kg in the intervention groups, whereas body weight increased in the control groups by 0.2–2 kg. Participants from the intervention group who at baseline had IGT improved their glucose control by 75.8%, and 10.6% developed T2D. By contrast, those in the control group with baseline IGT experienced 67.1% deterioration in glucose control, and 28.6% developed T2D. Weight reduction ($r=0.19$, $P<0.02$) and increased fitness ($r=0.22$, $P<0.02$) were positively correlated with improved glucose tolerance. This early study provided evidence
that changes in lifestyles might prevent diabetes even after the intervention ends.

In a second landmark study from Scandinavia called the Finnish Diabetes Prevention Study, 522 overweight men and women (mean age 55 yrs) with IFG/IGT were randomized to receive either intensive lifestyle intervention with exercise and diet modification or to a control arm involving standard care (81). Participants randomized to the lifestyle intervention arm received detailed advice regarding the five goals of the intervention: i) weight loss, ii) reduced total fat intake, iii) reduced saturated fat intake, iv) increased fibre intake, and v) exercise (for at least 30 minutes per day). After 1 and 2 years follow-up, there was a weight loss of 4.2±5.1 kg and 3.5±5.5 kg in the intervention group and 0.8±3.7 kg and 0.8±4.4 kg in the control group, respectively (both \( P<0.001 \)).

After an average of 3.2 years follow-up, the incidence of diabetes in the intervention group was 58% less compared with the control group (hazard ratio, 0.4; 95% CI: 0.3 - 0.7; \( P<0.001 \)). Although this study was not designed to assess the individual contributions of diet and exercise to diabetes risk reduction, the reduction in diabetes risk was directly proportional to adherence to the lifestyle recommendations. These results suggest that T2D can be prevented by changes in the lifestyles among high-risk individuals.

The Diabetes Prevention Program was a multicenter randomized controlled trial similar in design to the Finnish Diabetes Prevention Study. It involved an average follow-up of 2.8 years in 3,234 initially high-risk but non-diabetic individuals who were randomized to receive an intensive lifestyle intervention or standard care (control). The trial also included metformin and troglitazone arms, and in this respect differed in design from the Finnish Diabetes Prevention Study (46). The intensive lifestyle modification arm included goals to achieve at least 7 percent weight loss, dietary modification, and at least 150 minutes of physical activity per week. As with the Finnish Diabetes Prevention Study, the reduction in diabetes incidence attributable to the lifestyle intervention was 58% when compared with the T2D incidence rate in the control group. The efficacy of the lifestyle intervention was similar across ethnic groups and in men and women alike. These findings confirmed the findings from the Finnish Diabetes Prevention Study and showed that lifestyle intervention is more effective than metformin for preventing or delaying T2D in this population.

In one of two pioneering trials from Asia called the Da Qing IGT and Diabetes Study, 577 Chinese men and women (mean aged 46.5 years) with IGT were randomised to diet, exercise, diet + exercise interventions, or a control intervention comprising standard care (84). The cumulative incidence of diabetes at 6 years was significantly lower in the diet group (43.8%), the exercise group (41.1%), and the diet-plus-exercise group (46%) compare with the control group (67%). These effects were similar in lean and obese participants. The Da Qing Study is the only large-scale RCT to date where the impact of diet and exercise interventions have been compared as separate and combined treatments for diabetes risk reduction. The second large-scale Asian diabetes prevention study took place in India and was called the Indian Diabetes Prevention Programme (85). The design of this study mimicked that of the Diabetes Prevention Program and involved an average follow-up duration of 3 years, where 531 overweight men and women (mean age 45.9±5.7 years) with IGT at baseline were randomized to one of four arms: lifestyle modification, metformin, lifestyle modification + metformin, or placebo control. At the end of
the trial, the cumulative incidence of diabetes was 55.0% in the control group and 39.3% in the lifestyle intervention group. Similar effects were observed for metformin or metformin + lifestyle (40.5% and 39.5% respectively). The reduction in diabetes incidence attributable to lifestyle modification, metformin treatment, or lifestyle modification + metformin was 28.5%, 26.4%, and 28.2%, respectively, than in the control group. These findings reinforce the effectiveness of lifestyle modification in non-white ethnic groups.

One important question that remains largely unanswered is whether the expensive and tightly controlled interventions used in the Finnish Diabetes Prevention Study and the Diabetes Prevention Program can be translated to the primary care setting where resources are often limited. In a recent study from Northern Sweden called the Swedish Björknäs Study, 151 men and women at high risk of cardiovascular disease were randomized to receive a group-mediated intensive lifestyle intervention or standard care (control) (86). Follow-up lasted for 3 years on average, during which time waist circumference, blood pressure, and aerobic fitness had improved more in the intervention group than in the control group. No detectable changes in glucose or lipid levels were observed.

1.4 Features of the metabolic syndrome and mechanisms through which they influence diabetes risk

At the point of diagnosis, 80% to 95% of people with T2D are overweight or obese with adipose accumulation primarily centring round the abdominal region (87). This type of adiposity is strongly associated with insulin resistance. Although insulin resistance is an important mediator of the relationship between obesity and T2D, most obese insulin resistant individuals are able to maintain normal glucose regulation for many years. For example, in the U.S. about 20–25% of the ‘healthy’ population are estimated to be insulin resistant, but only 7% of the population has clinical diabetes (88). The key step in the progression from insulin resistance to T2D is the diminution of pancreatic beta-cell function, which signals a progressive and often irreversible decline in endogenous insulin production.

Two types of white adipose tissue are distinguishable, namely visceral adipose tissue (also referred to as intra-abdominal adipose tissue), located inside the peritoneal cavity, and subcutaneous adipose tissue, (i.e. adipose tissue located beneath the skin). Visceral adipose tissue appears to predispose greater T2D and cardiovascular risk than subcutaneous adipose tissue, as illustrated by several large epidemiological and physiological studies (89-92). Adipose tissue is an important endocrine organ (92-95). Several studies have shown that in obese individuals, adipose tissue secretes factors that interfere with insulin action in other tissues (including skeletal muscle and liver tissue) and increase the secretion of non-esterified fatty acids (NEFAs), glycerol, hormones (e.g. leptin and adiponectin), proinflammatory cytokines, and other factors that interfere in the development of insulin resistance (96). Among those factors, the most convincing data pertains to NEFAs because they are associated with both obesity and T2D. NEFAs induce insulin resistance. They do this by inhibiting insulin-stimulated glucose uptake in skeletal muscle and by stimulating gluconeogenesis in the liver. NEFAs also impair pancreatic beta-cell function, which in turn adversely affects glucose control.
T2D is more common in individuals with dyslipidaemia (11). The characteristic pattern of lipid abnormalities in patients with diabetes (often referred to as atherogenic dyslipidaemia or diabetic dyslipidaemia) involves elevated triglyceride concentrations, low levels of HDL-C, and elevated levels of small dense low-density lipoprotein cholesterol (sd-LDL-C) particles. In a 20-year follow-up of the Framingham Heart Study, hypertriglyceridemia and low HDL-C were associated with increased T2D risk in both men and women (97). The prevalence of high LDL-C levels in people with diabetes (9% in men and 15% in women) did not differ significantly from the rates in non-diabetic persons (11% in men and 16% in women). By contrast, the prevalence of elevated plasma triglyceride levels in persons with diabetes (19% in men and 17% in women) was significantly higher than in persons free of diabetes (9% in men and 8% in women). The prevalence of low HDL-C levels in people with diabetes (21% in men and 25% women) was almost twice as high as the prevalence in non-diabetic individuals (12% in men and 10% women). In addition, the incidence of cardiovascular disease was higher in people with diabetes compared with those without diabetes. Although T2D and dyslipidaemia are independent risk factors for cardiovascular disease, the combination of T2D with hypercholesterolemia, or with other risk factors such as hypertension and smoking, markedly increases CVD-related mortality rates (98, 99).

Diabetic dyslipidaemia results from lipoprotein dysmetabolism combined with abnormalities in insulin action. However, the mechanisms that underlie the relationship between dyslipidaemia and T2D are poorly understood. What is known is that the lipid changes that occur with diabetes include increased NEFA flux secondary to insulin resistance (100). Characteristically, this involves three steps: i) increased lipolysis from insulin-resistant adipocytes; ii) an increased flux of NEFAs into the liver, which in the presence of adequate glycogen stores promotes triglyceride synthesis and secretion of apolipoprotein B (ApoB) and large very low density lipoprotein cholesterol (VLDL-C) particles; and iii) an impaired ability of insulin to inhibit NEFA synthesis, which leads to enhanced hepatic VLDL-C production, which in turn correlates with the degree of hepatic fat accumulation (100). Hyperinsulinaemia is also associated with low HDL-C levels.

Elevated circulating NEFA levels derived from visceral adipocytes may result in an accelerated hydrolysis of stored triglycerides, which increases the delivery of NEFA to the liver via the portal vein, and insulin resistance ensues (101, 102). Obese individuals tend to have higher rates of NEFA and glycerol release into the portal circulation compared to non-obese individuals (103). Elevated NEFA concentrations promote ectopic fat storage in non-adipose cells such as hepatocytes and myocytes (104, 105). An excess of circulating NEFA is also linked to muscle insulin resistance (106).

High blood pressure is not generally considered a causal factor in the development of T2D. However, arterial endothelial dysfunction has been proposed as a causal link between elevated blood pressure and insulin resistance, which clearly could forge a causal link between blood pressure and diabetes (107, 108). Markers of inflammation such as C-reactive protein have also been related with both incident T2D (109-111) and increase levels of blood pressure (112, 113).

T2D and hypertension are strongly correlated, largely owing to their shared relationships with obesity and other lifestyle factors. Thus, the
relationship between high blood pressure and T2D may not be completely causal. Hypertension affects up to 40% or more of diabetic patients (114, 115). For example, in the Atherosclerosis Risk in Communities study, which is a prospective cohort study of 12,550 initially non-diabetic adults aged 45–64 years, an association between baseline hypertension and incident T2D was observed. During six years of follow-up, there were 1,146 new reported cases of diabetes (116). In multivariate analysis among the subjects who were not taking any antihypertensive medication, hypertensive individuals were at a 2.34-fold (95% CI: 2.16–2.73) higher risk of developing T2D compared with persons who were normotensive at baseline. Similarly, in the Women’s Health Study, which included 38,172 middle-aged women, baseline blood pressure and blood pressure progression were strong and independent predictors of incident T2D among initially healthy women; during 10.2 years of follow-up, 1,672 women developed T2D (117). After adjustment for BMI and other components of the metabolic syndrome, the incidence of diabetes was strongly related with baseline blood pressure. The Women’s Health Study and many others have also shown that elevated blood pressure is a major predictor of cardiovascular disease (117, 118).

1.5 Inherited factors in T2D

Heritability estimates provide an indication of the extent to which genetic and environmental factors influence the variance of specific traits (phenotypes) within populations. Heritability is formally defined as a ratio of variances, specifically as the proportion of total variance in a population for a particular measurement, taken at a particular time or age, that is attributable to variation in additive genetic or total genetic values, termed narrow-sense heritability (h^2) and broad-sense heritability (H^2), respectively (119). The H^2 for T2D ranges from 26% to 75% (120-122).

The offspring risk ratio is often used to express the heritable risk of developing a disease. In the Framingham Offspring Study the RR in offspring with one diabetic parent was ~3.5, and when both parents had diabetes the RR increased to ~6.0, compared with the risk in offspring of non-diabetic parents (123). The heritable risk of diabetes extends beyond the influence of the parents to other family members. In the Framingham study, a history of diabetes in any biologic ancesteral family member or sibling independently and progressively increased diabetes risk in the proband.

Multiple studies of twins also provide compelling evidence for a genetic component for T2D. Estimates for concordance rates range from 0.29 to 1.00 in monozygotic (MZ) twins, while in dizygotic (DZ) twins the range was 0.10–0.43(120, 121, 124-127). Lastly, the high levels of heritability for insulin sensitivity and insulin secretion also supports a genetic component to diabetes (128-130).

1.5.1 Ethnicity

Evidence for a genetic component of T2D comes in part from ethnic-specific differences in prevalence rates for T2D, which range from 1% in Chile Mapuche Indian, 2% among Caucasians in Europe, to frequencies as high as 41% in the Nauru (Pacific Island) and 50% among Pima Indians in Arizona (131). A 2004–2006 U.S. national survey including people aged 20 years or older indicated
that 11.8% of African-Americans, 10.4% of Hispanics-Americans, 7.5% of Asian-Americans, and 6.6% of non-Hispanic whites Americans had clinically manifest diabetes. Among Hispanics, rates were 12.6% for Puerto Rican Americans, 11.9% for Mexican Americans, and 8.2% for Cuban Americans (132). All these findings are age-adjusted. Ethnic variability can be partially attributed to non-genetic environmental and cultural factors. However, some studies show that diabetes prevalence differs markedly across ethnic groups, even when environmental exposures are similar. For example, Asians living in the UK have a prevalence of diabetes 3.8-fold higher than that in whites in the UK (133).

According to the WHO, between 2000 and 2030, Asia and Africa are likely to experience a 2–3-fold increased prevalence of T2D (32). The organisation further predicts that the most new cases of T2D will emerge from India, China, and the USA, partly because these countries have some of the world’s largest populations, but also because these are ethnically at risk populations that are rapidly adopting obesogenic lifestyles. According to the WHO, Bangladesh, Brazil, Indonesia, Japan, and Pakistan will also be heavily burdened by T2D in the future.

1.5.2 Genetics

In 2001, the draft sequence of the human nuclear genome was published (134, 135). The human genome consists of approximately 2.85 billion base pairs encoding about 20,000–25,000 genes (136). Although 99.9% of the human DNA sequence is thought to be identical between unrelated individuals, about 0.1% of coded DNA differs between the two chromosomal strands at the same base (137). It is these differences that account for the diversity in human phenotypes and their responsiveness to environmental exposures including, but by no means limited to, diet and physical activity.

DNA variation occurs in several known forms. Sequence variations occurring less frequently than in 1% of the population are often defined as mutations, whereas more common variants are defined as polymorphisms (138). Single nucleotide polymorphisms (SNPs) are the most commonly studied form of genetic variant. Other well known polymorphisms are variable number of tandem repeats (VNTRs) that include mini-satellites (repeat sequences of several nucleotides) that tend to cluster near ends of chromosomes and microsatellites (usually as di-, tri-, and tetra-nucleotide repeats) that are distributed throughout the genome (138), inversions, insertions, deletions, copy number variants (CNVs), and other complex rearrangements (139).

Polymorphisms are thought to occur because of selective pressures or randomly (genetic drift). Random variations eventually disappear as the contributing alleles become either fixed or extinct. Irrespective of the specific type of polymorphism, there is a plethora of common DNA sequence variants that can (and have) been used as disease predisposing markers in human population-based studies. Although many bona fide examples of disease-associated polymorphisms have been identified, by and large, these are non-functional variants, which merely ‘tag’ the (often) unobserved causal variant.

Approximately 12 million SNPs have been identified (140). More than 90% of the genomic variability between individuals is thought to be attributable to SNPs (141). The majority of these SNPs are biallelic and can be transition (purine-purine A↔G or pyrimidine-pyrimidine C↔T) or transversion (purine-pyrimidine or pyrimidine-purine) substitutions (142). Approximately two
thirds of SNPs are transition substitutions between double ring structured purines (A↔G) or between single ring structured pyrimidines (T↔C), whereas one third of SNPs are transversion substitutions between a purine and a pyrimidine nucleotide (143). The classification of SNPs is dependent on their genomic location. Coding SNPs (cSNP) are located in exons (a segment of a gene that is represented in the mature RNA product). Individual exons may contain coding DNA and/or non-coding DNA ([untranslated sequences]) and may be either synonymous or non-synonymous (144). Synonymous SNPs are typically silent and alter the DNA sequence, but do not change the aminoacid coding sequence. Non-synonymous cSNPs alter the DNA sequence in a coding region such that the aminoacid coding sequence of the protein is changed. These cSNPs are prioritized as genetic markers because a change in the aminoacid structure and function may impact the formation of the target protein. The majority of SNPs are located in the non-coding region of the genome (138). However, some of these intronic SNPs (an intron is a segment of DNA that is not represented in the mature RNA) have no known function but may play a regulatory role in modulating gene expression of coding regions. These SNPs are termed regulatory, or rSNPs. rSNPs located in the promoter region may affect transcription factor sites and rSNPs located in the 5'UTR and 3'UTR (untranslated regions) may also affect protein-binding sites by changing sequence motifs. rSNPs may still have consequences for gene splicing, transcription factor binding, or the sequence of non-coding RNA. Intrinsic SNPs and intergenic SNPs (regions between genes) lie in the non-coding regions. It is general thought that non-synonymous SNPs in a coding sequence are more likely to affect the function or availability of a protein than other SNP classes (145). However, all SNP types can cause disease, for example by altering the regulation of transcription of a critical protein. The true distribution of disease-associated variants between non-coding and coding sequences is unknown (145).

There are at least six established ways in which SNP genotyping might help advance our understanding of the molecular basis to human disease. These include: 1) hypothesis-free gene discovery and mapping; 2) association-based candidate polymorphism testing; 3) diagnostics and risk profiling; 4) prediction of response to environmental stimuli; 5) pharmacogenetics; and 6) homogeneity testing and epidemiological study design (141, 146).

1.6 Genetic association studies of T2D

Over the last few years, genome-wide association studies (GWAS) have been extremely successful in the detection of loci for complex disease traits such as obesity and T2D. The GWAS method involves testing associations with disease traits for a large number of genetic markers (usually more than 1,000,000 SNPs) over the whole genome. The method differs from the traditional biologic candidate gene approach, in that no specific hypothesis is tested; the approach instead relies heavily on replication of association signals across multiple populations and generally requires very large sample sizes to overcome the problems (related mainly to diminished power) inherent in conducting so many association tests (147). GWASs have confirmed the three previously identified signals for T2D which localize to TCF7L2 (148, 149), PPARG (150) and KCNJ11 (150) and identified many new susceptibility loci (149-154). Table 1 shows the T2D loci that have been discovered and replicated to date, most of which
localize to genes that appear to influence beta-cell function (151, 155, 156). This finding highlights the relative importance of inherited defects in insulin secretion on beta-cells rather than insulin resistance in the aetiology of T2D (157, 158).

Table 1 T2D-susceptibility loci for which there is genome-wide significant evidence for association

<table>
<thead>
<tr>
<th>Locus (nearest genes)</th>
<th>Year association ‘proven’</th>
<th>Approach</th>
<th>Probable mechanism</th>
<th>Index variant</th>
<th>Effect sizea</th>
<th>Risk-allele frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPARG</td>
<td>2000</td>
<td>Candidate</td>
<td>Insulin action</td>
<td>rs1801382</td>
<td>1.14</td>
<td>0.87</td>
</tr>
<tr>
<td>KCNJ11</td>
<td>2003</td>
<td>Candidate</td>
<td>β-cell dysfunction</td>
<td>rs6215</td>
<td>1.14</td>
<td>0.35</td>
</tr>
<tr>
<td>TCF7L2</td>
<td>2006</td>
<td>Large-scale association</td>
<td>β-cell dysfunction</td>
<td>rs7901385</td>
<td>1.37</td>
<td>0.31</td>
</tr>
<tr>
<td>FTO</td>
<td>2007</td>
<td>GWA</td>
<td>Altered BMI</td>
<td>rs250126</td>
<td>1.17</td>
<td>0.49</td>
</tr>
<tr>
<td>HHEX/IDE</td>
<td>2007</td>
<td>GWA</td>
<td>β-cell dysfunction</td>
<td>rs1111876</td>
<td>1.15</td>
<td>0.65</td>
</tr>
<tr>
<td>SLC30A8</td>
<td>2007</td>
<td>GWA</td>
<td>β-cell dysfunction</td>
<td>rs1326634</td>
<td>1.15</td>
<td>0.69</td>
</tr>
<tr>
<td>CDKAL1</td>
<td>2007</td>
<td>GWA</td>
<td>β-cell dysfunction</td>
<td>rs1094290</td>
<td>1.14</td>
<td>0.32</td>
</tr>
<tr>
<td>CDKN2A/2B</td>
<td>2007</td>
<td>GWA</td>
<td>β-cell dysfunction</td>
<td>rs1811961</td>
<td>1.20</td>
<td>0.83</td>
</tr>
<tr>
<td>HNF1B</td>
<td>2007</td>
<td>GWA</td>
<td>β-cell dysfunction</td>
<td>rs4403950</td>
<td>1.14</td>
<td>0.32</td>
</tr>
<tr>
<td>HHEX</td>
<td>2007</td>
<td>Large-scale association</td>
<td>β-cell dysfunction</td>
<td>rs4307956</td>
<td>1.10</td>
<td>0.47</td>
</tr>
<tr>
<td>WFS1</td>
<td>2007</td>
<td>Large-scale association</td>
<td>Unknown</td>
<td>rs10010731</td>
<td>1.12</td>
<td>0.69</td>
</tr>
<tr>
<td>JAZF1</td>
<td>2008</td>
<td>GWA</td>
<td>β-cell dysfunction</td>
<td>rs654846</td>
<td>1.10</td>
<td>0.59</td>
</tr>
<tr>
<td>CDC123/CAMK1D</td>
<td>2008</td>
<td>GWA</td>
<td>Unknown</td>
<td>rs4873790</td>
<td>1.11</td>
<td>0.48</td>
</tr>
<tr>
<td>TSPAN8/8/LGR5</td>
<td>2008</td>
<td>GWA</td>
<td>Unknown</td>
<td>rs765181</td>
<td>1.09</td>
<td>0.27</td>
</tr>
<tr>
<td>THADA</td>
<td>2008</td>
<td>GWA</td>
<td>Unknown</td>
<td>rs357387</td>
<td>1.15</td>
<td>0.80</td>
</tr>
<tr>
<td>ADAMTS9</td>
<td>2008</td>
<td>GWA</td>
<td>Unknown</td>
<td>rs4607103</td>
<td>1.09</td>
<td>0.76</td>
</tr>
<tr>
<td>NOTCH2</td>
<td>2008</td>
<td>GWA</td>
<td>Unknown</td>
<td>rs10803231</td>
<td>1.13</td>
<td>0.10</td>
</tr>
<tr>
<td>KCNQ1</td>
<td>2008</td>
<td>GWA</td>
<td>β-cell dysfunction</td>
<td>rs22237182</td>
<td>1.29</td>
<td>0.93</td>
</tr>
</tbody>
</table>

aAbbreviations: ADAMTS9, ADAM metallopeptidase with thrombospondin type 1 motif 9; CAMK1D, calcium/calmodulin-dependent protein kinase 1D; CDC123, cell division cycle 123 homologue (Saccharomyces cerevisiae); CDKAL1, CDK5 regulatory subunit-associated protein-like1; CDKN2B, cyclin-dependent kinase inhibitor 2B; FTO, fat mass and obesity associated; HHEX, haematopoietically expressed homeobox; HNF1B, hepatocyte nuclear factor 1 homeobox B; IDE, insulin degrading enzyme; IGFBP2, insulin-like growth factor 2 mRNA binding protein 2; JAZF1, juxtaposed with another zinc finger gene 1; KCNJ11, potassium inwardly rectifying channel, subfamily J, member 11; KCNQ1, potassium voltage-gated channel, KQT-like subfamily, member 1; LGR5, leucine-rich repeat-containing G-protein coupled; NOTCH2, Notch homologue 2 (Drosophila); PPARG, peroxisome proliferator-activated receptor gamma; SLC30A8, solute carrier family 30 (zinc transporter), member 8; TCF7L2, transcription factor 7 like 2; THADA, thyroid adenoma associated; TSPAN8, tetraspanin 8; WFS1, Wolfram syndrome.

bEstimates of effect size (given as per-allele odds ratios, i.e. the increase in odds of diabetes per copy of the risk allele) and risk-allele frequencies are all reported for European descent populations based on available data (157).

Figure 1 shows the effect sizes of known T2D-susceptibility loci from European populations. The TCF7L2 variant yields the largest effect on diabetes risk with a per-allele odds ratio of ~1.4 (157). These results show that individually, each variant confers only a small risk of developing T2D.
Figure 1 Effect sizes of known T2D-susceptibility loci. The T2D-susceptibility variants so far discovered have only modest individual effects. The X-axis gives the per-allele odds ratio (estimated for European-descent samples) for each locus listed on the y-axis. Loci are sorted by descending order of per-allele effect size from TCF7L2 (1.37) to ADAMTS9 (1.09) (Table 1). Loci shown in blue are those identified by GWA approaches, whereas those found by candidate-gene approaches and by large-scale association analyses are shown in yellow and red, respectively. Odds ratios are estimated from data in Refs (149, 150, 152-154, 159-163). These figures are approximate estimates of the true effect sizes at each locus and might be either overestimates (owing to winner’s curse) or underestimates (because the causal variant in most cases is not known yet) (157).

Several T2D-susceptibility variants are also associated with other traits. For example, the variants at HNF1B (TCF2) and JAZF1 are associated with susceptibility to prostate cancer (164, 165), the CDKAL1 variant has been associated with Crohn’s disease (166). A GCKR variant has also been associated with higher triglyceride levels and lower glucose levels (167, 168). Several of these T2D-susceptibility variants are also implicated in monogenic diseases such as PPARG (169), KCNJ11 (170), HNF1B (171), and WFS1 (172).

Recently, common variants in the KCNQ1 (159, 160) and MTNR1B (173, 174) genes have been added to this list, with many more in the pipeline (personal communication, Mark McCarthy, Oxford).

1.7 Overview of literature of genes of interest

1.7.1 NOTCH2

A recent meta-analysis of genome-wide scans (FUSION group, Wellcome Trust Case Control Consortium and, the Diabetes Genetics Initiative) and large scale replication has reported T2D susceptibility loci marked by the intronic rs10923931 of the NOTCH2 gene (OR [95%CI]: 1.13, P=4.1×10\(^{-8}\)) (25). NOTCH2 is located at chromosome position 1p13-p11 and is a transmembrane receptor involved in pancreatic organogenesis (175).
1.7.2 THADA

**THADA** is located at chromosome position 2p21, where a non-synonymous rs7578597 (T1187A) SNP implicated in T2D (OR [95%CI]: 1.15, \( P = 1.1 \times 10^{-9} \)) was identified by a recent meta-analysis of genome-wide scans and large scale replication (25). There is evidence that loss of function of **THADA** contribute to the development of the follicular neoplasias of the thyroid (176).

1.7.3 ADAMTS9

The rs4607103 SNP, located at chromosome position 3p14.1, ~38 kb upstream of the **ADAMTS9** gene has shown association with T2D susceptibility (OR [95%CI]: 1.09, \( P = 1.2 \times 10^{-8} \)) and was identified by a meta-analysis of genome-wide scans and large scale replication (25). **ADAMTS9** encodes a member of a large family of 19 metalloproteases that is involved in maturation of precursor proteins, extracellular matrix remodeling, cell migration and inhibition of angiogenesis (177, Porter, 2005 #428). **ADAMTS9** is widely expressed in skeletal muscle and pancreas (25).

1.7.4 PPARG

Multiple studies have reported T2D associations for the non-synonymous Pro12Ala (rs1801282) polymorphism located at chromosome position 3p25 of the gene **PPARG** (150, 178-181). In a meta-analysis the odds ratio per copy of the risk allele of the Pro12Ala variant (rs1801282) for T2D was 1.19 (\( P = 1.19 \times 10^{-7} \)) (182). **PPARG** is associated with impaired insulin sensitivity (178). This transcription factor is involved in adipocyte development (183).

1.7.5 IGF2BP2

Multiple GWAS have reported T2D susceptibility on the intronic rs4402960 SNP of the **IGF2BP2** gene located at chromosome position 3q27.2 (150, 152, 154). Subsequently, other studies have confirmed these results in different populations (151, 180, 181, 184-188). The odds ratio per copy of the risk allele for the intronic rs4402960 variant for T2D is 1.14 (\( P = 8.9 \times 10^{-16} \)) (189). The signal of **IGF2BP2** exerts its primary effect on insulin secretion (190, 191). IGF2BP is a paralog of **IGF2BP1**, which binds to the 5’ untranslated region of the insulin-like growth factor 2 (**IGF2**) mRNA and regulates **IGF2** translation. **IGF2** is involved in the development, growth, and stimulation of in insulin action (192).

1.7.6 WFS1

**WFS1** is located at chromosome position 4p16.1 and has been implicated in T2D susceptibility (161). Replication studies have confirmed this association (162). The odds ratio per copy of the risk allele for the rs10010131 variant for T2D was 1.15 (\( P = 4.5 \times 10^{-5} \)) (161). There is evidence that functions of **WFS1** include the regulation of membrane trafficking, protein processing and homeostasis in the endoplasmic reticulum of pancreatic beta-cells (193, 194). Disruption of these processes may lead to progressive pancreatic beta-cell loss and neuronal degeneration observed in Wolfram syndrome (195).
1.7.7 **CDKAL1**

Evidence from GWAS implicate *CDKAL1*, located at chromosome position 6p22.2, in T2D predisposition (149, 150, 152, 154). Replication studies from different ethnic backgrounds have confirmed this association (181, 184-188). The odds ratio per copy of the risk allele for the intronic rs7754840 variant for T2D was 1.12 \((P=4.1 \times 10^{-11})\) (189). *CDKAL1* is a gene of unknown function. However, this gene shares homology at protein domain level with CDK5 regulatory subunit associated protein 1 (CDK5RAP1), which has been implicated in the loss of beta cell function under glucotoxic conditions (196).

1.7.8 **JAZF1**

The rs864745 intron 1 SNP, located at chromosome position 7p15.2-p15.1 of the *JAZF1* gene was implicated in T2D (OR [95%CI]: 1.10[1.07-1.13], \(P=5.0 \times 10^{-14}\) under an additive model) and was identified by a recent meta-analysis of genome-wide scans and large scale replication (25). The risk allele in the *JAZF1* is associated with insulin release suggesting the contribution of abnormal pancreatic beta-cell function (156). JAZF1 is a transcriptional repressor and contributes to neoplastic phenotypes (197).

1.7.9 **SLC30A8**

Evidence from GWAS implicate *SLC30A8*, located at chromosome position 8q24.11, in T2D predisposition (149, 150, 152-154). Replication studies from different ethnic backgrounds have confirmed this association (184, 186, 188, 198). The odds ratio per copy of the risk allele for the nonsynonymous rs13266634 variant for T2D was 1.12 \((P=5.3 \times 10^{-8})\) (189). *SLC30A8* encodes a zinc transporter expressed solely in the secretory vesicles of beta-cells and is thus implicated in the final stages of insulin biosynthesis, which involve co-crystallization with zinc (189). Overexpression of *SLC30A8* in insulinoma (INS-1E) cells enhanced glucose-induced insulin secretion (189).

1.7.10 **CDKN2A/B**

Evidence from GWAS implicate *CDKN2A/B*, located at chromosome position 9p21, in T2D predisposition (150, 153, 154, 199). Replication studies (151) but not all (181) have confirmed this association in different ethnic groups. The odds ratio per copy of the risk allele for the rs10811661 variant for T2D was 1.2 \((P=8.8 \times 10^{-15})\) (189). The *CDKN2A* and *CDKN2B* genes are both tumour suppressors, working via inhibition of CDK kinases (200) and are highly expressed in adipocytes and pancreatic islets (152).

1.7.11 **CAMK1D**

The rs12779790 (intergenic region), located at chromosome position 10p13 of the *CAMK1D* gene was implicated in T2D (OR [95%CI]: 1.11 [1.07-1.14], \(P=1.2 \times 10^{-10}\)) and was identified by a recent meta-analysis of genome-wide scans and large scale replication. (25) This study also suggests that *CAMK1D* gene
may be involved through a T2D pathogenetic mechanism with cell cycle dysregulation.

1.7.12 *HHEX*

Evidence from GWAS implicates *HHEX*, located at chromosome position 10q23.33, in T2D predisposition (153). Replication studies from different ethnic backgrounds have confirmed this association (151). The odds ratio per copy of the risk allele for the nonsynonymous rs1111875 variant (7.7 kb downstream) for T2D was 1.13 \((P=5.7 \times 10^{-10})\) (189). *HHEX* is essential for hepatic and pancreatic development (201) and is a target of the Wnt signalling pathway (202).

1.7.13 *TCF7L2*

*TCF7L2* is located at chromosome position 10q25.3 and confers the strongest effect on T2D risk European populations (150, 154, 203). These results have subsequently been confirmed in multiple ethnic groups (180, 198). The odds ratio per copy of the risk allele for the intronic rs7903146 variant for T2D was 1.37 \((95\% CI: P=1.0 \times 10^{-48})\) (204). The precise mechanisms underlying the increased risk are poorly understood and that had no ‘track-record’ as a candidate for T2D. TCF proteins are transcription factors that affect cell proliferation and differentiation via the Wnt signalling pathway.

1.7.14 *KCNJ11*

Although initial smaller studies failed to replicate the association of the E23K polymorphism *KCNJ11* gene, located at chromosome position 11p15.1, with T2D, large scale studies and meta-analyses have consistently associated the lysine variant with T2D (205-210). The odds ratio per copy of the risk allele for the missense E23K rs5219 variant for T2D was 1.14 \((P=6.7 \times 10^{-11})\) (211). *KCNJ11* encodes the beta-cell potassium channel and is crucial to the regulation of glucose-induced insulin secretion in pancreatic beta cells (170).

1.7.15 *MTNR1B*

Recently, GWAS revealed that one of the strongest signals for glucose-stimulated insulin secretion emanated from MTNR1B located at chromosome position 11q21-q22 \((P= 7 \times 10^{-4}, \text{rank order 595})\) (173). The rs10830963 SNP was strongly associated \((P=3.2 \times 10^{-50})\) with elevated fasting glucose concentrations in a meta-analysis of the recent GWAS of T2D (174) and subsequent studies confirmed that this intronic variant was tightly associated with FPG and T2D risk (173, 212-216). Melatonin is an indoleamine formed from tryptophan via acetylation and it has primarily implicated in the regulation of circadian rhythms (217). *MTNR1B* mRNA is expressed in human islets, and is primarily localized in beta-cells in islets (173). Insulin release from clonal beta-cells in response to glucose was inhibited in the presence of melatonin suggesting that blocking the melatonin ligand-receptor system could be a way of therapy for T2D (1).
1.7.16 TSPAN8

TSPAN8 is located at chromosome position 12q14.1-q21.1. An intronic variant (rs7961581) has been implicated in T2D (OR [95%CI]: 1.09[1.06-1.12], \( P=1.1\times10^{-9} \)) and was identified by the previous recent meta-analysis of genome-wide scans and large scale replication (25). TSPAN8 is expressed in different carcinomas (e.g. colon, liver and pancreas) (218).

1.7.17 FTO

GWAS revealed that FTO located at chromosome position 16q12.2 exerts its primary effect on T2D risk through excess adipose accumulation (219). Each copy of the FTO risk allele, for which ~40% of whites are heterozygous and 16% are homozygous, corresponds to ~1 kg heavier weight. These results have been confirmed in multiple studies (150, 152, 180, 181, 187, 198, 220, 221). The odds ratio per copy of the risk allele for the intronic rs9939609 variant for T2D was 1.27 (\( P=5 \times 10^{-8} \)) (189). The function of the FTO protein is unknown but gene expression profiles show that FTO is expressed in several tissues especially parts of the brain and in muscle (219, 220).

1.7.18 HNF1B

HNF1B is located at chromosome position 17q12. Variation at this locus has been associated with MODY and T2D (163). The odds ratio per copy of the minor allele for the rs4430796 variant (intron 2) for T2D was 0.91 (95%CI: \( P=2.7 \times 10^{-7} \)) (165). HNF1B (also known as TCF2) encodes a homeobox transcription factor that is implicated in cell proliferation and differentiation in the kidney, pancreas, liver, and genital tract tissues.

1.7.19 PPARGC1A

The non-synonymous Gly482Ser variant (rs8192678) of the PPARGC1A gene, located at chromosome position 4p15.1 has showed T2D susceptibility in a meta-analysis (1.11, \( P=0.004 \)) (222). The protein encoded by the PPARGC1A gene co-activates as many as 30 transcription factors involved in thermogenesis, mitochondrial biogenesis, glucose/fatty acid metabolism, adipogenesis, regulation of reactive oxygen species and hepatic glucose production and is expressed predominant in tissues with high metabolic activity, such as brown adipose tissue, heart, kidney and exercising skeletal muscle (223). PPARGC1A is implicated in a marked reduction of skeletal muscle to generate mitochondria and oxidize fat that subsequently influence the development of insulin resistance, and T2D (224-226).

1.8 What is a gene x lifestyle interaction?

For the purpose of this section I will define gene x environment interactions only in the linear form, as this is the most commonly studied type of interaction effect. In epidemiology, the term interaction implies a mutual dependency of two or more risk factors contributing to disease risk (227). The use of the word interaction in biology sometimes differs from the epidemiological use of the word; thus, I will distinguish here between the two definitions.
In biology, interactions are defined as the interdependent operation of two or more factors that act together through the same causal biological mechanism (228, 229). This is somewhat similar to the concept of effect-mediation in epidemiology. An interaction in epidemiology is often defined as the interdependence of two or more exposure (independent) variables, where the effect of one exposure depends on the level of another (228). Thus, although statistical and biologic definitions of interactions have important distinctions, epidemiological models of interaction can sometimes be used to test biologically derived hypotheses of interaction and vice versa.

Gene x environment interactions occur when the magnitude of the effect of the environmental exposure on a disease trait differs in magnitude across genotypes at a given locus or set of loci. Different types of interaction exist. Typically these can be divided into two basic classes termed removable (or ‘quantitative’) interactions and non-removable (or ‘qualitative’) interactions. The former is dependent on the scale used to express the phenotype (i.e. additive or multiplicative scales), whereas the latter is not (228). In some cases a gene x environment interaction is apparent when the data are expressed on the log scale, but can be removed by expressing the same data on the normal scale. The design and purpose of the study influences the interpretation of interaction effects and the choice of whether to accept or refute an interaction that is evident on one scale but not another, assuming that is that the choice of scale does not violate the fundamental assumptions of the statistical model (227). Departure from an additive scale may be evaluated by risk difference (RD) which corresponds to how much risk of disease is added in subjects with the risk factor compared to those without the risk factor. Interactions expressed on multiplicative scales can be evaluated by relative risks or odds ratios. In these examples, the magnitude of the association between the first exposure variable (e.g. the genotype) and the disease trait is compared for individuals in discordant environmental exposure settings. The formal statistical approach to test interaction usually involves fitting a product term (gene x environment) in addition to the marginal effects to a regression model, where the phenotype is fitted as the dependent variable.
Several variations on these basic methods have been proposed (231), including the case only design. Using this approach a multiplicative model can be tested. A limitation is that it does not allow for the estimation of the separate effects of high-risk environments or genotypes (232). Where the case only design is applied in family pedigrees (233), the log-linear approach (234) and the transmission disequilibrium test can be used to test for interaction (235). Alternatively, a joint test of the genetic main effect and gene x environment interaction effects (also known as the two-degree of freedom model) is an appealing procedure as it is generally more powerful than conventional approaches (236).

Understanding the nature of gene x environment interactions may help improve power of genetic prediction models. As shown in figure 3, where linear interaction effects occur, genetic risk varies across the spectrum of the environmental exposure. Selecting individuals from the extreme of the environmental distribution where genetic effects are of greatest magnitude would enhance the statistical power of genetic prediction models (237).
Confounding in genetic association studies is generally less of a problem than in studies evaluating disease associations for non-genetic exposures. This is because genotypes are fixed throughout life and are unaffected by the level of other factors such as smoking, alcohol consumption, income, or social status which in non-genetic association studies may cause confounding. Nevertheless, at a population level age (through survival bias) and ethnicity (through population stratification) can cause confounding in genetic association studies. The latter can occur when allele frequencies and disease frequencies coincidently vary between ethnic subgroups, which when pooled for genetic association studies can lead to spurious conclusions about associations unless the models are appropriately controlled for ethnicity. It is important to bear in mind though that even if genetic association studies are resilient to confounding, studies of gene x environment interactions are not, as they simultaneously inherit the limitations and susceptibilities that beset studies of genetic AND non-genetic association studies. Moreover, confounding by parallel interaction effects can also occur. Therefore, it may be necessary to include a range of potential confounders in interaction models, in addition to product terms comprised of the genotype of interest and putative environmental confounding variables.

1.9 Literature review on studies of gene x lifestyle interactions

The purpose of the following section is to review the published literature on gene x physical activity interactions on diabetes-related traits. Observational and interventional studies are described. Most of the genes described in the studies outlined below were selected as biologic candidate genes, with a small number involving gene variants identified through GWAS.
1.9.1 Observational studies

Observational studies examining gene x physical activity interactions on obesity, T2D, and related traits are shown in Table 2a.

Several early studies of gene x physical activity interactions focused on variants in the adrenergic beta-2 receptor (ADRB2). The initial study was undertaken in French males and focused on obesity as the outcome (BMI and waist circumference). In that study, Meirhaeghe et al reported a strong association between the Gln27Glu variant and BMI and waist circumference in physically inactive men, but this effect was completely diminished in active men, which gave rise to a statistically significant gene x physical activity interaction. In a follow-up study conducted in UK white adults, Meirhaeghe et al reported gene x physical activity interactions for a different ADRB2 variant (Gly16Arg) on obesity and NEFA levels. This was the first gene x environment interaction study to use objective assessments of physical activity in a population-based cohort. The Gly16Arg variant was inversely related with fasting and post-glucose challenge NEFA levels in the population as a whole. However, the magnitude of this association was greater in people with the Arg16Arg genotype compared with carriers of the Gly16 allele. No evidence of interactions on obesity-related traits was reported. Elsewhere, Corbalan et al (239) found that recreational physical activity (ratio between MET-hours per week to time spent sitting down during leisure time) modulates the effect of the ADRB2 Gln27Glu polymorphism on the risk of obesity in Spanish women.

One of the most widely studied genes in the context of gene x lifestyle interactions on obesity and T2D is the peroxisome proliferator-activated receptor gamma gene (PPARG). Several studies have reported interactions between dietary fat intake and the PPARG Pro12Ala variant on insulin resistance and obesity (240-244). A number of subsequent studies examined the interaction between Pro12Ala and physical activity on these traits. For example, in a cross-sectional study of UK Caucasians, Franks et al. (245) reported that the association between physical activity and dietary fat composition (polyunsaturated to saturated fatty acid ratio) on fasting insulin levels is additive in persons with the Pro12Pro genotype, but is synergistic in carriers of at least one copy of the Ala12 allele. Elsewhere, in a cross-sectional study of 216 Hispanic pedigrees (1,850 nuclear families) and 236 non-Hispanic white (NHW) pedigrees (1,240 families), Nelson et al. (246) found a significant interaction between the Pro12 allele and physical activity on T2D risk. The Pro12 allele was only associated with T2D risk in physically inactive individuals, with active individuals carrying the Pro12 allele being at a similar level of risk as Ala12 allele carriers.

A third class of biologic candidate genes which has been widely studied for interactions with physical activity are the genes encoding uncoupling proteins (UCP-2 and UCP-3). Two studies examining the UCP-2 gene found no evidence of interactions with physical activity on obesity-related traits (247, 248). A further two studies focussing on the UCP-3 gene found no evidence of gene x physical activity interactions on obesity-related traits (248, 249).
<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Study Subjects</th>
<th>Lifestyle exposure</th>
<th>Gene</th>
<th>Polymorphism</th>
<th>Outcome</th>
<th>Int P-value</th>
<th>Main results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Franks et al. 2004</td>
<td>UK Caucasians</td>
<td>506 (226 men/281 women)</td>
<td>Free-living heart rate monitoring</td>
<td>PPARG</td>
<td>rs1801282 (Pro12Ala)</td>
<td>Fasting insulin</td>
<td>0.038</td>
<td>In carriers of the Ala allele there was an interaction between PA and P:S ratio on fasting insulin ($P_{interaction}=0.038$). In carriers of the Pro allele homozygotes no interaction was observed between PA and P:S ratio on fasting insulin ($P_{interaction}=0.929$).</td>
</tr>
<tr>
<td>Nelson et al. 2007</td>
<td>USA</td>
<td>216 Hispanic pedigrees (1850 nuclear families)</td>
<td>Questionnaire</td>
<td>PPARG</td>
<td>rs1801282 (Pro12Ala)</td>
<td>T2D</td>
<td>0.022</td>
<td>In non-Hispanic whites a significant interaction between low PA and the Pro12 allele on T2DM was observed ($P_{interaction}=0.022$). After stratifying by PA tertiles the Pro12 allele was significantly associated with T2DM in those with low PA.</td>
</tr>
<tr>
<td>Meirhaeghe et al. 2001</td>
<td>UK Caucasians</td>
<td>604</td>
<td>Objective measures of total energy expenditure</td>
<td>ADRB2</td>
<td>rs1042713 (Gly16Arg)</td>
<td>NEFA levels</td>
<td>0.05</td>
<td>The effect of the Arg16Arg genotype on the suppression of NEFA levels was modified by PA level ($P_{interaction}&lt;0.05$).</td>
</tr>
<tr>
<td>Corbalan et al. 2002</td>
<td>Spain</td>
<td>252 women</td>
<td>Questionnaire and interview</td>
<td>ADRB2</td>
<td>rs1042714 (Gln27Glu)</td>
<td>Obesity</td>
<td>0.005</td>
<td>A statistically significant interaction between recreational energy expenditure and the 27Glu allele on the risk of obesity was demonstrated ($P_{interaction}=0.005$).</td>
</tr>
<tr>
<td>Otabe et al. 2000</td>
<td>French Caucasians</td>
<td>68 morbidly obese and 120 controls of unrelated</td>
<td>Indirect calorimetry</td>
<td>UCP2</td>
<td>Exon 1</td>
<td>obesity</td>
<td>NI</td>
<td>These results failed to give evidence of a relationship between resting metabolic rate and polymorphisms in the UCP2 gene on obesity.</td>
</tr>
</tbody>
</table>
Changes in BMI

No interaction between the UCP2-genotype and PA in relation to change in BMI was found in the juvenile obese or control group. No interactions between PA and UCP2-genotype in relation to waist circumference, body fat mass index or obesity.

Changes in BMI

No interaction between the UCP3-genotype and PA in relation to change in BMI was found in the juvenile obese or control group. No interactions between PA and UCP3-genotype in relation to waist circumference, body fat mass index or obesity.

An interaction between the G-250A and PA on serum HDL-C ($P_{interaction}=0.002$) was observed. Vigorous physically active homozygous A-allele carriers had a 0.30 mmol/litre (95% CI 0.22–0.37) increase in HDL-C compared with homozygous G-allele carriers, whereas no major effect of an active lifestyle was seen comparing heterozygous carriers with homozygous G-allele carriers.

A significant increase risk of CHD for LIPC-480 TT was observed for subjects who did not participate in vigorous PA
Vigorous PA did not significantly alter the risk of CHD for subjects with the LIPC-480 CC or CT genotype ($P=0.62$); however, a reduced risk due to vigorous PA was found for subjects with the LIPC-480 TT genotype ($P=0.03$ comparing vigorous PA with sedentary or moderate PA among LIPC-480 TT subjects).

### Study Details

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Participants</th>
<th>Age (yr)</th>
<th>APOE Genotypes</th>
<th>HDL-C</th>
<th>Other Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prospective cohort</td>
<td>(14-y follow-up)</td>
<td>non-Hispanic whites</td>
<td>$51.0 \pm 12.6$</td>
<td></td>
<td></td>
<td>(Significantly higher compared to vigorous PA)</td>
</tr>
<tr>
<td>Corella et al. 2001</td>
<td>Cross-sectional study</td>
<td>Spanish Caucasians</td>
<td>909 (396 men/513 women)</td>
<td>Self-administered questionnaire</td>
<td>APOE ε2 isoform</td>
<td>HDL-C</td>
</tr>
<tr>
<td>Bernstein et al. 2002</td>
<td>Cross-sectional study</td>
<td>Switzerland</td>
<td>1,708 (846 men/862 women)</td>
<td>Self-administered questionnaire</td>
<td>APOE ε2 isoform</td>
<td>HDL-C</td>
</tr>
<tr>
<td>Boer et al. 1999</td>
<td>Cross-sectional study</td>
<td>Dutch</td>
<td>379 (167 men/212 women)</td>
<td>Questionnaire</td>
<td>LPL D9N mutation</td>
<td>Total cholesterol: HDL-C</td>
</tr>
<tr>
<td>Sentí et al. 2001</td>
<td>Cross-sectional study</td>
<td>Spain</td>
<td>520 men</td>
<td>Questionnaire</td>
<td>LPL HindIII</td>
<td>HDL-C</td>
</tr>
</tbody>
</table>
### Grove et al. 2007

**USA**

*3,728 African Americans and 10,988 whites*

**Questionnaire**

**GNB3 rs5443 (825C>T)**

**Obesity**

**Hypertension**

| **A significant interaction** (*P*<0.001) between 825C>T and PA in predicting obesity status in African Americans (AAs) was observed. In AAs who were active, each 825T allele was associated with a 20% lower prevalence of obesity (*P*=0.005), whereas each 825T allele was associated with a 23% greater prevalence of obesity for low-active individuals (*P*=0.008). There was no significant interaction between 825C>T and PA in predicting hypertension. However, an interaction between the 825C>T, obesity status and PA in predicting hypertension in the AA subjects was observed (*P*=0.018). AA homozygotes for the 825T allele who were both obese and had a low activity level were 2.7 times more likely to be hypertensive, compared to non-obese, active 825C homozygotes (*P*<0.02). In Whites there was no significant interactions between this variant and PA in relation to obesity or hypertension. |

### Franks et al. 2005

**UK Caucasians**

*706 UK (309 men/397 women)*

*53.4±10.7 Healthy nondiabetic participants (MRC Ely Study only)*

**Objective measures of total energy expenditure**

**NOS3**

1800783

1800779

2070744

IVS6+26

E298D

IVS11-30

IVS25+15

rs3800787

| **An interaction was observed between total energy expenditure and NOS3 haplotypes (rs2070744, IVS11-30, and rs3800787) on 2-h glucose (*P*=0.007). The total energy expenditure was inversely related with glucose intolerance in the 3 haplotype groups (*P*<0.0001).** |

### Vimaleswaran et al. 2008

**UK Caucasians**

*726 (298 men/428 women)*

*55±10 yr Healthy nondiabetic participants*

**Habitual/non-resting energy expenditure**

**NOS3**

rs10277237

rs1800779

rs2070744

<p>| <strong>The interaction between non-resting energy expenditure (NREE) and IVS25+15 was significant for both DBP (<em>P</em>=0.006) and SBP (<em>P</em>=0.026), in such a way that the effect of the GG-genotype on blood pressure was stronger in</strong> |</p>
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Sample Size</th>
<th>Methodology</th>
<th>Gene</th>
<th>SNP</th>
<th>Trait</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andreasen et al. 2008</td>
<td>Denmark</td>
<td>5722</td>
<td>Cross-sectional study</td>
<td>FTO</td>
<td>rs9939609</td>
<td>BMI</td>
<td>0.007</td>
</tr>
<tr>
<td>Rampersaud et al. 2008</td>
<td>USA</td>
<td>704</td>
<td>Cross-sectional study</td>
<td>FTO</td>
<td>rs1861868</td>
<td>BMI</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rs1477196</td>
<td></td>
<td>BMI</td>
<td>0.004</td>
</tr>
<tr>
<td>Andreasen et al. 2008</td>
<td>Denmark</td>
<td>5604</td>
<td>Cross-sectional study</td>
<td>INSIG2</td>
<td>rs7566605</td>
<td>BMI</td>
<td>0.004</td>
</tr>
<tr>
<td>Andreasen et al. 2008</td>
<td>Denmark</td>
<td>5604</td>
<td>Cross-sectional study</td>
<td>PFKP</td>
<td>rs6602024</td>
<td>BMI</td>
<td>NI</td>
</tr>
</tbody>
</table>

Individuals with higher NREE (DBP: -4.9 mm Hg, \(P=0.02\); SBP: -3.8 mm Hg, \(P=0.03\) for the third tertile). Similar results were observed when the outcome was dichotomously defined as hypertension.

An interaction between the FTO rs9939609 genotype and PA on BMI levels was observed in the Inter99 study sample (\(P_{\text{interaction}}=0.007\)). Physically inactive homozygous risk A-allele carriers had a 1.95±0.3 kg/m² increase in BMI compared with homozygous T-allele carriers.

The rs1861868 was associated with BMI to those subjects with low sex and age-adjusted PA scores (\(P<0.001\); in contrast, the SNP had no effect on those with above average PA scores (\(P=0.29\), \(P_{\text{interaction}}=0.01\)). The rs1477196 C allele was associated with a 1.22 (0.38) increase in BMI per risk allele in the low activity group (\(P=0.001\) but only a 0.27 (0.31) increase in BMI in the high-activity group (\(P=0.38\)).

An interaction between the INSIG2 rs7566605 genotype and PA level (\(P_{\text{interaction}}=0.004\)) was observed on BMI. A BMI difference of 0.53±0.42 kg/m² was found when comparing physically passive homozygous C-allele carriers with physically passive G-allele carriers.

No interaction between the PFKP rs6602024 variant and physical activity was demonstrated.
<table>
<thead>
<tr>
<th>Cross-sectional study</th>
<th>46 yr (Inter99 study sample)</th>
<th>528 (221 men/307 women)</th>
<th>Non-diabetic (NGT or IGT)</th>
<th>Questionnaire</th>
<th>HNF4A rs1885088 (G &gt; A)</th>
<th>2 h glucose</th>
<th>&lt;0.0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stephanie-May et al.</td>
<td>Canada</td>
<td></td>
<td></td>
<td></td>
<td>rs745975 (C &gt; T)</td>
<td>Glucose AUC</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>2009 Cross-sectional study</td>
<td></td>
<td>39.50 ± 14.80 yr</td>
<td></td>
<td></td>
<td>Insulin AUC</td>
<td>NI</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fasting C-peptide</td>
<td>NI</td>
<td></td>
</tr>
</tbody>
</table>

Significant genotype-by-PA interactions were found for glucose AUC and 2-h glucose levels with the HNF4A rs1885088 polymorphisms (<0.0001 for both). High level of PA (>2 h/week) was associated with lower glucose AUC and 2-h glucose values in A/A but not in G/A and G/G carriers. Interactions were also observed between rs745975 and PA on fasting C-peptide levels (Pinteraction=0.03) and insulin AUC (Pinteraction=0.003), however no statistical associations with PA were performed by genotype.

AUC - area under the curve  
BMI - body mass index  
CHD - Coronary Heart Disease  
DBP - diastolic blood pressure  
HDL-C - high density lipoprotein cholesterol.  
HOMA-IR - homeostasis model assessment–insulin resistance index  
IFG - impaired fasting glucose  
IGT - impaired glucose tolerance  
IRI - fasting immunoreactive insulin  
NEFA - nonesterified fatty acid  
NGT - normal glucose tolerance  
NI - no significant interaction  
NTI - no test of interaction  
PA - physical activity  
RMR - resting metabolic rate  
SAT - subcutaneous adipose tissue  
SBP - systolic blood pressure
SI-insulin sensitivity
VAT - visceral adipose tissue
Gene x physical activity interaction studies involving lipid enzyme encoding genes have also proven popular. For example, in the cross-sectional Inter99 study (N=5,585 Danish adults) Grarup et al. (250) observed an interaction \(P_{\text{interaction}}=0.002\) between the hepatic lipase gene \((LIPC)\) gene -250G>A polymorphism and self-reported physical activity on serum HDL-C concentrations. The authors reported that physical activity was more strongly associated with HDL-C levels in A-allele carriers than in G-allele homozygotes.

In a US study with 14 years follow-up of Hispanics and non-Hispanic whites, Hokanson et al. found that the relationship between the \(LIPC\)-480C>T variant modified the association between physical activity and coronary heart disease (CHD) (251) Overall, the TT genotype raised CHD risk, but this was offset in those reporting vigorous activity levels. Several studies have tested for interactions between the \(\varepsilon2\), \(\varepsilon3\) and \(\varepsilon4\) genotypes in the apolipoprotein E \((APOE)\) gene and physical activity on plasma lipids. For example, in a Spanish study, Corella et al. (252) found that the relationship between APOE variants and levels of HDL-C differed when comparing physically active and inactive men. No interactions between APOE polymorphisms and physical activity were observed in women, suggesting that additional sex-specific factors may further modify HDL-C levels, that the analyses in women were underpowered to detect an interaction, or that the findings in men are false-positive. In a study of Swiss men, Bernstein et al. (253) reported stronger associations between physical activity and HDL-C in the APOE4 genotype group compared with the APOE2 and APOE3 genotype groups. Cross-sectional studies examining interactions and associations between levels of physical activity and lipoprotein lipase \((LPL)\) polymorphisms on lipid profiles have yielded inconsistent results. In Dutch men and women, Boer et al. (254) found that the D9N genotype was associated with higher total cholesterol and lower HDL-C concentrations in physically inactive persons, but these relationships were not apparent in active individuals. In a study of Spanish men, Sentí et al. (255) reported an absence of interactions between physical activity and the HindIII variant on HDL-C and triglycerides levels.

In a cross-sectional study of African Americans and whites, Grove et al. (256) observed a significant interaction between the G-protein beta-3 (GNB3) variant 825C>T and physical activity on obesity risk; this interaction was evident in blacks but not in whites. In physically active blacks, each copy of the 825T allele was associated with a 20% lower prevalence of obesity, whereas in inactive individuals each copy of the 825T allele was associated with 23% increased prevalence of obesity. The authors also reported a three-way interaction between physical activity, obesity and the 825C>T variant on hypertension risk.

Nitric oxide is a molecule central to numerous vascular and metabolic processes owing to its pivotal role in vasodilation. Variants in the genes encoding the molecule that synthesizes nitric oxide \((\text{NOS}1, 2, 3)\) have been the focus of several interaction studies. For example, in studies of UK whites, Franks et al. (257) observed an interaction between physical activity energy expenditure and NOS3 haplotypes (comprised of the rs2070744, IVS1-30, and rs3800787 variants) on two hour glucose concentrations. Relative to the most common haplotype, the second most common haplotype was associated with a moderate increase in T2D risk, and the least common haplotype was associated with decreased T2D risk. Physical activity was inversely associated with two hour glucose concentrations within each of the three haplotypes, but the
magnitude of this association was significantly greater in the least common haplotype, which manifest as a statistically significant gene x physical activity interaction. In a study of Japanese adults (258), an intron 4 NOS3 variant modified the association between physical activity and systolic blood pressure levels. The authors reported a stronger inverse relationship between the genotype and blood pressure in physically inactive than in active individuals. Vimaleswaran et al. (259) also observed a significant interaction between physical activity and a NOS3 gene variant (IVS25+15) on diastolic and systolic blood pressures. In each of the three genotype groups, physical activity was inversely related with blood pressures, but the association was stronger in the major allele homozygotes than in carriers of the minor allele, indicating that these individuals’ blood pressures may be resistant to the effects of exercise. Similar findings were reported by Franks et al in a study of gene x physical activity interactions at the G-protein coupled receptor 10 (GPR10) locus (260). In that study, physical activity was inversely related with blood pressure levels in major allele homozygotes (G62G), but there was no such association in carriers of the minor A62 allele.

Several studies have exploited the recent discoveries from GWAS for obesity-related traits to undertake studies of gene x lifestyle interactions. Possibly the most striking example of a gene x physical activity interaction reported to date comes from a Danish study of FTO gene variants. In that study, Andreasen et al. (261) found an interaction between the FTO rs9939609 variant and self-reported physical activity on BMI levels in a large sample of treatment-naive Danish adults. Overall, they replicated the association between the rs9939609 variant and obesity. However, when stratified by physical activity level, they observed that the effect was substantially weaker in active compared with inactive individuals. In a report that emerged shortly after the Danish study from the Old Order Amish community, Rampersaud et al. (262) observed a statistically significant interaction between the rs1861868 variant (in low LD with the rs9930969 variant) and objectively assessed physical activity on levels of BMI. Similar interaction effects were reported for the rs1477196 variant. Since these initial studies, several other reports have emerged which reached variable conclusions (263-265).

The first GWAS report focused specifically on obesity was conducted in the Framingham Heart Study (266). The authors identified a variant in the INSIG2 gene which appeared to be associated with elevated BMI. Independent replication studies yielded conflicting results, with some suggesting that the risk allele reported in the original study was in fact protective in other populations (267). These inconsistencies might be attributable to gene x environment interactions, a hypothesis that has been tested by several groups. For example, in the cross-sectional Danish Inter99 cohort study, Andreasen et al. (198) observed an interaction between self-reported physical activity and the INSIG2 rs7566605 variant on BMI levels. The same authors conducted one of the largest studies of gene x physical activity interactions to date. In that study the gene x physical activity interactions were examined for a variant at the PFKP gene locus on BMI levels in 16,781 Danish adults. However, no evidence of interaction was found.

Several of the bona fide T2D genes confirmed to date were initially identified for their roles on monogenic diabetes. One such example is the hepatocyte nuclear factor 4 alpha (HNF4A) gene. Stephanie-May et al. (268) investigated interactions between the HNF4A rs1885088 G>A and rs745975
C>T variants with physical activity in a cross-sectional study of 528 Canadians. Significant interactions between the rs1885088 genotype and physical activity were found for both glucose AUC and 2-h glucose levels independent of BMI. High level of physical activity (>2 h/week) were associated with lower glucose AUC and 2-h glucose values in the A/A, but not in G/A or in G/G genotype groups. The A allele is generally perceived to be the T2D risk allele based on results from several large meta-analyses. Although Stephanie-May et al. (268) were unable to detect significant main effects of this variant on glucose levels, which is possibly owing to the small sample, they did find modest evidence to suggest that carriers of the A allele may be more sensitive to the beneficial effects of physical activity on glucose homeostasis. In the same study, these authors reported significant gene x physical activity interactions for the rs745975 and physical activity on fasting C-peptide levels and insulin AUC.

1.9.2 Clinical trials and other experimental studies

Intervention studies examining gene x physical activity interactions on obesity, T2D, and related traits are shown in Table 2b. PPARG has been extensively investigated for its potential modifying role in exercise-induced changes in obesity, insulin resistance, and T2D. Most published studies have focused exclusively on a single variant (Pro12Ala). Several studies have reported that Ala12 allele carriers are more responsive to the effects of exercise training as regards improvements in insulin resistance and glucose homeostasis (refs). For example, in a study of Japanese men, Kahara et al. (269) reported greater improvements in insulin action in Ala12 allele carriers than in Pro12 homozygotes after three months of aerobic exercise training. However, because the frequency of the Ala allele in this study population was very low and it was a small study, the results pivot on only a handful of individuals. Elsewhere, Adamo et al. (270) reported that after three months of supervised exercise training, initially sedentary T2D patients with the Ala12X genotype showed greater improvements in fasting plasma glucose homeostasis compared with the Pro12 homozygotes. In white Americans (32 men and 41 women) Weiss et al (271) reported that male Ala12 allele carriers experienced significant improvements in insulin levels following 6 months of supervised exercise training. No such effects were found in women. A major limitation of this study is that all but four women in the trial were Pro12 homozygotes. Thus, power to detect an interaction with exercise training on insulin levels in this trial is likely to be very low. Three randomized control trials have also analysed the effect of physical activity on risk of T2D in the Pro12Ala variant (272). The Finnish Diabetes Prevention Study involved a 3-year intensive lifestyle intervention in 490 overweight individuals with impaired fasting and two hour glucose concentrations. In a report from this study, Lindi et al. observed evidence of gene x treatment interactions on T2D incidence, where the Ala12 allele apparently protected against developing diabetes within the lifestyle intervention arm, but not in the control arm of the study. These findings suggest that although the Ala12 allele may predispose to the development of T2D under normal conditions, it may be protective in the context of a lifestyle intervention.
### Table 2b. Interventional studies of gene x physical activity interactions on T2D and related traits.

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Study Subjects</th>
<th>Lifestyle exposure</th>
<th>Gene</th>
<th>Polymorphism</th>
<th>Outcome</th>
<th>Int P-value</th>
<th>Main results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kahara et al. 2003</td>
<td>Japanese</td>
<td>123 men</td>
<td>3-month aerobic</td>
<td>PPARG</td>
<td>rs1801282</td>
<td>IRI and HOMA-IR</td>
<td>NTI</td>
<td>After exercise the Ala allele correlate with IRI and HOMA-IR (&lt;0.05). Ala allele carriers improved more in IRI and HOMA-R than those without it.</td>
</tr>
<tr>
<td>Intervention study</td>
<td>45.2±11.6 yr Healthy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weiss et al. 2005</td>
<td>USA</td>
<td>73 (32 men/41 women) 50-75 yr Healthy</td>
<td>6-month supervised endurance exercise training</td>
<td>PPARG</td>
<td>rs1801282</td>
<td>Insulin AUC</td>
<td>0.001</td>
<td>In men endurance training resulted in a significantly greater improvement in insulin AUC in Pro12Ala heterozygous as compared with Pro12 homozygous (P= 0.001) but not in women.</td>
</tr>
<tr>
<td>Intervention study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adamo et al. 2005</td>
<td>Canada</td>
<td>139 54.4±7.2 yr Diabetic patients (161 men/329 women)</td>
<td>3-month supervised exercise-training intervention</td>
<td>PPARG</td>
<td>rs1801282</td>
<td>Fasting glucose</td>
<td>NTI</td>
<td>After exercise-training intervention Ala carriers showed a greater improvement in glycaemia compared with Pro/Pro carriers (P=0.034)</td>
</tr>
<tr>
<td>Intervention study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lindi et al. 2002</td>
<td>Finland</td>
<td>490 (161 men/329 women) Overweight subjects with IGT 40–68 yr 2 groups: intensive diet and exercise group or a control group</td>
<td>3-yr follow-up of exercise (individually guide to increase PA levels) and diet (intervention group)</td>
<td>PPARG</td>
<td>rs1801282</td>
<td>T2D</td>
<td>NTI</td>
<td>Subjects with the genotype Ala12Ala didn’t develop T2D in the intervention group. However, subjects with the genotype Ala12Ala develop T2D in the control group</td>
</tr>
<tr>
<td>Randomized control trial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Florez et al. 2007</td>
<td>USA (Caucasian, African-American, Hispanic, Asian-American, American Indian)</td>
<td>3548 nondiabetic with IGT and elevated fasting glucose 51 yr 4 groups: lifestyle, metformin, or troglitazone vs. placebo</td>
<td>3-yr follow-up of diet exercise intervention</td>
<td>PPARG</td>
<td>rs1801282</td>
<td>T2D</td>
<td>NI</td>
<td>No interaction between genotype and intervention (lifestyle and metformin)</td>
</tr>
<tr>
<td>Randomized control trial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kilpeläinen et al. 2008</td>
<td>Finland</td>
<td>479 overweight individuals with IGT</td>
<td>4.2-yr follow-up of exercise (individually guide to increase PA)</td>
<td>PPARG</td>
<td>rs1801282</td>
<td>T2D</td>
<td>0.031</td>
<td>The change in the total amount of PA, stratified by median, modified the association of rs17036314 and</td>
</tr>
</tbody>
</table>
Randomized control trial

55 yr
2 groups: intervention and control group

Levels/questionnaires and diet (intervention group)

(His<sub>477</sub>His)

rs17036314 T2D 0.002
rs4135263 NI
rs2972162 NI
rs2938395 NI
rs1152003 NI

rs1801282 with the risk of T2D during the intervention (P= 0.002 and 0.031, respectively, for interaction between PA change and genotype); an increase in PA seemed to remove the effect of the risk alleles. Carriers of the rare allele of rs17036314 or rs1801282, who were in the lower half of the change in total PA tended to increase the risk of developing T2D than the carriers of the common homozygous genotype.

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Subjects</th>
<th>Duration</th>
<th>Intervention</th>
<th>Genotype</th>
<th>Trait</th>
<th>P-value</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macho-Azzarate et al. 2002</td>
<td>Spain</td>
<td>16 obese women (8 Glu27Glu and 7 Gln27Gln women)</td>
<td>60 min</td>
<td>ADRB2 rs1042714 (Gln27Glu)</td>
<td>Triglycerides Insulin</td>
<td>NTI</td>
<td></td>
<td>Plasmas of triglyceride (P&lt;0.001) and serum insulin (P&lt;0.05) were higher in the Glu27Glu group than the Gln27Gln group.</td>
</tr>
<tr>
<td>Kilpeläinen et al. 2008</td>
<td>Finland</td>
<td>487 overweight subjects with IGT</td>
<td>4.1 yrs</td>
<td>ADRB2 rs1042714 (Gln27Glu)</td>
<td>T2D.</td>
<td>NI</td>
<td></td>
<td>No reports on interactions between rs1042714 (Gln27Glu) and PA on the risk of developing T2D.</td>
</tr>
<tr>
<td>Shiwaku et al. 2003</td>
<td>Japanese</td>
<td>76 women</td>
<td>3 months</td>
<td>ADRB3 rs4994 (Trp64Arg)</td>
<td>Fasting glucose</td>
<td>NTI</td>
<td></td>
<td>After the exercise program the level of fasting plasma glucose showed a significant decrease in Trp/Trp (P&lt;0.01) and Trp/Arg (P&lt;0.05) genotypes.</td>
</tr>
<tr>
<td>Salopuro et al. 2004</td>
<td>Finland</td>
<td>490 overweight subjects with IGT</td>
<td>3 yrs</td>
<td>ADRB3 rs4994 (Trp64Arg)</td>
<td>Weight BMI Waist circumference Blood pressure</td>
<td>NTI</td>
<td></td>
<td>After intervention significant decreases were found in body weight, BMI, waist circumference and blood pressure in the wild type but not in the Trp64Arg mutants.</td>
</tr>
</tbody>
</table>

35
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Sample Size and Characteristics</th>
<th>Intervention and Study Duration</th>
<th>Genes and Variants</th>
<th>Disease</th>
<th>Associated Risk Factor</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kilpeläinen et al. 2008</td>
<td>Finland</td>
<td>487 overweight subjects with IGT</td>
<td>4.1-yr follow-up of exercise (individually guide to increase PA levels/questionnaires) and diet (intervention group)</td>
<td>ADRB3 rs4994 (Trp64Arg)</td>
<td>T2D</td>
<td>NI</td>
<td>No reports of interactions between rs4994 (Trp64Arg) and PA on the risk of developing T2D.</td>
</tr>
<tr>
<td>Laaksonen et al. 2006</td>
<td>Finland</td>
<td>506 overweight subjects with IGT</td>
<td>4.1-yr follow-up of exercise and diet (1-yr follow-up of active intervention plus 3-yr follow-up questionnaires)</td>
<td>ADRA2B 12Glu9</td>
<td>0.033</td>
<td></td>
<td>In the combined intervention and control groups, increased total leisure-time physical activity (LTPA) decreased the risk of diabetes in 12Glu carriers but not in 9Glu homozygotes (P for the interaction 0.033).</td>
</tr>
<tr>
<td>Kahara et al. 2002</td>
<td>Japanese</td>
<td>106 men</td>
<td></td>
<td>UCP-1 A-3826G</td>
<td>Fasting glucose</td>
<td>NTI</td>
<td>After the exercise program the level of fasting plasma glucose showed a significant decrease in the A/G heterozygote (P&lt; 0.01).</td>
</tr>
<tr>
<td>Salopuro et al. 2004</td>
<td>Finland</td>
<td>469 overweight subjects with IGT</td>
<td>3-yr follow-up of diet and exercise</td>
<td>UCP1 A-3826G</td>
<td>T2D</td>
<td>NI</td>
<td>No associations were observed between the A-3826G polymorphism and higher incidence of T2DM.</td>
</tr>
<tr>
<td>Todorova et al. 2004</td>
<td>Finland</td>
<td>490 overweight subjects with IGT</td>
<td>3-yr follow-up of diet and exercise</td>
<td>LIPC rs2070895 (-250G&gt;A)</td>
<td>T2D</td>
<td>0.024</td>
<td>In the control group, 23.0% of the subjects with the G-250G genotype and 19.4% of the subjects with the 250A allele converted to diabetes (P = 0.507). In the intervention group, 13.0% of the subjects with the G-250G genotype and 1.0% of the subjects with the -250A allele converted to...</td>
</tr>
<tr>
<td>Kilpeläinen et al.</td>
<td>Finland</td>
<td>487 overweight subjects with IGT 55 yr 2 groups: intervention and control group</td>
<td>4.1-yr follow-up of exercise (individually guide to increase PA levels/questionnaires) and diet (intervention group)</td>
<td>LIPC</td>
<td>rs2070895 (&lt;-250G&gt;A)</td>
<td>T2D</td>
<td>NI</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>------</td>
<td>----------------------</td>
<td>------</td>
<td>-----</td>
</tr>
<tr>
<td>Hagberg et al. 1999</td>
<td>USA Intervention study</td>
<td>51 men Healthy 45-80 yr</td>
<td>9 months of endurance exercise training</td>
<td>APOE</td>
<td>e2 isoform e3 isoform</td>
<td>HDL-C</td>
<td>NTI</td>
</tr>
<tr>
<td>Garenc et al. 2001</td>
<td>Canada Intervention study (black and white participants)</td>
<td>741 adult white and black healthy Healthy</td>
<td>20-wk of endurance training</td>
<td>LPL</td>
<td>S447X</td>
<td>BMI</td>
<td>NTI</td>
</tr>
<tr>
<td>Florez et al. 2006</td>
<td>USA Randomized control trial (Caucasian, African-American, Hispanic, Asian-American, American Indian)</td>
<td>3548 overweight or obese participants at high risk of T2D 51±11 yr 4 groups: lifestyle, metformin or troglitazone, vs placebo</td>
<td>3-yr follow-up of diet and exercise intervention</td>
<td>TCF7L2</td>
<td>rs12255372 rs7903146</td>
<td>Diabetes</td>
<td>NI</td>
</tr>
<tr>
<td>Franks et al. 2008</td>
<td>USA Randomized control trial (Caucasian, African-American, Hispanic, Asian, and American)</td>
<td>3451 men overweight or obese participants at high risk of T2D 51 yr 4 groups: lifestyle, metformin or troglitazone,</td>
<td>1-yr follow-up of diet and exercise</td>
<td>FTO</td>
<td>rs99399609</td>
<td>Obesity</td>
<td>NI</td>
</tr>
</tbody>
</table>
Indian vs placebo

Franks et al. 2008 Randomized control trial

USA (Caucasian, African-American, Hispanic, Asian, and American Indian)

3556 men overweight or obese participants at high risk of T2D

51 yr 4 groups: lifestyle, metformin or troglitazone, vs placebo

1-yr follow-up of diet and exercise

INSIG2 rs7566605 Obesity NI

Nominally significant interactions were observed between INSIG2 rs756605-lifestyle and weight change (P=0.02) (n=3556), subcutaneous adipose area at L2/3 (P=0.01) and L4/5 (P=0.03), and for visceral adipose area at L2/3 (P=0.02) (n=725). However, no statistical evidence of association with PA was observed for either genotype.

Kilpeläinen et al. 2007 Randomized control trial

Finland

479 overweight subjects with IGT 55 yr 2 groups: intervention and control group

4.1-yr follow-up of exercise (individually guide to increase PA levels/questionnaires) and diet (intervention group)

SLC2A2 rs5393, rs5394 T2D 0.027

The carriers of the common homozygous genotype of rs5393, rs5394, or rs5404 of SLC2A2 and rs3758947 of ABCC8 who were in the lower third of the change in moderate-to-vigorous PA during the follow-up had a 2.6- to 3.7-fold increased risk of developing T2D compared with the upper third, whereas the rare allele carriers seemed of genotype with change in PA, P= 0.022– 0.027 for the SNPs in SLC2A2, and P= 0.007 for rs3758947).

Kilpeläinen et al. 2008 Randomized control trial

Finland

487 overweight subjects with IGT 55 yr 2 groups: intervention and control group

4.1-yr follow-up of exercise (individually guide to increase PA levels/questionnaires) and diet (intervention group)

TNF rs1800629 (−308G/A) T2D NI

IL6 rs1800795 (−174C/G) T2D NI

IGF1R rs2229765 (Glu1013Glu) T2D NI

LEPR rs1137100 (Lys109Arg) T2D NI

GHRL rs27647 (−604G/A) T2D NI

No interaction between the polymorphisms and PA on the conversion to T2D was found during 4.1-year follow-up. The Leu72Met (rs696217) polymorphism in GHRP modified the effect of moderate-to-vigorous PA on the changes in SBP (P=0.001) and waist circumference (P=0.006), the −501A/C (rs26802) polymorphism in GHRH modified the effect of total (P= 0.005) and moderate-to-vigorous (P=0.024) PA on the change in HDL-C concentration, and the Lys109Arg (rs1137100) polymorphism in LEPR.
Moore et al.  
2008 Randomized control trial  
USA (Caucasian, African-American, Hispanic, Asian, and American Indian)  
3548 overweight or obese participants at high risk of T2D  
51 yr 4 groups: lifestyle, metformin or troglitazone, vs placebo  
1-yr follow-up of diet and exercise

<table>
<thead>
<tr>
<th>Gene</th>
<th>rs Number</th>
<th>Trait</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>−473G/A</td>
<td>(no rs number)</td>
<td>HDL-C</td>
<td>0.024</td>
</tr>
<tr>
<td>Arg51Gln</td>
<td>(no rs number)</td>
<td>Weight</td>
<td>NI</td>
</tr>
<tr>
<td>rs696217 (Leu72Met)</td>
<td></td>
<td>Waist circumference</td>
<td>0.001</td>
</tr>
<tr>
<td>rs4684677 (Gln90Leu)</td>
<td></td>
<td></td>
<td>NI</td>
</tr>
</tbody>
</table>

modified the effect of total PA on the change in systolic blood pressure \( (P=0.017) \).

A significant interaction between genotypes and treatment arm on insulin secretion was observed at CDKN2A/2B rs10811661 in crude \( (P=0.03) \), ethnicity adjusted \( (P=0.04) \), and BMI \( (P=0.04) \) analyses.

In ethnicity-adjusted analyses a nominal differential improvement in b-cell function for carriers of the protective genotype after 1 year of troglitazone treatment \( (P=0.01) \) and lifestyles modification \( (P=0.05) \) were observed. These results persisted when adjusted for BMI alone in both the troglitazone treatment \( (P=0.03) \) and lifestyles modification \( (P=0.02) \).

Moore et al.  
2009 Randomized control trial  
USA (Caucasian, African-American, Hispanic, Asian/Pacific Islander, and American Indian)  
3534 50-5 yr overweight or obese participants with IGT 3 groups: lifestyle or metformin vs placebo  
1-yr follow-up of diet and exercise

<table>
<thead>
<tr>
<th>Gene</th>
<th>rs Number</th>
<th>Trait</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENPP1</td>
<td>rs1044498 (K121Q)</td>
<td>Diabetes</td>
<td>0.03</td>
</tr>
</tbody>
</table>

There was a significant interaction between genotype at rs1044498 and intervention under the dominant model \( (P=0.03) \). In analyses stratified by treatment arm, a positive association with diabetes incidence was found in Q allele carriers compared to KK homozygotes \( (P=0.009) \) in the placebo arm \( (n=996) \). Lifestyle modification eliminated this increased risk. These findings persisted after adjustment for body mass index and race/ethnicity.
AUC - area under the curve
BMI - body mass index
CHD - Coronary Heart Disease
DBP - diastolic blood pressure
HDL-C - high density lipoprotein cholesterol.
HOMA-IR - homeostasis model assessment–insulin resistance index
IFG - impaired fasting glucose
IGT - impaired glucose tolerance
IRI - fasting immunoreactive insulin
NEFA - nonesterified fatty acid
NGT - normal glucose tolerance
NI – no significant interaction
NTI - no test of interaction
PA - physical activity
RMR - resting metabolic rate
SAT - subcutaneous adipose tissue
SBP - systolic blood pressure
SI-insulin sensitivity
VAT - visceral adipose tissue
A limitation of the studies described above is that none appropriately tested for gene x exercise interactions. In the Diabetes Prevention Program, which engaged a similar lifestyle intervention to the Finnish Diabetes Prevention Study, but included a much larger US multiethnic cohort (N=3,548 non-diabetic participants with elevated fasting and 2hr glucose), Florez et al. (273) reported an absence of statistical interaction between Pro12Ala (and five other PPARG variants) and intervention (lifestyle and metformin) on the incidence of T2D. However, statistically significant interactions between the Pro12Ala variant and dietary fat intake on weight change were observed in the Diabetes Prevention Program by Franks et al (274). In another report from the Finnish Diabetes Prevention Study, Kilpeläinen et al. (275) described statistical interactions between the rs17036314 and Pro12Ala SNPs with the change in total amount of self-report PA ($P_{\text{interaction}}=0.002$ and $P_{\text{interaction}}=0.031$, respectively). In these observational analyses, increased physical activity tended to decrease the effect of the rs17036314 and Pro12Ala risk alleles on T2D incidence.

Inconsistent results have been reported regarding the interactions between the genes encoding the adrenergic receptors and physical activity. Two studies examining the ADRB2 gene, including one randomized control trial, found no evidence of any effects of physical activity and Gln27Glu on plasma triglycerides and serum insulin (276) or T2D (277). Similarly, four studies of ADRB3 variants, including two randomized control trials, did not observe any statistical interactions between physical activity and the Trp64Arg polymorphism on fasting glucose (278), anthropometric measures of obesity, blood pressure (279), or T2D risk (277, 280). In contrast, in the combined intervention and control groups of the Finnish Diabetes Prevention Study, Laaksonen et al. observed a statistical interaction between changes in estimated total leisure-time physical activity (LTPA) and the ADRB2 12Glu9 polymorphism ($P_{\text{interaction}}=0.033$). In that study, high levels of LTPA tended to decrease the risk of developing diabetes in 12Glu carriers but not in 9Glu9 homozygotes (281).

Numerous other genes have been studied in exercise or lifestyle intervention studies. These include UCP-1, for which results on gene x exercise interactions are inconclusive (278, 280). As is the case with observational studies, several intervention studies have examined the interaction between exercise and LIPC variants. For example, in the Finnish Diabetes Prevention Study, an interaction between the −250G>A variant and lifestyle intervention was observed for T2D incidence, where the effect of the G250 allele was more pronounced in the intervention group as compared with the control group. In the control group the cumulative incidence of diabetes was 23% in subjects with the G-250G genotype, whereas in the intervention group only one participant with the -250A allele developed diabetes. This finding suggests that changes in lifestyle, such as diet, physical activity, and weight loss, modify diabetes risk predisposition in carriers of the -250A allele. However, in a follow-up observational analysis of the same cohort, Kilpeläinen et al. (277) were unable to detect interactions between the −250G>A variants and changes in physical activity or body weight on the risk of T2D at 4.1-years follow-up, suggesting that the interaction effect reported in the initial study was either false positive or involved aspects of lifestyle other than physical activity (such as diet). In one of the earliest studies of genes and exercise intervention, Hagberg et al. (282) reported that APOE genotypes influence the HDL-C response to exercise, with
APOE2 carriers being more responsive than APOE3 and APOE4 carriers following 9 months of endurance exercise. This was a small study (N=51), which lacked a control arm, and no formal tests of interaction were reported. Elsewhere, Garenc et al. (283) examined the LPL S447X variant in the context of a 20 week exercise intervention study (the HERITAGE Study); white women with the X447 allele exhibited greater reductions in BMI, fat mass, and percent body fat, following the intervention, and black women with the same genotype showed a greater reduction in abdominal visceral fat and a larger increase in post-heparin LPL activity following the intervention. However, these effects were not apparent in the male participants in the study.

The recent advances in the genetics of obesity and T2D have prompted analyses exploring whether these confirmed disease-predisposing loci modify the response to lifestyle interventions. In the Diabetes Prevention Program for example, Florez et al. (284) showed that carriers of the T2D-predisposing alleles at the transcription factor 7–like 2 (TCF7L2) gene (rs12255372 and rs7903146) experienced a similar reduction in diabetes risk following lifestyle intervention. In the same cohort, Franks et al. (285) analysed two genes selected from obesity GWASs. The authors reported a positive association between the FTO rs9930969 variant and one year weight change in the placebo group, which was reversed in the lifestyle intervention group, suggesting a gene x treatment interaction. Similar analyses conducted in the Finnish Diabetes Prevention Study did not yield a significant gene x treatment interaction effect for this FTO variant (286). A series of observational studies have since been published with inconsistent results (263-265, 287). In the Diabetes Prevention Program, the INSIG2 variant (rs7566605) initially identified in a GWAS conducted by Herbert et al (266) was studied in relation to weight change, and was found to convey discordant effects in the placebo and lifestyle intervention groups, suggesting the presence of a gene x lifestyle interaction at this locus.

Another gene, which has been the focus of studies of glucose homeostasis and has been studied in the Finnish Diabetes Prevention Study is the solute carrier family 2 (facilitated glucose transporter), member 2 (SLC2A2). Kilpeläinen et al. (288) reported significant interactions between SLC2A2 variants (rs5393, rs5394, and rs5404) and physical activity on T2D incidence. The same authors also reported tentative evidence of gene x physical activity interactions on diabetes risk for the rs3758947 variant in the ABCC8 gene and the E23K variant in the KCNJ11 gene when analysing observational data from the Finnish Diabetes Prevention Study. Kilpeläinen et al. (277) also reported an absence of statistically significant gene x physical activity interactions for variants in the genes encoding the leptin receptor (LEPR), ghrelin/obestatin prepropeptide (GHRL), tumor necrosis factor alpha (TNFA), interleukin 6 (IL6), and insulin-like growth factor 1 receptor (IGF1R). These authors did however observe nominal interactions for the GHRL rs696217 variant and physical activity on changes in waist circumference, body weight, and HDL-C levels. They also reported interactions for the LEPR rs1137100 variant on systolic blood pressure during the first year of the trial. A major limitation of the results from the Finnish Diabetes Prevention Study the study is relatively small, the results are often nominally statistically significant, and a wide range of hypotheses have been tested. Thus, many of these findings may be prone to type 1 error.

In the Diabetes Prevention Program, Moore et al. undertook gene x treatment interaction analyses for 10 previously confirmed T2D-predisposing
loci (289). The most promising finding emerged for a variant (rs10811661) proximal to the genes (CDKN2A/B) encoding the cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4) and cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4) proteins. A significant interaction was observed between the CDKN2A/B rs10811661 variant and lifestyle intervention on beta-cell function after 1 year of treatment. A borderline significant gene x lifestyle interaction on diabetes incidence was also reported. A second study by Moore et al. (290) in the Diabetes Prevention Program examined interactions between the K121Q variant in the ectonucleotide pyrophosphatase/phosphodiesterase 1 (ENPP1) gene and lifestyle intervention; in that report, the increased risk of diabetes observed in Q allele carriers in the placebo group was diminished in the lifestyle intervention group.

1.10 Methodological and statistical considerations in studies of interaction

1.10.1 Sample size and statistical power

The literature on gene x physical activity interactions on T2D and related traits is dominated by initial positive results which lack appropriate replication. A low level of statistical power would influence the replicability of a study’s findings if (as is likely) this influences the way in which analyses are performed and reported. For example, investigators may be more likely to over-analyses data from small, poorly designed studies and selectively report these results.

Power is defined as the probability of finding a significant association if it truly exists. By convention, many consider level of 80% to be a reasonable threshold for statistical power, which those studies powered above this level being considered appropriately powered, although this threshold is arbitrary (291). Studies that simulated the power to detect gene x environment interactions demonstrate that the sample size necessary to detect a an interaction is generally larger than the sample size required to detect the genetic or environmental main effects (292, 293). The main factors that affect power in studies of interaction are the strength of the association between exposure and outcome, the magnitude of the interaction effect, and the frequency of the minor allele, sample size, and measurement error (291). For example, a study of interaction with continuously distributed outcomes (e.g., glucose, insulin, BMI) with accurate measurements of the exposures and outcomes would be equally powered to detect gene x lifestyle interactions as one ten to 50 times larger which had incorporated weak measures of the exposures and outcomes (291).

Figure 4 shows an example of the samples sizes necessary to achieve 80% power at 0.01 significance level for varying effect sizes of the gene x environment interaction using a candidate gene association studies for different study designs (case–control, trio, case-only and cohort) assuming a recessive disease allele with frequency 0.3 (294).
Figure 4. Sample size requirements for 80% power to detect a gene–environment (GxE) interaction for different study designs depending on the strength of the interaction. Sample sizes for case–control, case–parent trio, and case-only designs were calculated using Quanto59 (http://hydra.usc.edu/gxe), assuming an analysis by (conditional) logistic regression. For the cohort design, sample sizes are estimated using Power (http://dcegqa.cancer.gov/bb/tools/power), which is based on a prospective binary response model. Shown are the number of individuals required to detect a significant GxE interaction effect at $\alpha=0.01$ with a power of 80%. Solid lines represent the case–control design, dotted lines the trio design, dashed line the case-only design and dotted-dashed lines the cohort design. The horizontal solid line represents the sample size required for 80% power to detect a genetic main effect using a case–control design. The interaction odds ratio was varied between 1.25 and 3 whereas the main effects of the genetic and environmental risk factors were 1.2 (a) and 1.5 (b). The disease model was defined by a recessive disease allele with frequency 0.3. The environmental risk factor had a prevalence of 30%. The baseline risk of the disease was 10%. The samples sizes to detect the genetic main effect, which were constant in the two scenarios, were 43 045, 16 196 and 19 860 in (a) for the cohort, case–control and trio design, and 7712, 3070 and 3423 in (b), respectively. For a dominant disease allele, similar relations between required sample sizes are observed for the different designs (294).

Most epidemiological studies of gene x physical activity interactions on T2D and related traits have used questionnaires in order to assess information on lifestyle behaviour. Although most of the commonly used physical activity questionnaires appropriately classify the behaviour when compared with gold standard methods (e.g. doubly-labelled water estimated free-living energy expenditure), the correlation between methods is usually low $r<0.25$ and reporting biases associated with disease outcomes can complicate the interpretation of results (295, 296). On the other hand objective measurements of physical activity (e.g., accelerometry) are generally more accurate ($r>0.5$) (297-300). Thus, studies of gene x physical activity interaction may benefit from the availability of objective assessment methods. However, it is important also to recognize that objective methods usually reflect a relatively short period of free-living activity and may be prone to Hawthorne Effect, whereas questionnaires often reflect activity levels during the past months or years.

1.10.2 Multiple hypothesis testing

In general, few studies of gene x physical activity interaction on T2D and related traits have corrected analyses for multiple hypothesis testing (multiple
statistical comparisons). Multiple testing is a particular problem in studies of gene x environment interactions because such studies often involve the comparison of numerous genetic markers, environment factors, and outcomes. Such tests may also be performed at a secondary level and represent exploratory analyses, rather than those identified during the analysis planning stage. Providing this information is reported in the publications documenting the results, this should not present major problems for the appropriate interpretation of the results; however, this may not always be the case.

A number of methods have been proposed to statistically correct for multiple testing. In studies of gene x physical activity interactions the methods most used to date are the Bonferroni correction and the false discovery rate (FDR). The classical Bonferroni correction compensates for $n$ independent tests by testing each individual hypothesis (setting for significance threshold: alpha=0.05) divided by the number of tests. It is generally agreed that this method may be overly conservative, largely because it assumes that each test is independent, which is rarely true (301). The FDR on the other hand is a study-specific estimate of type 1 error and may be more appropriate in studies of interaction (302). This method may reduce false positives findings and simultaneously preserves power to detect true findings providing a more consistent and robust measure to control the false positive rate.
2 AIMS

The overall aim of this thesis is to quantify the main genetic effects and gene x physical activity interactions on T2D and related traits in paediatric and adult cohorts. The *PPARGC1A* gene (Paper I) was selected because this gene has strong *a priori* biological evidence supporting its role as a candidate gene for interaction with physical activity on cardiometabolic trait predisposition (303). *TCF7L2* was included in this thesis (Papers II and III) because for both paediatric and adult cohorts variants in this gene have been robustly associated with T2D or its quantitative traits; thus this variant is considered a positive control with respect to T2D in the adult cohorts and glucose levels in the paediatric cohorts studied here. The remaining work described in this thesis (Papers III and IV) follows on from recent GWAS discoveries. The change in methodological emphasis in this thesis from the biologic candidate gene approach to one where GWAS derived variants were studied reflects a general change in methodological emphasis in the field of population genetics of complex traits during the past four years.

The specific aims of this thesis are as follows:

2.1 Paper I

To comprehensively assess the main genetic effects and gene x physical activity interactions for variants at the *PPARGC1A* gene on cardiovascular and metabolic disease traits in a cohort study of Danish and Estonian children from the European Youth Heart Study (EYHS).

2.2 Paper II

To assess the relationships between a previously associated *TCF7L2* gene variant and T2D-related traits, and to extend previous findings by examining gene x physical activity interactions on growth and insulin and glucose concentrations in a cohort study of Danish and Estonian children from the EYHS.

2.3 Paper III

To determine whether the effects of 17 previously associated T2D gene variants on the risk of impaired glucose regulation (IGR) or incident T2D are modified by physical activity in a prospective population-based cohort of initially non-diabetic middle-aged Swedish adults from the Malmö Preventive Project (MPP).

2.4 Paper IV

To examine whether the effect of *FTO* genotypes on obesity is modified by physical activity in two middle-aged cohorts of Swedish and Finnish adults from the MPP and the Prevalence, Prediction and Prevention of Diabetes in Botnia (PPP- Botnia), respectively.
3 METHODS AND MATERIALS

Table 3 lists the methods used in this thesis.

3.1 Study populations

This thesis is based on data collected from a multicenter school-based cohort (EYHS) (Papers I, II) and two adult population-based cohorts (the MPP - Papers III, IV) and the (PPP- Botnia Study - Paper IV). For EYHS (Papers I, II), written informed consent was obtained from a parent or guardian and all children gave verbal assent after a detailed explanation of the test procedures. Both EYHS study centres (Denmark and Estonia) obtained local research ethics committee approval. For Papers III and IV, all living participants gave written informed consent to take part in the studies, which were approved by the local research ethics committees of Lund University and Malmö University Hospital (MAS) (Papers III, IV) and Helsinki University Hospital (Paper IV) (MPP and PPP-Botnia Study, respectively). All research included in this thesis conducted according to the international guidelines for biomedical research outlined in the declaration of Helsinki.

3.1.1 European Youth Heart Study (EYHS) – Papers I, II

The EYHS started in 1997 and was designed as a multicenter school-based cohort of pre- and early-pubertal children (school grades 3 and 9, respectively) randomly selected by a two-stage sampling strategy. The main objectives of the EYHS were to examine the nature, strength, and interactions between personal, environmental, and lifestyle influences on CVD risk factors in healthy children (~4168 children) from 4 geographically-defined areas of Europe (the city of Odense, Denmark; the city of Oslo, Norway; the city and surrounding rural areas of Tartu, Estonia; and the island of Madeira, Portugal) (304). For genetic studies, DNA was available from the Danish, Estonian, and Portuguese centres. Owing to DNA contamination, only the Danish and Estonian DNA were suitable for genetic analyses. Briefly, at each study centre, schools were stratified according to location (urban, suburban, rural) and socioeconomic profile of the recruitment area (high, middle, low). From each stratum, children were randomly selected by a proportional, two-stage cluster sample. The first sampling criterion was schools, whereby the schools were clustered in samples (where sampling probability was proportional to school size). A minimum of 20 schools were randomly selected from the local authority’s school lists within appropriate age, sex, and socio-economic strata. The second selection criterion was classes (within schools) within school grades 3 and 9, where children registered in that grade and within age bands 8–10 and 14–16 years old were allocated with code numbers and randomly selected (using random number tables). Thus, all children who were randomly selected were invited to participate in the study along with their parents. The present analyses (Papers I, II) were based on a cross-sectional survey of the EYHS at the baseline examination (1997–2000) in participants for whom data on genotypes, anthropometry, blood pressure, biochemical traits, physical activity, and aerobic fitness were available. Paper I and II include 2,101 children and adolescents from Denmark and Estonia.
3.1.2 Malmö Preventive Project (MPP) – Papers III, IV

The MPP was a preventive case-finding programme started in 1974 at the Section of Preventive Medicine, Department of Medicine, University Hospital, Malmö, Sweden (83). This study comprised a large proportion of the adult population born in pre-specified adult age-groups and living in the city of Malmö in southern Sweden. The initial aim of the study was to find individuals at high risk of developing chronic diseases so that they could be provided with preventive interventions aimed to reduce cardiovascular risk factors, impaired glucose tolerance, breast cancer, and alcohol dependency. Participants were invited to participate in a general health screen, which included a physical examination, a glucose tolerance test, a mammography (for women), and a self-administered questionnaire on lifestyle habits. Between 1974 and 1992 a total of 22,444 men and 10,902 women attended the screening programme, with an overall attendance rate of 71.2%. The study began by studying only men (1974–1982). Women were studied from 1981 through 1992.

Figure 5 shows the recruitment schema for the MPP (26). During a median follow-up of 24.8 years, 2,063 subjects (12.8%) developed diabetes, with the highest conversion rate among those with impaired fasting glucose (IFG) levels or impaired glucose tolerance (IGT) at baseline. Diagnosis of T2D was confirmed from patient records.

![Figure 5](image)

**Figure 5** Recruitment schema and an outline of the data collection procedure in the Malmö Preventive Project, in which 16,061 subjects without diabetes were initially eligible for the study of the prediction of future diabetes; T2D developed in 2,063 (12.8%) of these subjects (26).
The present analyses are based on a prospective population-based cohort study of initially non-diabetic middle-aged adults (Papers III, IV). Data were analyzed during a median follow-up period of 24.5 years (Papers III) and 23 years (Papers IV) for incidence of T2D and change in BMI, respectively. Papers III and IV comprise 16,003 and 15,931 MPP participants, respectively.

3.1.3 Prevalence, Prediction and Prevention of Diabetes in Botnia (PPP-Botnia)

The PPP-Botnia Study started in 1990 and was initially designed as a family-based study from four primary health care centres (Narpes, Malax-Korsnas, Korsholm, and Jakobstad) in the Botnia region of western Finland (305). This region was settled about 1,000 years ago by Swedish and Finnish ancestors whose descendants have remained in a mixed rural and urban area. Briefly, at health care centres within the region, persons with T2D and their family members were invited to participate in the study. The main aim of the study was to examine genetic and other inherited risk factors for T2D.

Figure 6 shows the recruitment schema for the PPP-Botnia Study. During the follow-up period, 138 of the 2,770 initially non-diabetic subjects developed diabetes (5.0%). The analyses report in this thesis are based on all available PPP-Botnia samples in which anthropometric, lifestyle and genetic data were available (N=2,511). Diagnosis of T2D was confirmed from patient records.

![Figure 6](image_url)

**Figure 6** Recruitment schema for the Botnia Study. Progression to diabetes in the Botnia Prospective Study, which included 2,770 family members and spouses without diabetes; T2D developed in 138 (5.0%) of these individuals (26).

3.2 Measures of anthropometry, growth, and blood pressure

Weight and height were measured using standard anthropometric techniques with subjects in light-weight clothing without shoes (Papers I-IV). BMI was calculated as weight in kg divided by height in meters squared (kg/m²) (Papers
I-IV). Waist circumference was measured midway between the lower rib margin and the iliac crest at the end of gentle expiration using a metal anthropometric tape (Paper I). In children from the EYHS, sexual maturity was assessed with Tanner’s 5-stage scale for breast development in girls and pubic hair in boys (Paper I, II). The sum of five skinfold-thickness measurements (triceps, biceps, subscapula, suprailiac, and calf) were taken according to the criteria described by Lohman et al and was used as an indicator of body fat percentage (Paper II). Blood pressure (BP) was measured with a Dinamap paediatric/adult and neonatal vital signs monitor (model XL, Critikron, Inc, Tampa, FL, USA). Five measurements were taken at 2-minute intervals and the mean of the last three measurements was used in analysis (Paper I). All variables were additionally standardized by age and sex using a conventional z-score procedure. BMI was also standardised to age and sex using the methods described by Cole in order to overcome growth-related deviations in BMI that occur in childhood (Papers I, II).

3.3 Blood samples and biochemistry

Blood samples were taken from the antecubital vein (Papers I-IV). Plasma glucose was measured using a hexokinase method (Papers I-IV) and the glucose oxidase method (Paper IV). No significant difference was observed between the glucose oxidase method and the hexokinase method regarding venous blood glucose concentrations.

In the MPP study, blood samples were drawn at 0, 40, and 120 minutes during a 30 g glucose/m² body surface OGTT. Fasting samples were drawn at a follow-up visit for measurement of plasma glucose (Papers III, IV). The OGTT method used in the MPP predated the WHO 75 g OGTT method, but previous validation studies indicate that the two approaches are comparable (306). In the PPP-Botnia study, blood samples were drawn 10 minutes before the 75g OGTT and then at 0, 30, 60, and 120 minutes (Paper IV). The homeostasis model assessment (HOMA) index was used to estimate insulin resistance beta-cell function [HOMA-B = 20 x fasting insulin (μU/mL) / fasting glucose (mmol/L) - 3.5] (Papers I, II) and insulin sensitivity [HOMA-S = 22.5/ fasting insulin (μU/mL) x fasting glucose (mmol/L)] (Papers I, II). Total cholesterol, HDL-C, and triglyceride were measured by enzymatic methods (Olympus Diagnostica, Hamburg, Germany) (Papers I). Insulin was analysed using enzyme immunoassay (microtitre plate format; Dako Diagnostics, Ely, UK) (Papers I, II). LDL-C was calculated using the Friedwald equation (307)(Paper I).

3.4 Genetics

3.4.1 SNP selection

SNPs included in this these are those which tag common variation within a genomic region (Paper I) and those which have been previously associated with relevant cardiometabolic traits (Papers I-IV). T

3.4.2 Selection of tagging polymorphisms and definition of linkage disequilibrium
The International HapMap Project has provided a dense map of common genetic variants across the human genome from collections of trios from four diverse ethnic groups. These data can be used to select common SNPs that tag regions of genomic variation within ethically comparable populations to those featured in HapMap. This highly economic approach has proven effective in association studies examining the role of common genetic variants in a wide range of diseases and related traits (308).

The HAPMAP DNA samples were selected from a total of 270 people in whom detailed genotyping has been performed (approximately 2 million SNPs [minor allele frequency (MAF) ≥ 0.05]. The four populations included in HapMap comprise 45 unrelated Japanese-descent mother-father-child trios (Tokyo, Japan [JPT] descent); 45 unrelated Chinese-descent trios (Han Chinese, Beijing, China [CHB] descent); 30 unrelated African-descent trios (Yoruba in Ibadan Nigeria [YRI] descent); and 30 unrelated European-descent trios [U.S. population, Utah residents from Northern and Western European ancestry (CEU) collected by the Centre d’Etude du Polymorphisme Humain (CEPH) (309). The main goal of the International HapMap Project was initially to construct a haplotype map of the human genome because it was hypothesized that haplotype blocks and haplotypes (a series of closely linked alleles found on the same chromosome inherited as a unit) were shared within populations of the same ethnicity. The aim of this strategy was to economize on the number of variants that needed to be genotyped within a given study, by selecting only those which were most informative of common genomic variation a given locus. In other words, tagSNPs serve as proxies for groups of highly correlated neighbouring SNPs (310-312).

Linkage disequilibrium (LD) (or allelic association) exists when alleles at two or more neighbouring loci tend to be inherited together as a unit more often than would be expected to by chance (313). Polymorphisms that are in high LD share a common ancestry because they have not been separated by recombination or by the occurrence of new mutations within the region (310, 314, 315). In addition, LD is also influenced by events such as isolation, migration, admixture and population bottlenecks (the random sampling of gametes during restriction in population size). All these events may contribute to differences in allele frequencies between populations (316). The two most common measures used to describe the strength of LD between pairs of SNPs are the pairwise-disequilibrium coefficient ($D'$) and the square of the correlation coefficient ($r^2$). In this study we selected the latter method, as it has emerged as the most commonly used of these two approaches. Both measures range from 0 (no disequilibrium) to 1 (complete disequilibrium). A value of 0 means that the two markers are independent by contrast, a value of 1 implies that the markers have not been separated by recombination or new mutations. Thus $D' = 1$ reflects complete LD, because the rarer allele occurs exclusively with one of the two alleles at the other marker. By contrast and $r^2$ equal to 1 indicates complete LD because the allele frequencies are identical at every observed occurrence. Thus, either one of the pair of SNPs perfectly predicts the SNP at the other locus (Zondervan and Cardon 2004) (313). In Paper I, a total of 35 SNPs were selected. These included known common variants and additional tagSNPs. This was the first published study to examine all common PPARGC1A sequence variation (±10 kb). We undertook pair-wise tagging ($r^2 > 0.8$ and minor allele frequency >0.05) using the Tagger software in Haploview (http://www.broad.mit.edu/mpg/haploview, accessed 1 May 2008) from the

3.4.3 Genotyping

All TCF7L2 genotyping (Paper II) and almost all PPARC1A genotyping (Paper I) was performed using the Illumina Beadstation custom-array (Illumina, San Diego, CA, USA). The exception was for Gly482Ser (rs8192678), which was genotyped at the MRC Epidemiology Unit genotyping facility using TaqMan (Applied Biosystems, Foster City, CA, USA). For Paper III, genotyping was performed at the Clinical Research Center at Lund University Diabetes Center in Malmö using the following methods: the matrix-assisted laser desorption–ionization (MALDI-TOF) time-of-flight mass spectrometry method on the MassARRAY platform (Sequenom) (rs7903146, rs1801282, rs5219, rs7754840, and rs10811661); the allelic discrimination assay-by-design approach on ABI 7900 (Applied Biosystems) (rs4430796, rs4402960, rs10010131, rs1111875, rs864745, rs12779790, rs7961581, rs7578597, rs4607103, rs10923931, and rs10830963); and, the Allele-specific assay (KASPar, KBioscience) (rs13266634). FTO genotyping (Paper IV) was carried out using the TaqMan-assay on the ABI7900 platform (Applied Biosystems, Foster City, CA, USA).

3.4.4 Quality control – Genotyping success rate

The genotyping success rate is defined by the number of genotypes that were successfully obtained. A genotyping success rate > 95% was considered of sufficiently high-quality. Rates lower than this may result in a degree of selection bias if factors influencing genotyping success are also correlated with the traits of interest (e.g. low white blood cell count and immunosuppressive disorders). In Paper I, a total of 62 DNA samples were of inadequate quality and failed for all genotyping assays. Therefore, these were excluded from analyses. The genotyping success rates for all SNPs exceeded 98% with the exception of the genotyped Gly482Ser (rs8192678) variant that was 96.4% (Paper I). Genotyping success rates for the TCF7L2 (Paper II) and FTO (Paper IV) SNPs were >95% and >99%, respectively. The genotyping success rate and accuracy for all SNPs (11% of samples were regenotyped using Sequenom) exceeded 95% and 98.7%, respectively (Paper III).

3.4.5 Quality control – Hardy-Weinberg Equilibrium

Hardy-Weinberg equilibrium (HWE) was calculated as part of the genotyping quality control checks performed in this study. An underling assumption when calculating HWE is that in large population samples the frequency of alleles and genotypes will remain relatively constant from one generation to the next if the population structure is stable (317). The HWE model describes a mathematical relationship \[p^2+2pq+q^2= (p+q)^2 = 1; \] whereas at a single locus \( p \) represents the frequency of one allele and \( q \) the frequency of the other allele that allows the prediction of the frequency of offspring genotypes based on parental allele frequencies, which means that it is an equilibrium or a non-evolutionary model with no changes in the gene pool. This law provides a baseline to inform whether alleles frequencies have changed in a population and thus determine
whether evolution has occurred. Deviations from HWE in the population may be due to consanguinity (i.e. inbreeding of close relatives), genetic drift (the random appearance and disappearance of mutations within a population), migration, and environmental pressure (i.e. the selective advantage or disadvantage of a particular allele in a given environment that influences survival fitness). Most commonly, however, deviations from HWE in population-based samples related to methodological problems such as error in genotyping or population stratification. In this thesis each SNP was tested for departures from HWE using Haploview v4.0 (http://www.broad.mit.edu/mpg/haplovew). All polymorphisms were in HWE ($P>0.01$), indicating appropriate genotype distributions in these populations the absence of excessive genotype misclassification.

### 3.5 Physical activity and aerobic fitness

Free-living physical activity was assessed with a conventional uni-axial accelerometer (MTI Actigraph, model WAM 7164; Manufacturing Technology Inc, Fort Walton Beach, FL, USA) during 2 weekdays and 2 weekend days, as previously described (Paper I, II) (318, 319). Briefly, all the children wore the accelerometer on an elastic waistband on the right hip during the daytime, except while bathing and during other water-based activities. Physical activity data were stored on a minute-by-minute basis and were downloaded to a computer via a serial interface for data cleaning before analysis. Criteria for a successful recording were a minimum of 3 days of 600 min recording per day. In Papers I and II, physical activity was expressed as total counts per unit time (counts/min$^{-1}$), which is an indicator of the total volume of activity (i.e., average intensity of activity). It has been previously shown that this variable is significantly correlated with physical activity energy expenditure obtained by the doubly-labelled water method (320).

For Papers III, IV, habitual physical activity was assessed using a self-administered computer-based questionnaire. The MPP spanned a number of years and during that time the questions used to assess physical activity were changed (Papers III, IV). To overcome this limitation, a binary variable was constructed in order to classify individuals as ‘physically active’ or physically inactive’ based on their responses to six questions: (1) do you walk or cycle to and from work; (2) do you walk or cycle for recreation during weekdays; (3) do you walk or cycle for recreation during weekend days; (4) do you undertake at least 3 h/week of structured physical exercise; (5) do you walk to work or do yard work; or (6) do you perform light structured physical exercise each week.

In the PPP-Botnia Study, the Kuopio Ischemic Heart Disease (KIHD) 12-month LTPA questionnaire was used to assess physical activity (321, 322). This questionnaire was based on a modification of the Minnesota LTPA (323) and was adapted to reflect physical activity patterns and types in the Finnish population based on findings from the Mini-Finland survey. A validation study was performed against maximum oxygen uptake ($V_{O_{2max}}$) in 1,163 Finnish men (median age, 54 years) from the general population where the 12-month LTPA correlated with $V_{O_{2max}}$ ($r=0.275$). This questionnaire estimates the duration (hours and minutes per session), frequency, (number of sessions per month) and mean intensity (four categories of activity scored as 0 for recreational activity, 1 for conditioning activity, 2 for brisk conditioning activity, and 3 for competitive, strenuous exercise) of LTPA during the previous 12 months. The
intensity of physical activity was expressed in metabolic equivalents of resting oxygen consumption (MET) units. A MET unit is defined as the ratio of the metabolic rate during exercise to the metabolic rate at rest. Generally, one MET corresponds to an oxygen uptake of 3.5 ml per kilogram of body weight per minute, or approximately 1 kcal per kilogram of body weight per hour. An activity requiring a threefold increase in metabolism, such as walking would be assigned a MET score of 3, which for a 70 kg adult would correspond with an approximate energy expenditure of 270 kcal per hour. In the present analyses, three activity levels were defined using MET hours per week (MET-h/week) as follows: sedentary group (<10 MET-h/week), moderately active group (10–40 MET-h/week) or very active group (>40 MET-h/week). In both, the MPP and PPP-Botnia studies inverse correlations were evident between physical activity and BMI (kg/m²) or post-challenge glucose concentrations, indicating that both physical activity questionnaires classify the behaviour of physical activity appropriately on average.

In Paper I maximal aerobic fitness (VO₂max) was assessed during an incremental ergometer cycle test to exhaustion on an electronically braked ergometer (Monark 839 Ergomedic, Varberg, Sweden) (324). Initial and incremental workloads were 25 watts (W) for 9-year olds weighing <30 kg and 30 W for heavier children. For 15-year-old boys and girls the initial workloads were 40 and 50W respectively. Workloads were programmed to increase after every 3 min. The criteria used to define maximal exertion were i) heart rate of 185 beats per minute or more and a judgment by the observer that the participant could no longer continue even with verbal encouragement. Aerobic fitness is expressed as maximal power output relative to body weight (W/kg).

### 3.6 Statistical methods

All statistical analyses were conducted using the SAS software version 9.1 (SAS Institute, Carey, North Carolina, USA). A $P$-value $\leq 0.05$ was considered statistically significant. Continuously distributed data are described as means ± SD and categorical variables as number of individuals per category (N). Statistical power was calculated using QUANTO v1.2.3 (Paper I, III). The purpose of these power calculations was to determine the probability that a given test will reject the null hypothesis when the alternative hypothesis was true. In Paper I statistical power was calculated assuming plausible effect sizes which were estimated from previous publications in paediatric cohorts (325). These were ranked as standardized beta coefficients of 0.005, 0.01 and 0.015% phenotypic difference per copy of the effect allele for a given dependent variable and for a range of minor allele frequencies (0.05–0.45). In Paper, III, statistical power was calculated with the background population disease risk set at 5% and assuming a detectable interaction effect size (RGE) ranging between OR 1.25–1.50 per copy of the minor allele (326). Statistical power was not calculated for Papers II or IV because these studies were seeking to replicate previous observations derived from smaller studies.

Body mass index was standardised to age and sex using the methods described by Cole) in order to allow comparison of observations from different normal distributions (Papers I, II). Glucose was also standardised (Paper II).

The Shapiro-Wilk test for normality was used to examined if the quantitative dependent variables were normal distributed (Papers I,II). Variables with skewed distributions were logarithmically transformed and the
normality of the transformed distribution was checked. The Pearson product-moment correlation coefficient was used to examine the strength and direction of the relationship between physical activity, BMI and glucose (Paper IV). The two-sided independent samples Student’s t test was performed in order to compare means between groups (e.g. age, sex, or country groups). The equality of variance for the two sample distributions was checked to determine whether to select the Satterthwaite or pooled samples t-test P-values. A likelihood ratio test with 1 degree of freedom was used to calculate the difference between proportions for categorical traits such as sex (men/women), glucose regulation (impaired glucose regulation vs. normal glucose regulation), and diabetes (no/yes) (Paper III).

Generalised linear models were used to test genotype associations and interactions for models where the outcome trait was express as a continuous variable, such as BMI (Papers I, II, IV) glucose (Papers I-III), waist circumference (Papers I, II), systolic and diastolic blood pressure, aerobic fitness (Paper I), body fat percentage, HOMA-S, HOMA-B, and height (Paper II). For these models, beta-coefficients and S.E. are provided. Unconditional logistic regression analyses were used to test genotype associations and interactions for models where glucose tolerance was express as a binary variable (impaired glucose regulation vs. normal glucose regulation) (Paper III). Logistic regression analyses were also used to test gene x physical activity interactions on a categorical obesity variable [obese (≥ 30 kg/m²) versus non obese (< 30 kg/m²)]. For these models, OR and 95% confidence intervals are provided. Cox proportional hazards models, a regression method for modelling survival times that assume a maximum likelihood, was used to test T2D cumulative incidence stratified by level of physical activity and genotype (Paper III). Cumulative incidence was calculated within each genotype during a median follow-up of 24.5 years in order to analyse the shape and nature of the underlying survival function. For these models, hazard rate ratios and 95% CIs are provided in addition to stratified cumulative incidence plots. To assess the putative effect modifying roles of age (Paper I), country (Paper I) or physical activity (Papers I-IV), two-way interaction terms and their marginal effects were fitted to the linear and logistic regression models. For the Cox models, a dummy variable for the interaction term was calculated and fitted to the model.

Potential confounders were incorporated as covariates in the regression models in order to reduce bias and extrinsic variance (error). These included age, sex and BMI (Papers I-IV). In addition, generalized linear models were adjusted for country, sexual maturity, age group, height and weight (Papers I, II).

Adjustments for multiple statistical comparisons were made using the Holm procedure (Papers I, III). This correction method was applied to each dependent variable separately. This procedure requires that the probability statistics for each hypothesis test (P-value ≤ 0.05) is placed in rank order, with the highest P-value appearing first. The denominator represents the total number of statistically significant P-values within that rank-ordered list. Thus, the least significant hypothesis test is divisible by 1, the second least significant hypothesis test by 2 and so on until all P-values are corrected.

All models assumed an additive mode of genetic inheritance (Papers I-V). Where a given SNP had genotypes of AA (major homozygotes), Aa (heterozygotes) and aa (minor allele homozygotes), the genotypes were coded as AA=0, Aa=1 and aa=2 in an additive model.
4 RESULTS

4.1 Paper I

To capture all common variation in the PPARGC1A gene 35 tag-SNPs were selected (see Table 4) and their main genetic effects (and in a subset, gene x physical activity interactions) on cardiovascular and metabolic traits were assessed in a Danish and Estonian cross-sectional cohort consisting of 2,101 children and adolescents from the EYHS.

Table 4 Extent to which the 35 PPARGC1A tagging single nucleotide polymorphisms (tag SNPs) capture the PPARGC1A SNPs available in the latest release of the phase 2 HAPMAP (release 24; NCBI build 36).

<table>
<thead>
<tr>
<th>HAPMAP Tag SNP</th>
<th>Tag SNP</th>
<th>r²</th>
<th>HAPMAP Tag SNP</th>
<th>Tag SNP</th>
<th>r²</th>
<th>HAPMAP Tag SNP</th>
<th>Tag SNP</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1002477</td>
<td>rs1002477</td>
<td>1</td>
<td>rs2932976</td>
<td>rs2932965</td>
<td>0.91</td>
<td>rs3755863</td>
<td>rs3755863</td>
<td>1</td>
</tr>
<tr>
<td>rs1002521</td>
<td>rs7656250</td>
<td>0.85</td>
<td>rs2932964</td>
<td>rs1388332</td>
<td>1</td>
<td>rs3774902</td>
<td>rs3774902</td>
<td>1</td>
</tr>
<tr>
<td>rs1002590</td>
<td>rs6850464</td>
<td>1</td>
<td>rs2305681</td>
<td>rs3774909</td>
<td>1</td>
<td>rs3774907</td>
<td>rs2932965</td>
<td>0.92</td>
</tr>
<tr>
<td>rs1000750</td>
<td>rs10018239</td>
<td>1</td>
<td>rs2932966</td>
<td>rs2932976</td>
<td>1</td>
<td>rs3774908</td>
<td>rs31988332</td>
<td>1</td>
</tr>
<tr>
<td>rs10018239</td>
<td>rs10018239</td>
<td>1</td>
<td>rs2932968</td>
<td>rs2932965</td>
<td>0.91</td>
<td>rs3774909</td>
<td>rs3774909</td>
<td>1</td>
</tr>
<tr>
<td>rs10028665</td>
<td>rs17637318</td>
<td>0.89</td>
<td>rs2932970</td>
<td>rs2932976</td>
<td>1</td>
<td>rs3774921</td>
<td>not tagged</td>
<td>-</td>
</tr>
<tr>
<td>rs10213440</td>
<td>rs4452416</td>
<td>0.81</td>
<td>rs2932971</td>
<td>rs2932976</td>
<td>1</td>
<td>rs3796407</td>
<td>not tagged</td>
<td>-</td>
</tr>
<tr>
<td>rs10938663</td>
<td>rs3755863</td>
<td>1</td>
<td>rs2932974</td>
<td>rs2932965</td>
<td>0.91</td>
<td>rs3821952</td>
<td>rs3821952</td>
<td>1</td>
</tr>
<tr>
<td>rs11724368</td>
<td>rs11724368</td>
<td>1</td>
<td>rs2932975</td>
<td>rs2932976</td>
<td>0.93</td>
<td>rs4233508</td>
<td>rs12645360</td>
<td>1</td>
</tr>
<tr>
<td>rs11734408</td>
<td>rs17637318</td>
<td>1</td>
<td>rs2932976</td>
<td>rs2932976</td>
<td>1</td>
<td>rs4361373</td>
<td>rs4361373</td>
<td>1</td>
</tr>
<tr>
<td>rs11941854</td>
<td>not tagged</td>
<td>-</td>
<td>rs2932977</td>
<td>rs2932965</td>
<td>0.96</td>
<td>rs4383605</td>
<td>rs17576121</td>
<td>1</td>
</tr>
<tr>
<td>rs12374310</td>
<td>not tagged</td>
<td>-</td>
<td>rs2946385</td>
<td>rs2946385</td>
<td>1</td>
<td>rs4452416</td>
<td>rs4452416</td>
<td>1</td>
</tr>
<tr>
<td>rs12374408</td>
<td>not tagged</td>
<td>-</td>
<td>rs2946386</td>
<td>rs2907089</td>
<td>1</td>
<td>rs4469064</td>
<td>rs4469064</td>
<td>1</td>
</tr>
<tr>
<td>rs12500214</td>
<td>rs12500214</td>
<td>1</td>
<td>rs2970847</td>
<td>rs2970847</td>
<td>1</td>
<td>rs4550905</td>
<td>rs13131226</td>
<td>1</td>
</tr>
<tr>
<td>rs12503529</td>
<td>rs12500214</td>
<td>0.81</td>
<td>rs2970848</td>
<td>rs2970848</td>
<td>1</td>
<td>rs4697045</td>
<td>not tagged</td>
<td>-</td>
</tr>
<tr>
<td>rs2640088</td>
<td>rs4452416</td>
<td>1</td>
<td>rs29070849</td>
<td>rs2970848</td>
<td>0.88</td>
<td>rs4697426</td>
<td>not tagged</td>
<td>-</td>
</tr>
<tr>
<td>rs12645360</td>
<td>rs12645360</td>
<td>1</td>
<td>rs2970851</td>
<td>rs2932976</td>
<td>1</td>
<td>rs6448227</td>
<td>rs2932965</td>
<td>0.96</td>
</tr>
<tr>
<td>rs12650562</td>
<td>rs3755863</td>
<td>0.82</td>
<td>rs2970852</td>
<td>rs2932976</td>
<td>1</td>
<td>rs6448228</td>
<td>rs12645360</td>
<td>1</td>
</tr>
<tr>
<td>rs13171712</td>
<td>rs13171712</td>
<td>1</td>
<td>rs2970853</td>
<td>not tagged</td>
<td>-</td>
<td>rs6850464</td>
<td>rs6850464</td>
<td>1</td>
</tr>
<tr>
<td>rs13131226</td>
<td>rs13131226</td>
<td>1</td>
<td>rs2970855</td>
<td>not tagged</td>
<td>-</td>
<td>rs7437482</td>
<td>rs10018239</td>
<td>1</td>
</tr>
<tr>
<td>rs1388332</td>
<td>rs1388332</td>
<td>1</td>
<td>rs2970856</td>
<td>rs2932965</td>
<td>0.92</td>
<td>rs7656250</td>
<td>rs7656250</td>
<td>1</td>
</tr>
<tr>
<td>rs1472095</td>
<td>rs2970876</td>
<td>1</td>
<td>rs2970866</td>
<td>rs2970869</td>
<td>1</td>
<td>rs7657071</td>
<td>rs7657071</td>
<td>1</td>
</tr>
<tr>
<td>rs16874194</td>
<td>rs1388332</td>
<td>0.85</td>
<td>rs2970869</td>
<td>rs2970869</td>
<td>1</td>
<td>rs7675317</td>
<td>rs6850464</td>
<td>1</td>
</tr>
<tr>
<td>rs16874268</td>
<td>rs4469064</td>
<td>0.88</td>
<td>rs2970870</td>
<td>rs2970870</td>
<td>1</td>
<td>rs7665116</td>
<td>rs7665116</td>
<td>1</td>
</tr>
<tr>
<td>rs16874271</td>
<td>rs4469064</td>
<td>0.88</td>
<td>rs2970871</td>
<td>not tagged</td>
<td>-</td>
<td>rs7672915</td>
<td>not tagged</td>
<td>-</td>
</tr>
<tr>
<td>rs16874276</td>
<td>rs4469064</td>
<td>0.87</td>
<td>rs2970872</td>
<td>not tagged</td>
<td>-</td>
<td>rs7674429</td>
<td>rs7674429</td>
<td>1</td>
</tr>
<tr>
<td>rs17576121</td>
<td>rs17576121</td>
<td>1</td>
<td>rs2970873</td>
<td>not tagged</td>
<td>-</td>
<td>rs7670000</td>
<td>rs6850464</td>
<td>0.81</td>
</tr>
<tr>
<td>rs17637318</td>
<td>rs17637318</td>
<td>1</td>
<td>rs2970883</td>
<td>rs2970883</td>
<td>1</td>
<td>rs7682765</td>
<td>rs7682765</td>
<td>1</td>
</tr>
<tr>
<td>rs1873532</td>
<td>rs3755863</td>
<td>1</td>
<td>rs2970884</td>
<td>rs2970884</td>
<td>1</td>
<td>rs768665</td>
<td>not tagged</td>
<td>-</td>
</tr>
<tr>
<td>rs1878949</td>
<td>not tagged</td>
<td>-</td>
<td>rs3736265</td>
<td>rs3736265</td>
<td>1</td>
<td>rs8192678</td>
<td>rs8192678</td>
<td>1</td>
</tr>
<tr>
<td>rs2279525</td>
<td>not tagged</td>
<td>-</td>
<td>rs3755862</td>
<td>rs3736265</td>
<td>1</td>
<td>rs9996943</td>
<td>not tagged</td>
<td>-</td>
</tr>
<tr>
<td>rs2932965</td>
<td>rs2932965</td>
<td>1</td>
<td>rs3774923</td>
<td>not tagged</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Using pairwise tagging, the above tag SNPs captured 77 of 95 alleles at \( r^2 > 0.80 \), of which 81% have a mean \( r^2 = 0.971 \) in the latest HAPMAP release.
release. This compares with a near complete capture rate when using the earlier version of HAPMAP (release 19, NCBI build 34), which was available when this study was in progress.

Significant differences ($p<0.01$) by age-group and by sex were observed for all variables shown below in Table 5. Thus, analyses were performed include adjustments to account for the heterogeneity that might be introduced by these differences (see below).

**Table 5** Participant characteristics by age group and sex in the EYHS

<table>
<thead>
<tr>
<th>Variables</th>
<th>Children ($n=1,255$)</th>
<th>Adolescents ($n=846$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boys ($n=599$)</td>
<td>Girls ($n=656$)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>9.7(0.4)</td>
<td>9.6(0.4)</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>138.9(6.6)</td>
<td>138.7(6.7)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>33.2(6.1)</td>
<td>33.1(6.8)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>17.1(2.2)</td>
<td>17.1(2.6)</td>
</tr>
<tr>
<td>Overweight (n)</td>
<td>91</td>
<td>87</td>
</tr>
<tr>
<td>Obese (n)¹</td>
<td>38</td>
<td>41</td>
</tr>
<tr>
<td>Waist circumference (cm)</td>
<td>59.4(5.5)</td>
<td>58.3(6.6)</td>
</tr>
<tr>
<td>Diastolic BP (mmHg)</td>
<td>60.2(7)</td>
<td>60.1(6.3)</td>
</tr>
<tr>
<td>Systolic BP (mmHg)</td>
<td>102.7(8.9)</td>
<td>101(8.5)</td>
</tr>
<tr>
<td>Aerobic fitness (W/kg)</td>
<td>3.2(0.5)</td>
<td>2.8(0.5)</td>
</tr>
<tr>
<td>Plasma glucose (mmol/l)</td>
<td>5.1(0.36)</td>
<td>4.97(0.37)</td>
</tr>
<tr>
<td>Plasma insulin (pmol/l)</td>
<td>40.2(22.06)</td>
<td>47.92(30.66)</td>
</tr>
<tr>
<td>Total cholesterol (mmol/l)</td>
<td>4.44(0.73)</td>
<td>4.52(0.77)</td>
</tr>
<tr>
<td>Triacylglycerol (mmol/l)</td>
<td>0.72(0.33)</td>
<td>0.81(0.33)</td>
</tr>
<tr>
<td>HOMA-Bb</td>
<td>129.44(85.65)</td>
<td>143.12(90.92)</td>
</tr>
<tr>
<td>HOMA-Sb</td>
<td>100.23(62.51)</td>
<td>94.05(59.63)</td>
</tr>
<tr>
<td>HDL-cholesterol (mmol/l)</td>
<td>1.57(0.36)</td>
<td>1.47(0.29)</td>
</tr>
<tr>
<td>LDL-cholesterol (mmol/l)</td>
<td>2.54(0.61)</td>
<td>2.68(0.66)</td>
</tr>
<tr>
<td>Physical activity (counts/min)</td>
<td>742.3(234.7)</td>
<td>613.4(188.1)</td>
</tr>
</tbody>
</table>

Data are means (±SD)

¹ The number of individuals with BMI between the 85th and 95th (overweight), and ≥95th (obese) percentiles, respectively


HOMA-B, homeostasis model of beta cell function; HOMA-S, homeostasis model of insulin sensitivity

Aerobic fitness, the ability of the circulatory and respiratory systems to supply oxygen during sustained physical activity

Having adjusted for age, sex, age-group, and sexual maturity, several nominally significant associations were observed for BMI (rs10018239: beta=-0.06 kg/m² per minor allele copy, $P=0.039$), waist circumference (rs10018239: beta=-0.01 cm per minor allele copy, $P=0.043$; rs7656250: beta=0.01 cm per minor allele copy, $P=0.012$; rs8192678 [Gly482Ser] beta=0.01 cm per minor allele copy, $P=0.015$; rs3755863: beta=-0.01 cm per minor allele copy, $P=0.02$, SBP (rs2970869: beta=0.77 mmHg per minor allele copy, $P=0.018$), fasting glucose level (rs11724368: beta=-0.01 mmol/l per minor allele copy, $P=0.045$). More robust associations were observed for aerobic fitness (rs7656250: beta=-0.06 W/kg per minor allele copy, $P=0.005$; rs13117172: beta=0.06 W/kg per minor allele copy, $P=0.008$), and fasting glucose levels (rs7657071: beta=-0.01 mmol/l per minor allele copy, $P=0.002$).

After controlling for multiple statistical comparisons (each phenotype separately), none of the observed associations listed above remained statistically significant. The strongest corrected $p$-value was for rs7657071 ($P_{corrected}=0.07$).

See Tables 6a and 6b for results from all tests.
**Table 6a** Results for association tests between *PPARGC1A* genotypes and measures of body composition and aerobic fitness.

<table>
<thead>
<tr>
<th>Polymorphisms</th>
<th>Phenotypes</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BMI (kg/m²)</td>
<td>Waist circumference (cm)</td>
<td>Aerobic fitness (Watts/kg)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reference sequence</td>
<td>Sequence variation</td>
<td>Beta</td>
<td>P-value</td>
<td>Beta</td>
<td>P-value</td>
</tr>
<tr>
<td>rs10938972</td>
<td>-5452 A&gt;G</td>
<td>-0.045</td>
<td>0.314</td>
<td>-0.007</td>
<td>0.104</td>
<td>0.030</td>
</tr>
<tr>
<td>rs7657071</td>
<td>-4824 A&gt;C</td>
<td>0.010</td>
<td>0.757</td>
<td>-0.001</td>
<td>0.771</td>
<td>0.008</td>
</tr>
<tr>
<td>rs2970869</td>
<td>-1719 A&gt;G</td>
<td>-0.022</td>
<td>0.542</td>
<td>0.000</td>
<td>0.896</td>
<td>0.003</td>
</tr>
<tr>
<td>rs17576121</td>
<td>-1624 C&gt;T</td>
<td>-0.032</td>
<td>0.305</td>
<td>-0.004</td>
<td>0.210</td>
<td>0.018</td>
</tr>
<tr>
<td>rs2970870</td>
<td>-1367 C&gt;T</td>
<td>-0.001</td>
<td>0.975</td>
<td>0.003</td>
<td>0.268</td>
<td>-0.009</td>
</tr>
<tr>
<td>rs3774909</td>
<td>IVS1 +745 C&gt;T</td>
<td>-0.068</td>
<td>0.319</td>
<td>-0.010</td>
<td>0.103</td>
<td>0.049</td>
</tr>
<tr>
<td>rs2946385</td>
<td>IVS2 +52 G&gt;T</td>
<td>-0.053</td>
<td>0.092</td>
<td>-0.005</td>
<td>0.109</td>
<td>0.012</td>
</tr>
<tr>
<td>rs12500214</td>
<td>IVS2 +807 A&gt;G</td>
<td>-0.007</td>
<td>0.860</td>
<td>-0.002</td>
<td>0.625</td>
<td>-0.018</td>
</tr>
<tr>
<td>rs4469064</td>
<td>IVS2 + A&gt;G</td>
<td>0.086</td>
<td>0.101</td>
<td>0.004</td>
<td>0.397</td>
<td>-0.017</td>
</tr>
<tr>
<td>rs7674429</td>
<td>IVS2 +5027 C&gt;T</td>
<td>-0.018</td>
<td>0.665</td>
<td>-0.006</td>
<td>0.144</td>
<td>0.058</td>
</tr>
<tr>
<td>rs17637318</td>
<td>IVS2 +5390 C&gt;T</td>
<td>-0.034</td>
<td>0.285</td>
<td>-0.003</td>
<td>0.804</td>
<td>0.029</td>
</tr>
<tr>
<td>rs13117172</td>
<td>IVS2 +5860 C&gt;T</td>
<td>-0.017</td>
<td>0.684</td>
<td>-0.006</td>
<td>0.153</td>
<td>0.060</td>
</tr>
<tr>
<td>rs4361373</td>
<td>IVS2 +10700 C&gt;T</td>
<td>-0.041</td>
<td>0.302</td>
<td>-0.005</td>
<td>0.158</td>
<td>0.009</td>
</tr>
<tr>
<td>rs13131226</td>
<td>IVS2 +11394 C&gt;T</td>
<td>0.041</td>
<td>0.207</td>
<td>0.000</td>
<td>0.992</td>
<td>-0.017</td>
</tr>
<tr>
<td>rs7656250</td>
<td>IVS2 +20359 C&gt;T</td>
<td>0.072</td>
<td>0.051</td>
<td>0.008</td>
<td>0.012</td>
<td>-0.057</td>
</tr>
<tr>
<td>rs4452416</td>
<td>IVS2 +22502 G&gt;T</td>
<td>-0.020</td>
<td>0.671</td>
<td>-0.003</td>
<td>0.493</td>
<td>0.014</td>
</tr>
<tr>
<td>rs11724368</td>
<td>IVS2 +23442 A&gt;T</td>
<td>0.043</td>
<td>0.230</td>
<td>0.004</td>
<td>0.177</td>
<td>0.001</td>
</tr>
<tr>
<td>rs12645360</td>
<td>IVS2 +25539 A&gt;G</td>
<td>-0.045</td>
<td>0.149</td>
<td>-0.004</td>
<td>0.201</td>
<td>-0.003</td>
</tr>
<tr>
<td>rs6859046</td>
<td>IVS2 +26893 A&gt;G</td>
<td>-0.006</td>
<td>0.893</td>
<td>-0.004</td>
<td>0.345</td>
<td>-0.008</td>
</tr>
<tr>
<td>rs10002477</td>
<td>IVS2 +29331 A&gt;G</td>
<td>-0.024</td>
<td>0.423</td>
<td>0.000</td>
<td>0.891</td>
<td>-0.004</td>
</tr>
<tr>
<td>rs10018239</td>
<td>IVS2 +31484 A&gt;G</td>
<td>-0.063</td>
<td>0.039</td>
<td>-0.006</td>
<td>0.043</td>
<td>0.028</td>
</tr>
<tr>
<td>rs7665116</td>
<td>IVS2 +33634 C&gt;T</td>
<td>-0.052</td>
<td>0.255</td>
<td>-0.004</td>
<td>0.315</td>
<td>-0.008</td>
</tr>
<tr>
<td>rs1388332</td>
<td>IVS5 +709 C&gt;T</td>
<td>-0.011</td>
<td>0.845</td>
<td>-0.004</td>
<td>0.461</td>
<td>-0.012</td>
</tr>
<tr>
<td>rs2932976</td>
<td>IVS7 +3575 A&gt;G</td>
<td>0.023</td>
<td>0.476</td>
<td>0.002</td>
<td>0.400</td>
<td>-0.021</td>
</tr>
<tr>
<td>rs2970848</td>
<td>IVS7 +8876 A&gt;G</td>
<td>-0.029</td>
<td>0.354</td>
<td>-0.004</td>
<td>0.146</td>
<td>0.014</td>
</tr>
<tr>
<td>rs2970847</td>
<td>Thr394Thr</td>
<td>-0.002</td>
<td>0.965</td>
<td>0.004</td>
<td>0.299</td>
<td>0.017</td>
</tr>
<tr>
<td>rs1912678</td>
<td>Gly482Ser</td>
<td>0.052</td>
<td>0.131</td>
<td>0.008</td>
<td>0.015</td>
<td>-0.011</td>
</tr>
<tr>
<td>rs3755863</td>
<td>Thr528Thr</td>
<td>-0.035</td>
<td>0.280</td>
<td>-0.007</td>
<td>0.020</td>
<td>0.001</td>
</tr>
<tr>
<td>rs3736265</td>
<td>Thr612Met</td>
<td>0.023</td>
<td>0.716</td>
<td>0.008</td>
<td>0.174</td>
<td>0.019</td>
</tr>
<tr>
<td>rs2932965</td>
<td>IVS10 +8884 A&gt;G</td>
<td>0.013</td>
<td>0.719</td>
<td>0.004</td>
<td>0.213</td>
<td>0.013</td>
</tr>
<tr>
<td>rs7682765</td>
<td>IVS12 +146 C&gt;T</td>
<td>-0.037</td>
<td>0.450</td>
<td>-0.006</td>
<td>0.154</td>
<td>0.045</td>
</tr>
<tr>
<td>rs3821952</td>
<td>IVS12 +5620 G&gt;T</td>
<td>-0.076</td>
<td>0.265</td>
<td>-0.011</td>
<td>0.089</td>
<td>0.002</td>
</tr>
<tr>
<td>rs2970884</td>
<td>3513 C&gt;T</td>
<td>0.044</td>
<td>0.179</td>
<td>0.003</td>
<td>0.373</td>
<td>-0.023</td>
</tr>
<tr>
<td>rs2970883</td>
<td>4844 C&gt;T</td>
<td>-0.005</td>
<td>0.885</td>
<td>0.001</td>
<td>0.820</td>
<td>0.030</td>
</tr>
<tr>
<td>rs1491378</td>
<td>5121 A&gt;G</td>
<td>-0.045</td>
<td>0.151</td>
<td>-0.004</td>
<td>0.165</td>
<td>0.026</td>
</tr>
</tbody>
</table>

All models were adjusted for country, gender, age-group, age, and sexual maturity. All beta-coefficients are raw units, except BMI, which was standardised using Cole's LMS method (327).
### Table 6b Results for association tests between PPARGC1A genotypes and blood pressures and fasting glucose concentrations

<table>
<thead>
<tr>
<th>Polymorphisms</th>
<th>Phenotypes</th>
<th>DBP (mmHg)</th>
<th>SBP (mmHg)</th>
<th>Fasting glucose (mmol/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Beta</td>
<td>P-value</td>
<td>Beta</td>
</tr>
<tr>
<td>Reference sequence</td>
<td>Sequence variation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10938972</td>
<td>-5452 A&gt;G</td>
<td>0.464</td>
<td>0.124</td>
<td>0.407</td>
</tr>
<tr>
<td>rs7657071</td>
<td>-4824 A&gt;C</td>
<td>-0.298</td>
<td>0.191</td>
<td>-0.594</td>
</tr>
<tr>
<td>rs2970869</td>
<td>-1719 A&gt;G</td>
<td>0.336</td>
<td>0.162</td>
<td>0.771</td>
</tr>
<tr>
<td>rs17576121</td>
<td>-1624 C&gt;T</td>
<td>0.178</td>
<td>0.401</td>
<td>0.206</td>
</tr>
<tr>
<td>rs2970870</td>
<td>-1367 C&gt;T</td>
<td>0.105</td>
<td>0.621</td>
<td>0.426</td>
</tr>
<tr>
<td>rs3774909</td>
<td>IVS1 +745 C&gt;T</td>
<td>0.648</td>
<td>0.164</td>
<td>0.472</td>
</tr>
<tr>
<td>rs2946385</td>
<td>IVS2 +52 G&gt;T</td>
<td>0.027</td>
<td>0.899</td>
<td>0.079</td>
</tr>
<tr>
<td>rs12500214</td>
<td>IVS2 +807 A&gt;G</td>
<td>-0.329</td>
<td>0.204</td>
<td>-0.513</td>
</tr>
<tr>
<td>rs4469064</td>
<td>IVS2 + A&gt;G</td>
<td>-0.199</td>
<td>0.578</td>
<td>0.452</td>
</tr>
<tr>
<td>rs7674429</td>
<td>IVS2 +5027 C&gt;T</td>
<td>0.139</td>
<td>0.631</td>
<td>0.366</td>
</tr>
<tr>
<td>rs17637318</td>
<td>IVS2 +5390 C&gt;T</td>
<td>0.263</td>
<td>0.218</td>
<td>0.35</td>
</tr>
<tr>
<td>rs13117172</td>
<td>IVS2 +5860 C&gt;T</td>
<td>0.174</td>
<td>0.531</td>
<td>0.4</td>
</tr>
<tr>
<td>rs4361373</td>
<td>IVS2 +10700 C&gt;T</td>
<td>-0.149</td>
<td>0.578</td>
<td>-0.456</td>
</tr>
<tr>
<td>rs13131226</td>
<td>IVS2 +11394 C&gt;T</td>
<td>-0.236</td>
<td>0.289</td>
<td>-0.421</td>
</tr>
<tr>
<td>rs7656250</td>
<td>IVS2 +20359 C&gt;T</td>
<td>-0.261</td>
<td>0.293</td>
<td>-0.515</td>
</tr>
<tr>
<td>rs4452416</td>
<td>IVS2 +22502 G&gt;T</td>
<td>0.543</td>
<td>0.081</td>
<td>0.684</td>
</tr>
<tr>
<td>rs11724368</td>
<td>IVS2 +23442 A&gt;T</td>
<td>-0.119</td>
<td>0.624</td>
<td>-0.294</td>
</tr>
<tr>
<td>rs12645360</td>
<td>IVS2 +26539 A&gt;G</td>
<td>-0.137</td>
<td>0.517</td>
<td>-0.099</td>
</tr>
<tr>
<td>rs6850464</td>
<td>IVS2 +26893 A&gt;G</td>
<td>0.073</td>
<td>0.803</td>
<td>0.151</td>
</tr>
<tr>
<td>rs10002477</td>
<td>IVS2 +29331 A&gt;G</td>
<td>-0.193</td>
<td>0.342</td>
<td>-0.138</td>
</tr>
<tr>
<td>rs10018239</td>
<td>IVS2 +3148 A&gt;G</td>
<td>0.239</td>
<td>0.246</td>
<td>0.354</td>
</tr>
<tr>
<td>rs7665116</td>
<td>IVS2 +33364 C&gt;T</td>
<td>0.049</td>
<td>0.874</td>
<td>-0.51</td>
</tr>
<tr>
<td>rs1388332</td>
<td>IVS5 +709 C&gt;T</td>
<td>-0.074</td>
<td>0.848</td>
<td>-0.215</td>
</tr>
<tr>
<td>rs2932976</td>
<td>IVS7 +3575 A&gt;G</td>
<td>-0.15</td>
<td>0.493</td>
<td>-0.39</td>
</tr>
<tr>
<td>rs2970848</td>
<td>IVS7 +8876 A&gt;G</td>
<td>0.052</td>
<td>0.804</td>
<td>0.302</td>
</tr>
<tr>
<td>rs2970847</td>
<td>Thr394Thr</td>
<td>-0.063</td>
<td>0.804</td>
<td>-0.215</td>
</tr>
<tr>
<td>rs8192678</td>
<td>Gly482Ser</td>
<td>-0.348</td>
<td>0.136</td>
<td>-0.386</td>
</tr>
<tr>
<td>rs3753863</td>
<td>Thr528Thr</td>
<td>0.048</td>
<td>0.826</td>
<td>0.103</td>
</tr>
<tr>
<td>rs3736265</td>
<td>Thr612Met</td>
<td>0.112</td>
<td>0.794</td>
<td>0.062</td>
</tr>
<tr>
<td>rs2932965</td>
<td>IVS10 +8884 A&gt;G</td>
<td>0.066</td>
<td>0.79</td>
<td>0.012</td>
</tr>
<tr>
<td>rs7682765</td>
<td>IVS12 +146 C&gt;T</td>
<td>0.288</td>
<td>0.383</td>
<td>0.137</td>
</tr>
<tr>
<td>rs381952</td>
<td>IVS12 +5620 G&gt;T</td>
<td>-0.126</td>
<td>0.786</td>
<td>0.232</td>
</tr>
<tr>
<td>rs2970884</td>
<td>3513 C&gt;T</td>
<td>0.005</td>
<td>0.982</td>
<td>-0.028</td>
</tr>
<tr>
<td>rs2970883</td>
<td>4844 C&gt;T</td>
<td>-0.045</td>
<td>0.86</td>
<td>-0.08</td>
</tr>
<tr>
<td>rs1491378</td>
<td>5214 A&gt;G</td>
<td>-0.119</td>
<td>0.579</td>
<td>-0.022</td>
</tr>
</tbody>
</table>

All models are adjusted for country, gender, age-group, age and sexual maturity. All beta-coefficients are for raw units. In addition, blood pressures are adjusted for height and weight and glucose for analysis centre. SBP = systolic blood pressure; DBP = diastolic blood pressure.

The non-synonymous Gly482Ser SNP is a widely studied variant. Therefore, we show here the associations between this polymorphism and a range of previously studied cardiovascular and metabolic traits which we did not examine in relation to other PPARGC1A SNPs owing to insufficient power.
(see Table 7). In these analyses, waist circumference was the only trait that showed a nominally statistical significant association ($P_{uncorrected}=0.02$).

**Table 7** Summary of association analyses for the *PPARGC1A* Gly482Ser polymorphism

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>Phenotypes</th>
<th>Gly/Gly</th>
<th>Gly/Ser</th>
<th>Ser/Ser</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BMI (kg/m$^2$)</td>
<td>0.08(0.03)</td>
<td>0.12(0.03)</td>
<td>0.2(0.07)</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Waist circumference (cm)</td>
<td>62.03(0.18)</td>
<td>62.27(0.19)</td>
<td>63.3(0.43)</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Diastolic BP (mmHg)</td>
<td>61.72(0.22)</td>
<td>61.46(0.23)</td>
<td>60.92(0.5)</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>Systolic BP (mmHg)</td>
<td>105.81(0.29)</td>
<td>105.51(0.31)</td>
<td>104.92(0.68)</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>Aerobic fitness (W/kg)</td>
<td>3.07(0.02)</td>
<td>3.06(0.02)</td>
<td>3.06(0.04)</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Glucose (mmol/l)</td>
<td>5.05(0.01)</td>
<td>5.05(0.01)</td>
<td>5.1(0.03)</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Data are means (±SD)

All results were adjusted for country, sex, age group, age and maturity. Diastolic BP and systolic BP were also adjusted for height and weight, glucose was additionally adjusted for analysis centre BMI was standardised using Cole’s LMS method (327).

Interaction terms (gene * physical activity, age, or country) were tested for each of the polymorphisms that showed nominal evidence of association in the main effect models. None of these interaction tests approached statistical significance. We did not perform interaction tests for variants that did not show at least nominal evidence of association in the main effect models, as this would have substantially raised the risk of false discovery (type 1 error).

In summary, we observed limited support for the hypothesis that variants at the *PPARGC1A* gene influence metabolic or cardiovascular risk factors in this cohort of European children. The most promising results were for the rs7657071 and rs7656250 SNPs, which merit further investigation.

**4.2 Paper II**

In this study, we investigated the main genetic effects and gene x physical activity interactions of the *TCF7L2* rs7903146 variant and T2D-related traits in a Danish and Estonian cross-sectional paediatric cohort from the EYHS. Participant characteristics are shown in Table 8.

**Table 8** Participant characteristics by age group and sex in the EYHS

<table>
<thead>
<tr>
<th>Variables</th>
<th>Children ($n=1,255$)</th>
<th>Adolescents ($n=846$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boys ($n=599$)</td>
<td>Girls ($n=656$)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>9.7(0.4)</td>
<td>9.6(0.4)</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>138.9(6.6)</td>
<td>138.7(6.7)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>33.2(6.1)</td>
<td>33.4(6.8)</td>
</tr>
<tr>
<td>BMI (Kg/m$^2$)</td>
<td>17.1(2.2)</td>
<td>17.1(2.6)</td>
</tr>
<tr>
<td>Overweight (n)$^a$</td>
<td>91</td>
<td>87</td>
</tr>
<tr>
<td>Obese (n)$^b$</td>
<td>38</td>
<td>41</td>
</tr>
<tr>
<td>Waist circumference (cm)</td>
<td>59.4(5.5)</td>
<td>58.3(6.6)</td>
</tr>
<tr>
<td>Plasma glucose (mmol/l)</td>
<td>5.1(0.36)</td>
<td>4.97(0.37)</td>
</tr>
<tr>
<td>Plasma nulin (pmol/l)</td>
<td>40.2(22.06)</td>
<td>47.92(30.66)</td>
</tr>
<tr>
<td>HOMA-B$^b$</td>
<td>129.44(85.65)</td>
<td>143.12(90.92)</td>
</tr>
<tr>
<td>HOMA-S$^b$</td>
<td>100.23(62.51)</td>
<td>94.05(59.03)</td>
</tr>
<tr>
<td>Physical activity (counts/min)</td>
<td>742.3(234.7)</td>
<td>613.4(188.1)</td>
</tr>
</tbody>
</table>

Data are means (±SD)
All the models were adjusted for country, age, sex, age-group, and sexual maturity. The minor T allele at the rs7903146 variant was associated with higher glucose levels in older (beta=−0.098 mmol/l per minor allele copy, \( P=0.029 \)) but not younger children (beta=−0.001 mmol/l per minor allele copy, \( P=0.972 \)). A significant inverse association between the minor allele at rs7903146 and height was evident in boys (beta=−1.073 cm per minor allele copy, \( P=0.001 \)), but not in girls. The test of interaction between the TCF7L2 rs7903146 variant and physical activity on HOMA-B was nominally statistically significant (beta=0.022, \( P_{\text{interaction}}=0.015 \)), whereby physical activity reduced the effect of the risk allele on estimated beta-cell function (Figure 7).

![Figure 7](image)

**Figure 7** Interactions between TCF7L2 rs7903146 variant and physical activity on Homa-B. Quintiles of physical activity from the least active to the most active.

In summary, these findings suggest that physical activity modifies the effects of the TCF7L2 rs7903146 variant on HOMA-B in a cohort of healthy European children.

### 4.3 Paper III

Main genetic effects and gene x physical activity interactions for 17 previously associated T2D gene variants were assessed. Dependent variables were impaired glucose regulation (IGR) and incidence of T2D. The cohort used in these analyses was a prospective population-based collection of 16,003 initially non-diabetic middle-aged Swedish adults from the Malmö Preventive Project (MPP). Table 9 reports participant characteristics.
Table 9 Participant characteristics stratified by level of physical activity (n=16,003)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Physically inactive (n=3,455)</th>
<th>Physically active (n=12,548)</th>
<th>P-difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N or mean (SD)</td>
<td>N or mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Sex (N=M/F)</td>
<td>2.287/1.168</td>
<td>8.115/4.433</td>
<td>0.097</td>
</tr>
<tr>
<td>Baseline age (years)</td>
<td>44.7 (7.3)</td>
<td>45.7 (6.8)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Baseline BMI (kg/m²)</td>
<td>24.6 (3.7)</td>
<td>24.2 (3.1)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Baseline fasting glucose (mmol/l)</td>
<td>4.87 (0.48)</td>
<td>4.81 (0.49)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Baseline two hour glucose (mmol/l)*</td>
<td>5.67 (1.54)</td>
<td>5.64 (1.44)</td>
<td>0.480</td>
</tr>
<tr>
<td>Baseline glucose regulation (N=NGR/IGR)</td>
<td>2.585/870</td>
<td>9.601/2.947</td>
<td>0.038</td>
</tr>
<tr>
<td>Developed diabetes (N=no/yes)</td>
<td>2.958/497</td>
<td>10.982/1.566</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Data are N or means ± SD. Differences between means were calculated using a two-sided independent-samples Student’s t test. Differences between proportions were tested using a likelihood ratio test with 1 degree of freedom. *Available in subsamples of n=1,647 (physically inactive) and n=7,217 (physically active).

After adjustment for age, sex, and BMI, 10 of the 17 polymorphisms showed evidence of association with IGR and/or two hour glucose concentrations, in accordance with previously reports. These variants were: SLC30A8 (rs13266634), TCF7L2 (rs7903146), CDKAL1 (rs7754840), NOTCH2 (rs10923931), KCNJ11 (rs5912), IGFBP2 (rs4402960), JAZF1 (rs864745), HHEX (rs1111875), MTNR1B (rs10830963) and TSPAN8 (rs7961581) (see Paper III). No evidence of gene x physical activity interactions for IGR risk or two hour glucose concentrations emerged for the majority of the polymorphisms tested (see Paper III). The exceptions were HNF1B (rs4430796), CDKN2A/B rs1081661, and PPARG (rs1801282). The following sections describe these results in detail.

4.3.1 Interactions with physical activity on IGR and 2 hour glucose levels

The gene x physical activity interaction terms for HNF1B (rs4430796) on IGR risk and two hour glucose concentrations were both statistically significant (Pinteraction=0.026 and 0.0009, respectively). The latter remained significant after correction for multiple hypothesis testing (Pinteraction=0.015). As shown in Figure 8, the minor ‘A’ allele was associated with lower two hour glucose levels (beta=-0.13 mmol/l per allele; P=0.005) in physically inactive individuals, whereas the opposite effect was observed in physically active individuals (beta=0.04 mmol/l per allele; P=0.056). Concordantly, the minor ‘A’ allele tended to be associated with lower risk of IGR in physically inactive individuals (OR: 0.92 per allele; 95% CI 0.82-1.03; P=0.13) and with increased risk in physically active individuals (OR: 1.06 per allele; 95% CI 1.00-1.12; P=0.066).
The interaction terms for the **CDKN2A/B** rs10811661 polymorphism and physical activity with IGR risk or two hour glucose concentrations as outcomes were both nominally statistically significant ($P_{\text{interaction}} = 0.015$ and 0.013, respectively). The minor allele at rs10811661 tended to be associated with higher risk of IGR in physically inactive individuals (OR: 1.10 per allele; 95% CI 0.95-1.28; $P=0.20$) ($n=3,468$), but was protective of IGR risk in active individuals (OR: 0.89 per allele; 95% CI 0.82-0.97; $P=0.0075$) ($n=12,525$). Similarly, in physically inactive individuals ($n=1,618$) the minor allele tended to be associated with elevated two hour glucose concentrations (beta=0.12 mmol/l per allele; $P=0.064$), whereas in physically active individuals ($n=6,680$) a contrasting effect was observed (beta=–0.06 mmol/l per allele; $P=0.070$), although neither main effect was statistically significant.

The interaction between the **PPARG** rs1801282 polymorphism and physical activity on IGR risk was nominally statistically significant ($P_{\text{interaction}} = 0.04$). In inactive individuals, the minor allele appeared to be protective of IGR (OR: 0.88 per allele; 95% CI 0.75-1.02; $P=0.097$), whereas this effect was reversed in active individuals (OR: 1.05 per allele; 95% CI 0.97-1.15; $P=0.22$), albeit neither stratified effect was statistically significant. No evidence of statistical interaction was observed for two hour glucose concentrations ($P_{\text{interaction}}=0.78$).

### 4.3.2 Interactions with physical activity on T2D incidence

No statistical evidence of gene x physical activity interaction was observed on the incidence of T2D when stratified by level of physical activity (active versus inactive) for the majority of the polymorphisms tested (Figure 9). **HNF1B** (rs4430796) was the only exception. In this case, carriers of the risk allele who were physically active conveyed an increased risk for T2D relative to carriers of the risk allele who were physically inactive (uncorrected $P_{\text{interaction}}=0.0004$). This interaction effect remained statistically significant after adjusting for multiple hypothesis testing (corrected $P_{\text{interaction}}=0.0068$).
Figure 9 Gene x physical activity interaction effects for 17 previously associated T2D variants. Effects are ranked by the statistical significance of the interaction effect. Data are hazard rate ratios (95% CI) stratified by level of physical activity (inactive vs. active). The median follow-up duration was 24.5 yrs.

Figure 10 shows T2D cumulative incidence plots stratified by genotype at the \textit{HNF1B} (rs4430796) locus. The panel A shows major allele homozygotes. As anticipated, physically inactive individuals were at increased risk of developing T2D compared with active individuals. However, with each additional copy of the minor allele (panels B and C), the protective effect of baseline physical activity is progressively diminished.
Figure 10 T2D cumulative incidence plots stratified by level of physical activity and genotype at the HNF1B rs4430796 locus. Panel A: G/G. Panel B: G/A. Panel C: A/A. Cumulative incidences are calculated within each genotype group. Follow-up is truncated at the median duration for the cohort (24.5 years).
Because physical activity and BMI are related, interactions between the gene variants and BMI were also assessed. These interactions terms did not approach statistical significance for any the three variants showing evidence of gene x physical activity interactions.

In summary, this work suggests that common variants at *HNF1B* and *CDKN2A/B* interact with physical activity levels to influence the risk of IGR. For the *HNF1B* variant, consistent interactions on T2D risk are also apparent.

### 4.4 Paper IV

In this study, we attempted to replicate previously reported main genetic effects and gene x physical activity interactions for the *FTO* rs9939609 variant on obesity (261). The cohorts used in these analyses were two primarily non-diabetic population-based cohorts of middle-aged Swedish (MPP) and Finnish (PPP-Botnia Study) adults. Participant characteristics are shown in Table 10.

**Table 10** Participant characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>MPP</th>
<th>PPP-Botnia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline (n=15,931)</td>
<td>Follow-up (n=15,844)</td>
</tr>
<tr>
<td>Sex (M/F)</td>
<td>10,335/5,965</td>
<td>10,278/5,566</td>
</tr>
<tr>
<td>Age (years)</td>
<td>45.5 (6.9)</td>
<td>68.5 (5.6)</td>
</tr>
<tr>
<td>BMI (kg/m2)</td>
<td>24.3 (3.3)</td>
<td>27.1 (4.1)</td>
</tr>
<tr>
<td>Physically active</td>
<td>12,480/3,451</td>
<td>12,414/3,430</td>
</tr>
<tr>
<td>rs9939609 (TT/TA/AA)</td>
<td>5,557/7,656/2,718</td>
<td>5,528/7,813/2,703</td>
</tr>
</tbody>
</table>

Values are n or means (SD). Physical activity in the MPP is coded as ‘sedentary’ or ‘physically active’. Physical activity in the PPP-Botnia Study is coded as a ‘low’ (<10 MET-h/week), ‘moderate’ (10–40 MET-h/week) or ‘high’ (>40 MET-h/week) level of physical activity. M, male; F, female.

As in previous reports, the minor A allele at rs9939609 is associated with a 0.13 kg/m² and 0.43 kg/m² unit increase in BMI per copy of the minor allele in the MPP and PPP Botnia cohort, respectively (328). No interaction effect was evident between the rs9939609 variant and physical activity on BMI in either study cohort (see Figures 11a and 11b).
Figures 11a (left panel) 11b (right panel) Absence of a detectable interaction between the FTO rs9939609 SNP and physical activity on BMI in men and women. Panel A shows data from 15,925 non-diabetic Swedish men and women ($P_{\text{interaction}}=0.71$) and Panel B shows data from 2,511 primarily non-diabetic Finnish men and women ($P_{\text{interaction}}=0.18$). Data are means adjusted for age and sex (95% CIs). Genotype groups: black bars, T/T; grey bars T/A; white bars A/A

In models where obesity was defined using the WHO categorical definitions for obesity (52) there was no evidence of a statistically significant interaction in either cohort, nor was there evidence of an effect of interaction on change in BMI from baseline to follow-up in the MPP with ($P=0.27$) or without ($P=0.49$) the inclusion of incident cases of diabetes. The exclusion of individuals with T2D from the Finnish cohort did not materially affect the results.

Because the MPP is considerably leaner on average than the Danish cohort featured in the original report (261) we also restricted our analyses to the 5,705 individuals with a BMI $>25.0$ kg/m$^2$. The results from these analyses did not differ materially from the analyses in the entire cohort ($P_{\text{interaction}}=0.64$ for baseline models and $P_{\text{interaction}}=0.54$ for prospective models).

In summary, the risk of obesity associated with the FTO variant (rs9939609) was replicated in largely non-diabetic population-based cohort of 18,442 middle-aged Swedish and Finnish adults, but we were unable to replicate the previously reported interaction between this FTO variant and physical activity.
5 DISCUSSION

5.1 Summary of principal findings

Many believe that T2D is caused by the interaction between genetic and environmental factors. However, few adequately replicated examples of gene x lifestyle interactions on T2D risk have been reported, suggesting that if such interactions exist, they may be small in magnitude or considerably more complex in nature (i.e. involving more than two dimensions) than initially thought.

5.1.1 Paper I

PPARGC1A is one of the strongest biological candidate genes for T2D, particularly with respect to interactions with physical activity; PPARGC1A encodes a protein which coactivates at least 30 transcription factors involved in the regulation of complex metabolic cascades. Largely because of this, DNA variation at PPARGC1A has been extensively studied in relation to cardiometabolic traits, with almost all studies focusing on a single non-synonymous variant (Gly482Ser). The study reported in Paper I of this thesis is the first to comprehensively assess the role of common PPARGC1A sequence variation with respect to cardiometabolic quantitative traits. The strongest associations were for the rs7656250 and rs13117172 variants (with aerobic fitness) and the rs7657071 variant (with fasting glucose concentrations), which in this cohort are in very low linkage disequilibrium (LD) (r²=0.13) (Figure 12).

Figure 12 Pairwise linkage disequilibrium (r²) between PPARGC1A markers in the European Youth Heart Study (n=2101)

Since the initial report of association between the Gly482Ser variant and T2D (329), many others have reported statistical associations not only with T2D
(222), but also with other traits that include obesity (330, 331), dyslipidaemia (332-334), aerobic fitness (335-337), insulin resistance (338), and hypertension in adults (339, 340) and adolescents (341). A recent meta-analysis supports a modest role for the Gly482Ser variant at PPARGC1A with T2D (222). However, conflicting results have also been obtained. For example, the association of this polymorphism with T2D was not detectable in several populations such as French, Pima Indians or Austrians (332, 342, 343). Moreover, associations were undetectable between Gly482Ser and diabetes-related traits, such as glucose tolerance, insulin sensitivity, glucose disposal, insulin secretion, maximal oxygen consumption, or intra-myocellular lipids content, in non-diabetic German and Dutch populations (344).

Here we observed a nominal statistical association between Gly482Ser and waist circumference, but no associations were observed for other traits. We also detected nominal statistical associations for several other PPARGC1A SNPs with other candidate traits (i.e., BMI, waist circumference, diastolic and systolic blood pressures, aerobic fitness, and fasting glucose concentrations). However, after correcting for multiple statistical comparisons none remained significant.

Applying rigorous corrections for multiple statistical comparisons, such as we have done here, may be overly conservative when testing genes with strong a priori biological evidence of association with the traits on interested. Thus, one should bear in mind that some of the corrected results may be prone to type 2 error. For example, the association between fasting glucose and rs7657071 remained of borderline statistical significance (P<0.07) even after a conservative correction was applied, suggesting that this variant may warrant further investigation in other cohorts.

In conclusion, our findings suggest that common variation at the PPARGC1A gene is unlikely to be a major factor in the regulation of metabolic or cardiovascular traits in this European paediatric cohort.

5.1.2 Paper II

In Paper II, we found a significant interaction between the TCF7L2 rs7903146 variant and physical activity on HOMA-B in a well characterized cohort of 2,101 healthy children from the EYHS, with objective measures of physical activity. Physical activity attenuated the effect of the rs7903146 risk allele on HOMA-B across the fourth and fifth quintiles of physical activity.

We also found evidence of association between the risk allele of rs7903146 with hyperglycaemia and height. The risk allele of rs7903146 was associated with higher glucose levels in older but not younger children. Boys carrying the minor allele at the TCF7L2 rs7903146 were shorter in stature during childhood but no such effect was observed in girls. No consistent evidence of associations with HOMA-B or HOMA-S emerged. Similarly, we found no evidence of association with measures of obesity.

In conclusion, our findings suggest that physical activity may modify the effect of the TFC7L2 rs7903146 variant on beta-cell function in a cohort of healthy European children. This observation may have implications for public health because the processes that underpin the effects of TCF7L2 gene variant on adult diabetes risk may begin early in life, suggesting that the initiation of lifestyle interventions aimed at offsetting these effects may need to be timed accordingly.
5.1.3 Paper III

In Paper III, we showed that physical activity modifies the effect of common \textit{HNF1B} and \textit{CDKN2A/B} variants on risk of IGR and also modifies the effect of the \textit{HNF1B} on T2D risk. We also showed that 10 of the 17 polymorphisms examined showed evidence for association with IGR risk and/or 2-h glucose concentrations in directions consistent with previous reports (\textit{SLC30A8} rs13266634, \textit{TCF7L2} rs7903146, \textit{CDKAL1} rs7754840, \textit{NOTCH2} rs10923931, \textit{KCNJ11} rs5912, \textit{IGFBP2} rs4402960, \textit{JAZF1} rs864745, \textit{HHEX} rs1111875, \textit{MTNR1B} rs10830963 and \textit{TSPAN8} rs7961581) (25, 150, 153, 173, 174, 203, 205). The risk of glucose intolerance and T2D conveyed by \textit{HNF1B} (rs4430796) appears to be offset in people reporting physically active lifestyles. These putative interactions were observed in \~16,000 people, of whom more than 2,000 developed T2D during a median follow-up period of 24.5 yrs, making this the largest prospective study of gene x lifestyle interactions on T2D risk reported to date. Consistent interaction effects emerged in the sub-group of nearly 9,000 individuals in whom 2-h glucose concentrations had been recorded during an OGTT. These tests remained statistically significant after correction for multiple statistical comparisons. The interaction effect for a second variant at the \textit{CDKN2A/B} (rs10811661) was nominally statistically significant when baseline IGT or 2-h glucose were the outcomes. The nature of these effects is consistent with those reported in a recent clinical trial of intensive lifestyle intervention (289). However, when T2D was the outcome, the gene x physical activity interaction term was not statistically significant in this cohort. A borderline uncorrected statistical interaction on IGR risk for a third variant at the \textit{PPARG} gene (rs1801282) was also observed, but this was not apparent within the subcohort in which 2-h glucose concentrations were available, nor was an interaction evident when T2D was the outcome.

In conclusion, we found that physical activity modifies the effects of \textit{CDKN2A/B} on IGR risk. To our knowledge this is the first study to report interactions between variants at the \textit{HNF1B} locus and physical activity on IGR or T2D risk. These findings suggest that the genetic predisposition to T2D is partially dependent on a person’s lifestyle.

5.1.4 Paper IV

In this study, we were unable to replicate previous reports of interaction effects between the \textit{FTO} variant (rs9939609) and physical activity on levels of BMI in a large population-based study of 18,442 primarily non-diabetic Swedish and Finnish adults. We were, however, able to replicate the association between the \textit{FTO} variant and BMI (219).

5.2 Strengths and limitations of the studies

5.2.1 Paper I

There are several possible explanations for the apparent inconsistency between our findings for \textit{PPARG1A} Gly482Ser and those previously reported. For example, most existing studies were conducted in adults of advanced age, from a variety of ethnic backgrounds. By contrast, the cohort studied here consisted of relatively healthy children of Northern European ancestry. It may also be that
our findings are false negative owing to suboptimal statistical power for some of the tests performed, or that the previous reports contain false positive results owing to uncorrected multiple statistical comparisons which might have been performed in some studies.

**PPARGC1A** is a relatively large gene, comprised of eight exons. In European whites, >30 tagging variants are required to capture the common variation across the gene. Thus, a caveat to undertaking a comprehensive study of **PPARGC1A** is that it involves multiple statistical comparisons, which decreases statistical power. Most previous studies report nominally statistically significant findings that are rarely adjusted for multiple statistical comparisons, predisposing them to type 1 error. It is also possible that a degree of publication bias exists, where association studies with statistically significant findings are disproportionally represented in the literature. These points, combined with the fact that our study was underpowered for some of the association tests performed, should be taken into consideration when comparing the findings reported here with those reported elsewhere.

5.2.2 Paper II

To our knowledge this is the first study to report formal statistical evidence of a gene x physical activity interaction at the **TCF7L2** locus. Physical activity attenuated the effect of risk (T) allele of rs7903146 HOMA-B. This finding suggests that physical activity modifies the effect of the rs7903146 variant on insulin secretion.

Although our study is the largest of its kind to date and physical activity was objectively measured, this study has several potential limitations. Glucose and insulin were measured only in the fasting state; a noteworthy study with euglycaemic-clamp-derived insulin sensitivity in children found no significant associations between rs7903146 and insulin sensitivity (345). An additional limitation is that as previously described, our study is underpowered for the detection of gene x lifestyle interaction effects (346). Thus, the absence of such interactions should be interpreted with caution.

5.2.3 Paper III

This is the first study prospective cohort study to our knowledge to assess gene x physical activity interactions across a wide-range of previously associated T2D gene variants. It is also the largest study of gene x lifestyle interactions on pre-diabetic and diabetic traits to date, and hence one of the most well powered to test related hypotheses. Nevertheless, because realistic interaction effect sizes are difficult to determine (owing to the few *bona fide* interactions effects which have been reported it is difficult to estimate the magnitude of a potential interaction effect) it is difficult to interpret power calculations when it comes to gene x physical activity interactions. Here we used what we believe to be a conservative estimate of the interaction effect (i.e. from a risk ratio of 1.25 to 1.50 – see Paper III). Our study was well powered to detect interaction effects of 1.50 per allele or greater when IGR and T2D are the outcomes, but less so if the true interaction effect is closer to 1.25 per copy of the risk allele (see Paper III).

A limitation of this report is that physical activity was assessed using a basic questionnaire which, as with all questionnaires, is prone to error and bias. Measurement error diminishes power to detect interactions owing to
misclassification of the environmental exposure. However, our measure of physical activity was inversely correlated with IGR and T2D risk in this cohort, which is consistent with the direction of association reported in clinical trials involving exercise as part of a program of intensive lifestyle intervention (46, 81) or epidemiological studies where physical activity was objectively assessed (347), instilling confidence that the questionnaire correctly classified the behaviour of physical activity. Moreover, we purposefully studied people free from diabetes at baseline to minimize the extent to which changes in behaviour and cognition attributable to disease labelling and treatment might have biased the reporting of physical activity levels by study participants.

5.2.4 Paper IV

This study is about fourfold larger than the original report, meaning that with all else being equal it has sufficient power to detect the interaction effect reported previously (261). It is thus interesting that no interaction effect was observed between the FTO variant and physical activity on BMI. It is possible that our inability to replicate the initial study’s findings is attributable to differences in the physical activity measures across studies or that the cohorts differ in other respects. The Danish and MPP cohorts are geographically proximal, comprising largely white, middle-aged northern Europeans, and similar physical activity metrics were used; although it should be highlighted that the estimate of physical activity used in our study was limited to two levels (active and inactive), whereas in the Danish study three activity levels were defined. The most substantial differences between cohorts included those related to the level of obesity. For example, the MPP participants were leaner on average than those in the Danish study and the effect of the FTO variant in the MPP is roughly half the effect reported elsewhere (348). Because physically active individuals tend to be leaner than their inactive counterparts, this may indicate that overall MPP participants had relatively higher habitual physical activity levels than in the Danish cohort, was may have limited our ability to detect an interaction. However, this is unlikely to explain the lack of a significant interaction effect in the PPP-Botnia Study, as this cohort is of comparable mean BMI to the Danish cohort. Moreover, restricting the MPP sample to people of similar BMI to the Danish cohort had no bearing on the interaction between rs9939609 and physical activity. Since our papers were published, several other reports have emerged studying closely related research questions (263-265), with inconsistent results. Thus, the next step is to undertake an adequately powered meta-analysis of published and unpublished materials.

5.3 Biological interpretation of the studies’ findings

The main purpose in Paper I was to investigate PPARGC1A gene x physical activity interactions for both cardiovascular and metabolic disease traits in a well characterized paediatric cohort that had undergone objective measures of physical activity. We chose the PPARGC1A gene because the existing biologic evidence shows that PPARGC1A is responsive to changes in exercise and mediates several key metabolic pathways relevant to obesity and T2D (349). Several studies suggest that in patients with T2D the expression of PPARGC1A and the sub-set of genes involved in oxidative phosphorylation that PPARGC1A
coactivates (known as OXPHOS genes) is reduced in skeletal muscle compared with normal glucose tolerant controls (224, 225), although this may be due to lower physical activity levels in the diabetic individuals (350). Despite its plausibility, the notion that PPARGC1A interacts with physical activity to modify the risk of T2D has limited empirical support in human populations.

The PPARGC1A gene regulates key functions in the dynamic transcriptional control of metabolic pathways. The protein encoded by the PPARGC1A gene co-activates multiple transcription factors involved in thermogenesis, mitochondrial biogenesis, glucose/fatty acid metabolism, adipogenesis, regulation of reactive oxygen species and hepatic glucose production (303, 351-353). PPARGC1A is expressed predominantly in tissues with high metabolic activity, such as brown adipose tissue, heart, kidney and exercising skeletal muscle (223). PPARGC1A deficient mice show profound derangements which include reduced exercise capacity, cardiac dysfunction, hepatic steatosis, insulin resistance and glucose intolerance (354). In humans, sequence variation at PPARGC1A is associated with reduced PPARGC1A mRNA levels (335).

In skeletal muscle PPARGC1A may be up-regulated by acute and chronic exercise through multiple pathways such as CaMK, calcium, myocyte enhancer factor 2 (MEF2), p38 mitogen-activated protein kinase and AMP-activated protein kinase (355-362). Interestingly, the mechanisms that underlie PPARGC1A expression appear to be down-regulated in skeletal muscle by aging (335), inactivity (350, 363), accumulation of fatty acids (Benton, 2006 #521) (364-366), insulin resistance, and T2D (224, 225, 335). In particular, PPARGC1A induction following exercise up-regulates the expression of nuclear respiratory factors (NRF-1 and NRF-2) and mitochondrial transcription factor A (TFAM) to promote mitochondrial replication and to increase the expression of several other genes involved in oxidative phosphorylation (225); Baar, 2004 #526). PPARGC1A may also regulate muscle fibre type determination; in PPARGC1A transgenic mice, the type II muscle fibres express proteins characteristic of type I fibres, such as troponin I (slow) and myoglobin, and show a much greater resistance to electrically stimulated fatigue (355). In rodents (359, 367, 368) and humans (369, 370), induction of PPARGC1A signalling through short-term exercise and endurance training can enhance mitochondrial oxidative enzymes and mitochondria count in skeletal muscle, improving oxidative capacity. Exercise induce cellular glucose uptake through largely insulin-independent mechanisms, which is why insulin-deficient humans with T2D can improve their glucose homeostasis by being physically active. Exercise stimulates GLUT4 expression in skeletal muscle, which is a key glucose transported, in part through the activation of PPARGC1A (371, 372). Other genes thought to be involved in this process include mitofusin 2 (Mfn), which is induced directly by PPARGC1A and the orphan nuclear receptor estrogen-related receptor (ERR) (373). ERRα (Estrogen-related receptor α) interacts directly with PPARGC1A to induce the expression of genes involved in glucose and lipid metabolism. The process of PPARGC1A-mediated glucose disposal with exercise also involves the products of glucose-6-phosphatase (G6Pase) and hepatocyte nuclear factor-4α (HNF-4α), which lie downstream of the forkhead box O1A gene (FOXO1) (374). Therefore, reducing the hepatic insulin requirements in insulin-resistant patients through exercise may indirectly influence hepatic PPARGC1A activity.
In paper II, we selected a now widely studied variant (rs7903146; IVS3 +46983 C>T) at the TCF7L2 gene. This variant conveys the largest individual risk effect of all known common gene variants discovered to date. TCF7L2 belongs to a family of transcription factors that contain high mobility group (HMG) box DNA-binding domains that affect cell proliferation and differentiation via the Wnt signalling pathway. Although TCF7L2 is the gene which has been most extensively replicated with respect to its association with T2D, the underlying mechanisms are not fully understood. Tcf7l2-null mice lack gut epithelial stem cells leading to marked deficiencies in intestinal cells including enteroendocrine cells, which are responsible for incretin production (375). Incretins, including glucagon-like peptide 1 (GLP-1), are hormones secreted by the gut in response to the presence of nutrients and have been shown to regulate gut motility, satiety and energy homeostasis (376). In addition, GLP-1 plays a crucial role in stimulating insulin secretion by the pancreatic beta cell; it turns out that TCF4 (the protein product of TCF7L2) trans-activates the gene that encodes GLP-1 (377), thus defining a link between TCF7L2 and GLP-1 in the beta cell. Concordantly, carriers of the risk T allele at rs7903146 exhibit lower insulin secretion after an oral glucose load than non-carriers (284, 378-381). Intronic sequences may contain elements that enhance gene expression. Such a relationship could explain the association of TCF7L2 variants with incretins within pancreatic islets, since there is a binding site for TCF7L2 in the promoter of the preproglucagon gene (381). Although, TCF7L2 genotypes do not seem to differ regarding the concentrations of glucose-dependent insulinotropic polypeptide (GIP), GLP-1, and glucagon (382), higher levels of GIP and glucagon have been observed in carriers of the risk T allele (381). These findings suggest that TCF7L2 may increase risk of diabetes by influencing insulin secretion and incretin action, and altered expression of the gene in islets (1). In addition, TCF7L2 may also cause a defect in insulin processing (383). Expression of TCF7L2 is correlated with expression of the gene for insulin (INS), where human islets isolated from T allele carriers show increased TCF7L2 expression (383). Knocking down the expression of TCF7L2 result in lower insulin gene expression and lower glucose-mediated insulin secretion in mouse insulinoma cells (384), and impaired glucose-stimulated insulin secretion in human beta cells (385). Despite higher measurable proinsulin in serum and increased gene expression (383, 386), T allele carriers, both in vivo (284, 378-380) and in vitro (381), have lower insulin secretion. These findings suggest that TCF7L2 may insert a block between INS gene expression and post-translational processing or secretion of insulin. However, this increase in proinsulin levels may also be a marker of beta cell dysfunction.

In paper III, we studied all of the T2D gene variants that have been reproducibly shown to be associated with T2D risk. The most promising result from this study emerged for HNF1B. This gene (also known as TCF2) encodes a homeobox transcription factor and is a member of the homeodomain-containing superfamily of transcription factors. Mutations in this gene are known to cause various phenotypes including rare monogenic forms of Maturity Onset Diabetes of the Young (MODY5) and abnormalities in kidney, pancreas, liver, and genital tract formation (387-392). More recently, polymorphisms in HNF1B were reported to be associated with T2D in Caucasians (163, 165, 393). Defects in the metabolism of glucose in pancreatic beta cells may be the mechanism by which insulin secretion is impaired in T2D. Of particular interest is the SNP rs4430796 located in the second intron of the HNF1B gene. This
variant simultaneously protects against prostate cancer and increases risk of T2D, a link that requires further investigation (165).

The variant that shows the strongest association with T2D elsewhere and was studied here (rs4430796) is located in a DNA sequence with no obvious functional effect. This SNP may well be in high LD with causal SNPs in the coding regions. Interestingly, another SNP (rs7501939) was reported to be associated with T2D (165) which also turned out to be located in the noncoding region of the HNF1B (first intron). In fact, no common coding SNPs within this gene were revealed by the HapMap Phase II. Noncoding variants may also affect gene expression patterns, which provides an alternative explanation for the association between the rs4430796 variant and T2D.

A variant proximal to the CDKN2A and CDKN2B genes also appeared to interact with physical activity in our study. These genes localize to neighbouring genomic regions and function as tumour suppressors via inhibition of cyclin-dependent kinases and are best known for their role in the development of cancers and key physiological processes such as replicative senescence, apoptosis and stem-cell self-renewal (200, 394). As with the HNF1B variant, it is likely that high LD with a coding variant may explain the association with T2D, as the variant lies about 200 kb from the nearest known coding regions (157).

CDKN2A/B is highly expressed in adipocytes, pancreatic islets, and vascular endothelial cells (150, 152, 395). In transgenic mice, overexpression of cdkn2a with ageing contributes to decreased islet proliferation suggesting that T2D, in part, results from cdkn2a-induced replicative failure of islets (396). A possible explanation for the association with T2D could be gene x gene interaction, owing that regulatory effects on CDKN2A expression could be mediated through a non-coding RNA gene, the ANRIL, which maps to the region of maximal association with T2D (157, 397). In future, it would be interesting to examine if the known T2D-susceptibility loci that influence cell-cycle regulation are also associated with cancer predisposition.

The third variant that showed tentative evidence of interaction in our study localises to PPARG. This gene is a member of the nuclear hormone receptor family and is a transcription factor that significantly influences insulin sensitivity (398, 399). PPARG has been identified as a functional receptor for the thiazolidinedione (TZD) class of insulin sensitizing drugs (400). PPARG represents a highly replicated biologic candidate gene for T2D that it is expressed in human beta-cells (401) and in adipose tissue (402, 403), where it modulates the expression of target genes implicated in adipocyte differentiation and glucose homeostasis. The variant that has conveyed the strongest associations at this locus is the missense rs1801282 SNP (encoding a proline to alanine amino acid substitution at codon 12: Pro12Ala) in exon B of the gene (404). Unlike most other associated variants, Pro12Ala is likely to functionally impact gene expression (178).

In epidemiological studies, the Ala12 allele has been associated with increased insulin sensitivity and a moderately decreased risk of T2D in various populations (405). The largest meta-analysis showed a 19% reduction in the risk of T2D for the Ala12 carriers, but it also indicated large ethnic heterogeneity between populations, with frequencies range from 2 to 23% (182). In our study (Paper III) no evidence for association with IGR risk or 2-h glucose concentrations was observed. We did however observe a nominal statistical gene x physical activity interaction on IGR risk, which may be of interest given that
many other studies have reported evidence of gene x lifestyle interactions at this locus. Nevertheless, despite epidemiological evidence indicating that PPARG may interact with physical activity to modify the risk of T2D, the mechanisms that underlie these relationships are unclear.

Whilst one cannot exclude a possible functional interaction of physical activity and Pro12Ala, an alternative explanation for our findings could be that dietary fatty acids (12, 178), which are correlated with physical activity levels, may underlie the observed interaction. This would fit well with previous studies on gene x lifestyle interactions at this locus, most of which have focused on dietary fatty acids (12). Unfortunately, we did not have measures of dietary fat intake available in this cohort and could not test this possibility.

The fourth and final paper in this thesis examined the interaction between the rs9939609 SNP at the FTO locus and physical activity on levels of BMI. This interaction had been reported in three other studies when we embarked on these analyses (refs). The exact function of FTO is unknown. FTO encodes a 2-oxoglutarate-dependent nucleic acid demethylase, which is highly expressed in hypothalamic nuclei and in muscle (219, 220). FTO is also expressed in adipose tissue, but FTO mRNA is not correlated with rs9939609 genotypes (406, 407). Epidemiological studies have indicated that this gene is robustly associated with obesity-related traits in humans (219-221, 408, 409). We were able to replicate this association; the minor allele at rs9939609 was associated with increased levels of BMI.

We have extended our study to FTO rs9939609 x physical activity interactions since this gene is expressed in muscle (219), it is possible that the FTO rs9939609 genotype might affect insulin-mediated glucose uptake in muscle.

5.4 Strengths and weaknesses in relation to other studies

Genetic association studies have been proposed as a powerful tool to identify common genetic variants involved in complex polygenic diseases, such as T2D. However, the role of genetic association studies in failure to replicate findings remains controversial.

Lack of reproducibility of genetic associations has been commonly attributed to insufficient statistical power, population stratification or various forms of between-study heterogeneity or environmental influences (410, 411). Physical activity and diet have been reported to be important modifiers of the association of genes with T2D and might explain partially why some genetic association results are not always replicated (12). In this thesis we have extended our study to gene x physical activity interactions. Furthermore, due to the high number of reported SNPs, genetic association studies requires successful replication with well established genotype–phenotype association to provide credibility that the initial finding is valid (411). In this thesis SNPs were selected on the basis of a mixed approach based on candidate gene approach (Paper I, II) and GWAS (Paper III, IV) for T2D related traits. The change in methodological emphasis in this thesis from the biologic candidate gene approach to GWAS reflects the change in emphasis in the field of human genetics in general during the past three or four years.

Recently, critical focus has been directed towards biologic candidate gene approach in complex disease due to conflicting and unreplicable findings. For example, in Paper I PPARGC1A was selected because this gene is considered a
strong biological plausible candidate gene for a variety of metabolic and cardiovascular disease endpoints (303). However, conflicting results have been reported (332, 342-344). In our study (Paper I), a thorough exploration of common PPARGC1A variants was performed by selecting 35 tagging SNPs in order to capture all common variation across the gene in a well characterized cohort of Scandinavian whites children. However, we were unable to replicate associations with cardiovascular or metabolic traits. Four criteria for candidate gene approach in complex disease are required as follows: location of the gene in a chromosomal area of linkage; change in protein level or function by the mutation; identification of gene based on biological plausibility that might have a role in the aetiology of the disease; and consistent results (412, 413). However, these criteria appear to be rarely fulfilled. For example, whether the known Gly482Ser PPARGC1A is a functional variant remains to be shown (330). Candidate gene approach is questionable by its reliance on the priori knowledge about the physiological, biochemical or functional aspects of possible candidates (414). In addition, the variety of phenotypes and ethnic differences between the populations makes the interpretation difficult. Another difficulty is that a single gene makes only a small contribution to the susceptibility of multifactorial disease. Publication bias in the literature could be another problem: there is evidence that it is easier to publish a positive study than a negative finding and hence the true significance of findings reported is unclear. In a recent meta-analysis of published studies, a statistical association between the PPARGC1A Gly482Ser polymorphism and T2D was reported (222). However, this finding was not supported by a meta-analysis of T2D genetics conducted in a somewhat larger total sample (25). Similarly, in a recent meta-analysis of Gly482Ser and blood pressure, statistical associations were evident when including only previously published data, but these effects were abolished when unpublished data were also included (339). Lastly, the majority of the published studies have been based on small samples.

In the last three years the availability of data from GWAS has contributed dramatically to a growing number of common genetic variants increasing susceptibility to T2D providing strong evidence for the potential power of GWASs (148-150, 152, 153). In Paper III and IV the variants were selected because they were previously associated with T2D and obesity in GWAS, respectively. In our findings (Paper III, IV) the majority of the main effects confirm the previous reports from GWAS. However, absence of gene x physical activity interactions was observed for the majority of these interactions effects. These findings are not unsurprising, given that GWAS methods are designed to detect only the main effects and ignore heterogeneity between exposure classifications (150, 152). Our studies (Paper III, IV) are likely to be adequately powered to detect realistic interaction effects but GWAS are insensitive to detecting variants with heterogeneous effects, owing to the strict significance thresholds used to determine genome-wide statistical significance. Thus, the most strongly associated variants in GWAS are likely to be those with the least heterogeneous effects, which by definition are likely to exclude many variants that exert their genetic effects through interactions with environmental (or other genetic) exposures. In summary, our strategy to investigate gene x physical activity interaction was based on association studies in unrelated individuals. Future investigations may adopt alternative strategies to select genes for lifestyles interactions terms.
In this thesis, cross-sectional (Papers I, II), and prospective population-based cohorts (Papers III, IV) designs were used, to quantify the main genetic effects and gene x lifestyle interactions on T2D and related traits in paediatric and adult Scandinavian cohorts, respectively. Paediatric cohorts represent an interesting population for identifying genetic susceptibility for complex traits that are intermediate to diseases such as T2D. This is because those traits are less influenced by adverse environmental exposures in childhood than in adulthood, and may help to understand the sequence of events leading to diabetes across life. The nature of these cohorts is observational (papers I-IV) and causal inferences are difficult to make. However, prospective cohort studies (besides experimental studies) generally provide the best epidemiological evidence about causation and the most direct measurement of the risk of developing disease because subjects are followed up in the same way. The use of a longitudinal study design is that minimizes or eliminates selection, survivor, and recall biases, and the prospective collection of biomarker data is feasible. The major disadvantage of this approach is the large number of subjects required at baseline in order to guarantee a sufficient number of cases for reliable analysis in order to be informative as incident cases. In addition, it may be difficult to detect interactions related to rare phenotypes in traditional cohort studies due to large data sets and great expenses. In papers III and IV, we have used a hybrid design by combining large prospective cohorts with the efficiency of a case-control study which provides the lower likelihood of bias. The main problem in cross-sectional cohort studies is the difficulty to separate cause from effect, because measurement of exposure and disease is conducted at the same time, limiting their predictive value. This means that it is not possible to say anything regarding causality. We have decided to use a cross-sectional design using all cohort (Paper I and II) instead a case-control design in order not to lose power.

A dilemma on the study design for investigating gene x environment interactions, as with all existing epidemiological investigations, are that measurements of lifestyle exposures are often imprecise and the studies are often small. In this thesis, papers I and II include small sample sizes, but physical activity was measured using objective methods (uniaxial accelerometer). By contrast, papers III and IV include bigger samples and subjective measures of physical activity (questionnaires).

Small sample sizes result in reduced power to detect minor contribution of alleles. In the case of gene x environment interactions analysis this is particularly important because they require larger samples compare with association studies. The sample size of our cohort from the EYHS (papers I, II) was not very large, which might have increased type 2 error rates. However, small studies are likely to overestimate the true effect size. Owing to the ethnic diversity of the EYHS cohort, it is possible that the allelic substructure of the two cohorts varies which could lead to population stratification and confounding genotype-phenotype associations when pooled analyses are undertaken (papers I, II). To address this issue, initially we have analysed the LD structure of each population separately confirming that the LD structure was similar. Subsequently, combination of LD were undertaken in order to augment the sample size for increasing statistical power. In addition, all the analyses were adjusted for country. We were also conscious that the magnitude of the genotype associations may differ by age or sex. Hence we tested these hypotheses using the relevant interaction terms.
The weak measure of physical activity in papers III and IV is a limitation of the study. However, our measure of physical activity was inversely correlated with IGR, T2D risk (Paper III), and BMI (Paper IV) which is consistent with the direction of association reported in clinical trials involving exercise as part of a program of intensive lifestyle intervention (46, 81) or epidemiological studies (3) where physical activity was objectively assessed (347), instilling confidence that the questionnaire correctly classified the behaviour of physical activity. Nevertheless, it would be important to validate these findings with more direct measures of physical activity in the future.

Beyond sample size and measures of environmental exposure such as physical activity, the power to detect the main effects and gene x environment interactions is also influenced by the accuracy of the measures of the outcome (291). In papers I and II, the analyses were performed on a well-characterized cohort of healthy children, with metabolic traits obtained from fasting measurements. In papers III and IV, the analyses were performed on well-characterized cohorts of initially non diabetic adults, with metabolic traits obtained from fasting and/or OGTT measurements. In addition, diagnosis of diabetes was confirmed from Hospital patient records. The use of such measures could provide information on possible causal relationships between lifestyle and regulation of glucose homeostasis (418). One important point to keep in mind is that, throughout these papers we have had an extensive quality control to assure reliable genotyping results. The genotyping success rates exceeded 95% (papers I-IV).

Most previous genetic association studies report nominally statistically significant findings that are rarely adjusted for multiple statistical comparisons, predisposing them to type I1 error. In this thesis we have included corrections for multiple testing in papers I and III by using the Holm procedure (419) because this method may reduce false positive findings. It should be noted though that in the analyses involving a high number of multiple hypotheses testing (paper I include 35 tag SNPs) strictly corrected significance levels could also lead to false negative associations suggesting that observed trends for association should not be completely disregarded. In this thesis the most robust interaction terms were for the HNF1B SNP (paper III). In papers II and IV we have chosen not to include corrections for multiple testing because these genes (TCF7L2 and FTO) are among the most extensively replicated and are tests are largely confirmatory.

There are some potential confounders that we have considered in this thesis such as BMI, age, sex. Adjustment for these potential confounders did not change our estimates in general. However, we have not considerer confounding from any dietary factors in this thesis owing to a lack of such data. The absence of a statistical interaction between PPARG gene (rs1801282) and physical activity in the cohort with 2-h glucose concentrations may be a consequence of lower statistical power owing to the smaller sample size, but inadequate power is unlikely to explain the lack of effect where diabetes was the outcome. For this locus, numerous prior reports of gene x lifestyle interaction exist, although most have focused on gene-nutrient interaction effects (12). Any epidemiological study is not in and of itself sufficient to establish causality; for this a well-designed experimental study such as a randomized clinical trial is required. Even in well designed epidemiological studies, it is impossible to account for all sources of bias, misclassification, and confounding.
We tested for gene x physical activity interactions but few statistically robust findings emerged. This may reflect the complexities associated with testing interactions or the lack of statistically detectable interactions for the selected gene variants. Another issue that should be raised is that we only considered associations of individual SNPs. However, susceptibility for T2D may be modulated by epistasis (gene x gene interactions) (420). The failure to take into account such interactions in our studies may have led to false negative results.

5.5 Comparison of findings with previously reported studies

In Paper I, we did not identify any statistically significant gene x physical activity interactions at the PPARGC1A locus. This is despite strong biologic plausibility for such interactions. We were also unable to confirm numerous prior reports of genetic association. In humans, the Gly482Ser variant at PPARGC1A has been associated with physical activity (331) and reported to interact with physical activity to modify levels of aerobic fitness (335-337). It is possible that this variant may also modify an individual's cardiorespiratory response to exercise training, which has relevance for the prevention and treatment of T2D. This is because low cardiorespiratory fitness is associated with impaired fasting glucose, dyslipidaemia, insulin resistance, hypertension and T2D (421). The Gly482Ser variant has also been associated with blunted NEFA clearance following oral glucose challenge, which is most evident within obese individuals (333). The latter may be related with obesogenic behaviours, such as low physical activity levels and high saturated fatty acid diets, both of which inhibit PPARGC1A expression (422). Numerous other reports of associations between Gly482Ser and cardiometabolic traits also exist, but we found no evidence to indicate such effects are detectible in the paediatric cohort studied here.

As mentioned above, the main difference between the studies reported in Paper I and Paper II and previous studies is that our studies were conducted in healthy children. Paper I also focused on characterizing all common variation at the PPARGC1A gene, which required multiple statistical comparisons. By contrast, previous studies were conducted in adults of advanced age and focus only on a few PPARGC1A SNPs. Because the biological plausibility for association between PPARGC1A variants and cardiometabolic traits is high, one could argue that the borderline statistically significant associations we report in this thesis should not be discarded. Nevertheless, all such associations require replication in appropriate cohorts to confirm or refute our findings. The association between fasting glucose and the rs7657071 SNP is one example of a finding that warrants follow-up. Data from the Diabetes Prevention Program and the Finnish Diabetes Prevention Study (284, 423) suggest that the risk conveyed by the TCF7L2 variants can be reduced following intensive lifestyle modification in high-risk adults. However, neither study formally tested for gene x treatment interactions; indeed the P-value for gene x lifestyle interaction on T2D incidence in the Diabetes Prevention Program was not statistically significant (P=0.15) (JC Florez, personal communication). In the present study (paper II) physical activity attenuated the effect of the risk (T) allele of rs7903146 on estimated beta cell function. To our knowledge this is the first study to report formal statistical evidence of gene x physical activity interaction at the TCF7L2 locus. This finding suggests that physical activity modifies the
effect of the rs7903146 variant on insulin secretion, but no specific mechanism for this effect is known.

Paper III reported a novel interaction effect for an HNF1B variant (rs4430796). In the initial reports of association between the HNF1B rs4430796 polymorphism and T2D (165), the major G allele was associated with decreased T2D risk, as observed in physically active individuals in the current study. However, any beneficial effect of this allele is essentially lost, and possibly reversed, in the physically inactive individuals studied here. In our study the initially strong inverse association between physical activity and diabetes incidence diminishes in magnitude with each additional copy of the minor HNF1B allele. This result suggests that the protective effects of physical activity on diabetes risk may be attenuated in carriers of the rs4430796 minor allele. Interestingly, the same variant was associated in the opposite direction with prostate cancer risk in the initial report (165). Thus, it would be important to determine whether gene x physical activity interactions are also relevant for that disease, and if so how lifestyle intervention influences prostate cancer risk in carriers of the different rs4430796 genotypes. No prior studies to our knowledge have reported evidence of gene x physical activity interactions at the HNF1B rs4430796 locus.

In paper III we also confirmed a previous report of gene x lifestyle interaction for a CDKN2A/B variant (rs10811661). In physically active individuals, the major T allele was associated with increased T2D risk, as in prior reports of non-stratified populations (151, 165, 184, 185, 188, 199, 386). Two hour glucose concentrations have also been shown to be elevated in non-diabetic individuals carrying the rs10811661 T allele (151). The CDKN2A/B x physical activity interaction observed here provides partial confirmation of a recent report from the Diabetes Prevention Program (289), where the effect of the rs10811661 polymorphism on diabetes risk was abolished with intensive lifestyle intervention. Thus, the interaction effects reported in the present study and the study from the Diabetes Prevention Program are directionally consistent.

Data from several previously published intervention and association studies suggest that physical activity modifies the effects of the Pro12Ala (rs1801282) variant on the risk of T2D (289). In our study (Paper III) the data supporting gene x physical activity interactions at the PPARG rs1801282 locus was of nominal statistical significance, but did not withstand correction for multiple testing. Of note, PPARG is a strong biologic candidate for gene x nutrient (i.e. fatty acids) interactions (12, 178). Hence, because dietary patterns and physical activity tend to coalesce within free-living populations, our observation of PPARG gene x physical activity interaction (and previous reports) may be attributable to dietary correlates of physical activity, rather than direct biologic interactions between PPARG and physical activity. Unfortunately measures of dietary fat intake were unavailable in the present study. Because the interaction between PPARG rs1801282 and physical activity is of nominal statistical significance for IGR risk and no interaction was observed on 2-h glucose concentrations or T2D incidence, one should also consider the possibility that the interaction on IGR-risk is attributable to chance.

The last paper in this thesis focused an FTO gene variant (Paper IV). The two initial reports suggested that physical activity modifies the effects of two separate FTO variants (rs9939609 and rs1861868) on BMI (261, 262). In a third report (285), a 1 year programme of intensive lifestyle intervention did not
influence the effects of the FTO rs9939609 variant on weight change; however, the genetic effect on gain in subcutaneous adipose tissue during this period tended to be attenuated by intensive lifestyle intervention ($P$ for gene x treatment interaction = 0.05). These reports are interesting as they suggest that it may be possible to offset the deleterious health effects of FTO variants through behavioural change. In our study we found no such evidence of interaction. Since Paper IV was published, a large cohort study reported a weak non-significant interaction effect between the FTO rs1121980 genotype and physical activity on BMI; the nature of the interaction effect reported therein was consistent with the original report (263). Two other smaller cohort studies have also been published, reaching similar conclusions (264, 265). There are now around ten published reports on this topic, but none has provided conclusive support for the initial observation. It is possible that the replication attempts have been inconclusive because the initial report was false positive, the interaction effect is smaller than observed in the initial study (a concept referred to as the winner’s curse) and replication studies are underpowered to detect small interaction effects, the replication studies are different in design to the initial study, or that other factors modify the interaction effect that are not accounted for in the replication studies.

5.6 How does this work aid our understanding of the human biology of T2D?

This thesis provides evidence of gene x physical interactions on T2D or related traits for several genetic markers. To further explore these interactions, replication in independent cohorts and studies in vivo as well as in vitro are required in order to elucidate the role of these interactions in disease. Paper II and III show a possible gene x environment interaction which may provide better insight into the mechanisms of T2D. Gene x physical activity interactions may arise due to differential gene expression caused by sequence polymorphisms in the regulatory elements of the gene or differences among individuals in the amino acid sequences of a gene product (424). Although research on molecular genetics of physical activity-related phenotypes is still in its infancy these interactions may indicate a phenomenon of biological interest where a particular genetic effect may operates only in the presence of physical activity, or vice versa. Elucidating such effects may lead to a better understanding of possible molecular mechanisms and pathways in T2D progression (425). Alternative explanations for gene x environment interactions relate to epigenetic effects (426); but no epidemiological examples of gene x physical activity interactions in T2D risk have yet been identified as having epigenetic origins.

5.7 How might this work aid in the prevention of obesity or T2D?

This thesis provides some evidence that genetic components combine with physical activity to influence the development of T2D. In order to prevent any disease, first we need to identify the modifiable risk factors for that disease, and, ideally, we should know something about how people respond to interventions designed to prevent or treat the disease. Genetic predisposition and physical inactivity are two major risk factors for T2D. From a public health and clinical perspective, this suggests that genetic information may be used to identify
individuals most susceptible to the adverse consequences of sedentary lifestyle (Paper III) and to predict their response to lifestyle intervention, thereby providing a rationale for targeted intervention in genetically susceptible subgroups of the population. However, these studies have many limitations and the effect sizes are small. Thus, targeted intervention or personalized gene exercise therapy is remains unrealistic (416).

There are now around 20 published common variants with robust evidence of association with T2D. Individually, these variants only have a small effect on disease risk, and even in combination they account for only ~3% of the heritability of T2D (427). Thus, many have questioned the clinical relevance of these findings. These studies have been very expensive as they tend to be large and have involved extensive genotyping. Partly because of the costs involved, many anticipate that within the near future these studies and their extensions will lead to clinically useful discoveries. If common genetic risk variants interact with lifestyle behaviours and act in context-dependent ways, it is possible that genetic effects will be enhanced in subsections of the population. If this is true, those effects may be of clinically relevant, as they may highlight ways in which genetic risk can be offset.

5.8 Role of GWAS in studies of gene x environment interaction

Despite the potential relevance of gene x environment interactions for complex diseases traits, such as T2D, little work has been published on new methods for detecting these types of interactions effects using GWAS data. The GWASs published to date have tested only for the main effects of each genetic marker on disease (428, 429) and ignore heterogeneity between exposure classifications caused by interactions (150, 152, 154). Thus, it is possible that scanning for the main effects using current approaches might miss important genetic variants that act in context-dependent ways and have specific effects in subgroups of the population that are not present in other subgroups; for example, genetic effects that may be evident only in physically inactivity persons. It could be argued that testing gene x environment interactions will be too cumbersome, given the power constraints of such analyses (238). However, if there is evidence that an environmental factor contributes to risk and possibly modifies a genetic effect it is relevant to define a testing strategy and develop methods. Thus, the combination of traditional biological candidate gene methods and current GWAS methods (i.e. to prioritize specific genes based on their known biology) might be an effective approach for the discovery of loci involved in gene x environment interactions. Other approaches has also been suggested; for example, recently, a two-step analysis involving the GWAS method was proposed to identify loci involved in gene x environment interactions. This approach is performed independently of initial scans for main effects, with emphasis in SNPs that demonstrate heterogeneity between subgroups defined by an environmental exposure (430). Other studies might seek to test time-dependent genetic effects, as these are suggestive of gene x environment interactions in populations where the environment has changed with time; to our knowledge only one GWAS of longitudinal birth cohorts has reported joint on the investigation of environmental and genetic influences on complex traits (431).
5.9 Unanswered questions and future research

Complex diseases are based on the implicit assumption of a dynamic cooperation between shared genetic and lifestyle risk factors. However, several questions remain unanswered such as: “Are there gene x lifestyle interactions that play a role in T2D and obesity?”; “What are the mechanisms that might underlie gene x lifestyle interactions on T2D and obesity?”; “How might gene x lifestyle interactions vary across age?”; “How might gene x lifestyle interactions vary in health and disease?”; “Can we use information from gene x lifestyle interactions to target individuals most susceptible to the adverse consequences of sedentary behaviour?”; and “Can we use information from studies of gene x lifestyle interaction to predict people’s responses to lifestyle interventions?”. These are some of the main questions relevant to this field that presently lack good answers. These questions and the work that has led to them also highlight how complex the architecture of gene x lifestyle interactions on obesity and T2D is likely to be.

The work contained in this thesis can be used to form several recommendations for future work in this field (300, 432):

- Study designs - identify genes through GWAS and subsequently test them as candidate genes with priority in intervention studies; examine genes from large-scale observational studies (with priority in cohorts followed prospectively);
- Develop study designs that can discover new genes responsible for gene x physical activity and gene x nutrient interactions on obesity and T2D (i.e. genome-wide interactions);
- Develop studies that examine whether there are gene x obesity and T2D interactions influencing physical activity and nutrient intake;
- Improve the precision with which phenotypes and lifestyle behaviours (nutrient and physical activity components) are measured in studies of gene x lifestyle interaction;
- Develop methods of standardizing lifestyle data in order to compare with other studies;
- Quantify the phenotypic responses to a standardized change in lifestyle;
- Investigate molecular mechanisms;
- Target candidate genes derived from animal studies for subsequent human testing;
- Strive toward undertaking studies that include large sample sizes, statistical power, corrections for multiple testing and replication in independent cohorts, as these will be necessary for the detection of small interaction effects.

Post-genomic approaches have been proposed for gene x environment interactions as a strategy to advance the scientific understanding of health and disease (433). These approaches include the use of genomics (genetic variation), transcriptomics (gene expression), proteomics (gene products), metabolomics (metabolic effects), epidemiological data and microbial (gut microflora) data. The purpose of this research is to integrate these multiple types of data and
perspectives, based on biological models, into a global model in order to dissect the mechanisms and prognostic markers of disease, all of which have therapeutic implications.
6 CONCLUSIONS

6.1 Paper I

This study provides suggestive evidence of association between PPARGC1A gene variation and cardiovascular or metabolic traits, but none of these findings was conclusive. The strongest effect is of interest because of the biologic evidence implicating PPARGC1A in glucose regulation and may be worthy of follow-up in other populations. This study does not support a role for associations between the Gly482Ser and cardiometabolic traits in Northern European white youth. Lastly, no interactions between PPARGC1A SNPs and physical activity were observed for cardiovascular or metabolic traits.

6.2 Paper II

This study supports the notion that physical activity modifies the effects of the TCF7L2 gene variants on HOMA-B, and that these variants also influence growth and energy metabolism early in life.

6.3 Paper III

This study supports the notion that physical activity modifies the effects of the HNF1B and CDKN2A/B gene variants on the risk of impaired glucose regulation and also modifies the effects of the HNF1B on T2D risk.

6.4 Paper IV

This study does not support the notion that physical activity modifies the effect of the FTO variant on body mass in a large population-based cohort of non-diabetic Scandinavian adults.
7 ACKNOWLEDGEMENTS

No thesis is the accomplishment of one author alone. Many persons have been involved in the subprojects of this thesis and have contributed with their time and professional skills to make this thesis possible. Thank you all!

In particular, I would like to express my gratitude to:

**Paul Franks**, my supervisor, whose wisdom, knowledge and commitment to the highest standard have guided me through the research project. Thank you for sharing your scientific experience even with a “stubborn student”. It has been a true privilege!

**Frida Renström**, my co-supervisor and dear friend, for all the excellent support, encouragement, enthusiasm, and expertise. Your organisation skills are truly ‘Swedish’!

**Owe Johnson** and **Tommy Olsson**, Professors from the Department of Public Health & Clinical Medicine, Section for Medicine, Umeå University Hospital, Umeå, Sweden, for providing me with excellent support and research facilities.

**José Alberto Duarte**, Professor of Sports Faculty of the University of Porto, Portugal, whose broad knowledge in physiology and genetics initially involved me in the PhD project of genetics.

**Nick Wareham**, Director of the Medical Research Council (MRC) Epidemiology Unit, Cambridge, UK, for giving me the opportunity to work for one year at the MRC Unit in Cambridge as a Visiting Scientist.

**Ruth Loos**, Programme Leader in Genetic Epidemiology at the **MRC Epidemiology Unit**, Cambridge, who opened the door for me to study “Population Genetics”.

**Jing Hua and Jian'an Luan**, for teaching me statistics with SAS Software.

**Matt Sims** and the technical team (MRC Epidemiology Unit Cambridge, UK), for expert technical assistance with genotyping.

Technical support from **the staff at the CRC in Malmö** for genotyping assistance.

**Soren Brage, Ulf Ekelund, Karsten Fröberg, Valeriya Lyssenko, Göran Berglund, Peter M. Nilsson, Leif Groop, Lars Bo Andersen, Anna Jonsson, and Bo Isomaa** for collaboration on the papers in this thesis.

**Kerstin Rosenqvist, Eva Karlsson, Catrin Johansson, and Tobias Näslund** for administrative help with remarkable efficiency.

**Camilla Franks**, for making my English so much more English!
Sven-Anders Bergström, Ingegerd Söderström, Caroline Blomquist, Agneta Sandström, Xiaolei Hu, Juliet Evans, Therese Andersson, Stefan Söderberg, Inger Arnesjö, Sigbritt Rasmuson, Per Dahlqvist, Lars-Göran Sjöström, Susanne Sandberg, Magnus Strand, Caroline Mellberg, Mats Ryberg, Kotryna Simonite, Jonas Buren, Malin Alvehus, Jonas Andersson, Julia Otten, Anna Gradmark, Monica Holmgren, and Nicklas Fridberg for nice group meetings in Sweden.

Luis Sardinha, my “X”, thanks for everything.

Vimaleswaran Santhanakrishnan, my dear Indian friend, for all the stimulating discussions, seminars and social events that we have attended together, and all the joy we have had together in Cambridge.

Lena Larsson, my dear Swedish friend, for your company during lunches and dinners and seminars that we have attended together, but most of all the happiness and laughs we have shared in Sweden.

Margarida Espanha, my sister from the heart and friend with faith, for all the support with my children.

Aldir Alba, my special friend, thank you for being a part of my life!

In memory of my father Carlos and my brothers Baltasar and Carlos “You would have liked this small project”.

Baltasar, my dear nephew, always with good advice, “like an adult”.

My mother Luisa, “the rock”, for unconditional love and support in all aspects of life.

My son Luis Afonso, and daughter Inês, for supporting me even when I was not at home. I love you!

I am very grateful to the children (European Youth Heart Study) and adults (Malmö Preventive Project and Prevalence, Prediction and Prevention of Diabetes in Botnia) who participated in this work.

This thesis was carried at the Genetic Epidemiology & Clinical Research Group, Department of Public Health & Clinical Medicine, Section for Medicine, Umeå University Hospital, Umeå, Sweden. This work was supported by grants from the Påhlssons Foundation, the Novo Nordisk, the Swedish Heart-Lung Foundation, the Swedish Diabetes Association, the Knut and Alice Wallenberg Foundation, the Sigrid Juselius Foundation, The Folkhälsoan Research Foundation, the Nordic Center of Excellence in Disease Genetics, an EU grant (EXGENESIS), the Signe and Ane Gyllenberg Foundation, the Swedish Cultural Foundation in Finland, the Finnish Diabetes Research Foundation, the Foundation for Life and Health in Finland, the Finnish Medical Society, the Paavo Nurmi Foundation, Helsinki University Central Hospital Research...
Foundation, the Perklén Foundation, the Ollqvist Foundation, the Närpes Health Care Foundation, the Danish Heart Foundation, the Danish Medical Research Council Health Foundation, the Danish Council for Sports Research, the Foundation in Memory of Asta Florida Bolding Renée Andersen, the Faculty of Health Sciences, University of Southern Denmark, the Estonian Science Foundation grant numbers 3277 and 5209, and the Medical Research Council, UK. ECB was supported by a studentship from the Portuguese Foundation for Science and Technology. PWF was supported with grants from the Swedish Diabetes Association, the Swedish Heart-Lung Foundation, Novo Nordisk and Västerbottens Health Authority (ALF).
REFERENCES


132. (CDC) CfDCaP. Number of people with Diabetes continues to increase. 2008.


199. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007 Jun 7;447(7145):661-78.


211. Florez JC, Burtt N, de Bakker PI, Almgren P, Tuomi T, Holmkvist J, et al. Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor


