

Genes, Lifestyle and Coronary Heart Disease Risk

Epidemiological Interaction Studies

Jaana Gustavsson

Department of Public Health and Community Medicine
Institute of Medicine
Sahlgrenska Academy at University of Gothenburg



UNIVERSITY OF GOTHENBURG

Gothenburg 2015

Cover illustration:

Picture collage by Jaana Gustavsson

Genes, Lifestyle and Coronary Heart Disease Risk

© Jaana Gustavsson 2015

jaana.gustavsson@amm.gu.se

ISBN (printed) 978-91-628-9344-6

ISBN (e-publication) 978-91-628-9345-3

The e-version of this thesis is available at <http://hdl.handle.net/2077/37999>

Printed in Gothenburg, Sweden 2015

Kompendiet

“It is a fundamental truism, of logic more than of genetics, that the phenotypic ‘effect’ of a gene is a concept that has meaning only if the context of environmental influences is specified, environment being understood to include all the other genes in the genome. A gene ‘for’ A in environment X may well turn out to be a gene for B in environment Y. It is simply meaningless to speak of an absolute, context-free, phenotypic effect of a given gene.”

Richard Dawkins, *The Extended Phenotype*

Oxford University Press Inc., New York, 1999

Genes, Lifestyle and Coronary Heart Disease Risk Epidemiological Interaction Studies

Jaana Gustavsson

Department of Public Health and Community Medicine
Sahlgrenska Academy at University of Gothenburg, Sweden

ABSTRACT

Coronary heart disease (CHD) has multifactorial background involving both genetic and lifestyle factors, but much is still unknown about their interactions. The aim of this thesis was to study interactions focusing on apolipoprotein E (*APOE*), fat mass and obesity-related (*FTO*) and ghrelin/obestatin prepropeptide (*GHRL*) genes, as well as smoking, physical activity and diet. The study sample included 1831 cases with CHD (myocardial infarction or unstable angina) and 5175 population controls from two population-based studies: SHEEP, Stockholm and INTERGENE, Gothenburg. Interaction was assessed on the relative risk (RR) and risk difference (RD) scales.

APOE-smoking interaction was found both on the RR and RD scales, so that subjects carrying the $\epsilon 2$ allele had lower smoking-related CHD risk, adjusted OR 1.35 (95% CI 0.92-1.97) than non-carriers, with OR 2.17 (95% CI 1.82-2.59) in subjects with common genotype $\epsilon 3\epsilon 3$ and OR 2.43 (95% CI 1.88-3.14) in $\epsilon 4$ carriers. Women carrying the $\epsilon 4$ allele had particularly high smoking-related CHD risk with OR 3.69 (95% CI 2.33-5.83). A potential *APOE*-physical activity interaction was also observed, where the $\epsilon 2$ allele counteracted while the $\epsilon 4$ allele (vs $\epsilon 3\epsilon 3$) potentiated CHD risk from physical inactivity.

Carriers of the *FTO* single nucleotide polymorphism (SNP) rs9939609 A allele (TA/AA vs TT) had increased CHD risk with OR 1.20 (95% CI 1.06-1.37), independent of body mass index (BMI). No evidence of interaction between *FTO* and physical activity was found, indicating that *FTO*-related CHD risk is not counteracted by increased physical activity. No clear interactions between *FTO* and macronutrients were found with a dichotomous variable of below/above median energy% intake. With a continuous energy% variable, excluding subjects reporting diet change, however, interaction was observed on the RR scale for *FTO*-fat and *FTO*-saturated fatty acids, suggesting slightly increased *FTO*-related CHD risk with lower energy% of fat or saturated fatty acids.

Finally, a gene-gene interaction was found for SNPs *FTO* and *GHRL* rs35680 in a subsample of 420 INTERGENE controls, where the minor alleles had synergistic effects on BMI, supporting a mechanistic *FTO*-*GHRL* link behind obesity.

To conclude, identification of gene-lifestyle interactions may contribute to enhanced understanding of mechanisms causing CHD.

Keywords: Coronary heart disease, gene-lifestyle interaction, *APOE*, *FTO*, *GHRL*, smoking, physical activity, diet, obesity

ISBN (printed): 978-91-628-9344-6 **ISBN (e-publ):** 978-91-628-9345-3

SAMMANFATTNING PÅ SVENSKA

Kranskärslssjukdom, d.v.s. hjärtinfarkt och angina (kärlekkramp), beror av en inflammatorisk process med åderförfattning i hjärtats kranskärsl som leder till att blodflödet helt eller delvis förhindras, vilket orsakar syrebrist. Sjukdomen orsakas av en mängd faktorer, såväl gener som livsstil (t.ex. rökning, fysisk inaktivitet och ohälsosam kost). Dessa faktorer kan samverka (interagera) men mycket är ännu okänt om sådana interaktioner.

Syftet med denna avhandling var att studera interaktioner mellan varianter i generna *APOE* (Apolipoprotein E), den fetma-relaterade genen *FTO* och *GHRL* (för aptit-stimulerande hormon grelin) samt livsstils-faktorerna rökning, fysisk aktivitet på fritiden och kost, på risken för kranskärslssjukdom. Studiepopulationen bestod av sammanlagt 1831 patienter med kranskärslssjukdom och 5175 kontroller från två populations-baserade studier: SHEEP i Stockholm och INTERGENE i Göteborg.

Vi fann en interaktion mellan *APOE* genen och rökning, så att personer med $\epsilon 2$ allelen (ca. 15% av befolkningen) var delvis skyddade mot den rökning-relaterade risken för kranskärslssjukdom med en oddskvot på 1.35 (95% konfidensintervall 0.92-1.97). Personer med andra *APOE* alleler hade en betydligt högre rökning-relaterad risk med oddskvot 2.17 (1.82-2.59) hos personer med den vanligaste genotypen $\epsilon 3\epsilon 3$ (ca. 60%). Kvinnor med $\epsilon 4$ allel hade en särskilt hög risk från rökning med oddskvot 3.69 (2.33-5.83). En tendens till interaktion av liknande karaktär fanns mellan *APOE* och fysisk inaktivitet (d.v.s. mestadels stillasittande fritid), så att bärare av $\epsilon 2$ allelen var delvis skyddade mot en ökad sjukdomsrisk relaterad till fysisk inaktivitet, medan bärare av $\epsilon 4$ allelen var mer utsatta för risken från fysisk inaktivitet.

En *FTO* gen-variant med känt samband med fetma, gav även en något förhöjd risk för kärslssjukdom, oddskvot 1.20 (1.06-1.37). Detta samband tycktes dock inte bero av fetma-effekten, vilket tyder på att någon annan, ännu okänd, effekt av *FTO* förklarar sambandet med kärslsjukdom. En ökad nivå av fysisk aktivitet har tidigare setts minska *FTO*-fetma sambandet, men vi fann inga sådana tendenser för sambandet med kärslsjukdom, som var oberoende av nivån av fysisk aktivitet. Det fanns heller inga starka tecken på att kostens sammansättning av fett, kolhydrater och proteiner påverkade *FTO* sambandet med kärslsjukdom.

Slutligen fann vi en interaktion mellan varianter i *FTO* och *GHRL* generna för sambandet med övervikt/fetma (mätt som *BMI=body mass index*), som visade att risk-varianterna av båda gener förstärker varandras effekt.

Identifiering av interaktioner mellan gen-varianter och livsstils-faktorer kan bidra till en ökad förståelse för orsaker till kranskärslssjukdom, som i förlängningen kan möjliggöra individanpassad prevention och behandling.

LIST OF PAPERS

This thesis is based on the following papers, referred to in the text by their Roman numerals.

- I. Jaana Gustavsson, Kirsten Mehlig, Karin Leander, Elisabeth Strandhagen, Lena Björck, Dag S. Thelle, Lauren Lissner, Kaj Blennow, Henrik Zetterberg, Fredrik Nyberg.
Interaction of apolipoprotein E genotype with smoking and physical inactivity on coronary heart disease risk in men and women.
Atherosclerosis 2012; 220: 486-492
- II. Jaana Gustavsson, Kirsten Mehlig, Karin Leander, Lauren Lissner, Lena Björck, Annika Rosengren, Fredrik Nyberg.
FTO genotype, physical activity and coronary heart disease risk in Swedish men and women.
Circulation Cardiovascular Genetics 2014; 7: 171-177
- III. Jaana Gustavsson, Kirsten Mehlig, Karin Leander, Christina Berg, Gianluca Tognon, Elisabeth Strandhagen, Lena Björck, Annika Rosengren, Lauren Lissner, Fredrik Nyberg.
FTO gene variation, macronutrient intake and coronary heart disease risk: a gene-diet interaction analysis.
European Journal of Nutrition doi:10.1007/s00394-015-0842-0
- IV. Jaana Gustavsson, Kirsten Mehlig, Elisabeth Strandhagen, Karin Leander, Kaj Blennow, Henrik Zetterberg, Annika Rosengren, Dag S. Thelle, Fredrik Nyberg, Lauren Lissner.
FTO and GHRL gene-gene interaction on body mass index.
Submitted manuscript

CONTENT

ABBREVIATIONS	V
DEFINITIONS	VI
1 INTRODUCTION.....	1
1.1 Coronary heart disease	1
1.1.1 Etiology of CHD	2
1.1.2 Genetic factors.....	3
1.1.3 Lifestyle factors.....	6
1.1.4 Gene-lifestyle interaction	8
1.2 Definitions of interaction	9
1.2.1 Biological interaction	9
1.2.2 Statistical interaction	11
1.2.3 Notes on interaction analysis.....	12
2 AIMS	14
3 METHODS	15
3.1 Study sample	15
3.1.1 The SHEEP study.....	16
3.1.2 The INTERGENE study.....	16
3.1.3 Pooling of SHEEP and INTERGENE.....	17
3.2 Exposure assessment and definitions	17
3.2.1 Data collection and clinical examination	17
3.2.2 Smoking	18
3.2.3 Physical activity	18
3.2.4 Dietary data	19
3.2.5 Genetic data.....	20
3.2.6 Other variables	21
3.3 Statistical analyses	22
3.3.1 Interaction analyses	23
3.3.2 Power considerations.....	24

4	RESULTS	26
4.1	Characteristics of the study population	26
4.2	Main effects of lifestyle factors on CHD	27
4.3	APOE and lifestyle factors	28
4.4	FTO and physical activity	31
4.5	FTO and dietary macronutrients	33
4.6	FTO and GHRL interaction on BMI	36
5	DISCUSSION	38
5.1	APOE and lifestyle factors	38
5.2	FTO and physical activity	39
5.3	FTO and dietary macronutrients	40
5.4	FTO and GHRL interaction on BMI	42
5.5	General discussion of interaction findings	43
5.6	Methodological considerations	44
5.6.1	Study design	44
5.6.2	Potential systematic bias	45
5.6.3	Residual confounding	49
5.6.4	Reverse causation	50
6	CONCLUSIONS	51
7	FUTURE PERSPECTIVES	53
	ACKNOWLEDGEMENT	55
	REFERENCES	56
	APPENDIX	68

ABBREVIATIONS

<i>APOE</i>	Apolipoprotein E gene
BMI	Body mass index
BMR	Basal metabolic rate
CHD	Coronary heart disease
E%	Energy percentage (from total, non-alcohol, energy intake)
FFQ	Food frequency questionnaire
FIL	Food intake level
<i>FTO</i>	Fat mass and obesity associated gene
<i>GHRL</i>	Ghrelin/obestatin prepropeptide gene
GWAS	Genome wide association study
LDL-C	Low density lipoprotein cholesterol
MI	Myocardial infarction
MUFA	Mono-unsaturated fatty acids
PAL	Physical activity level
PUFA	Poly-unsaturated fatty acids
RERI	Relative excess risk due to interaction
SFA	Saturated fatty acids
SNP	Single nucleotide polymorphism

DEFINITIONS

Genome	The complete DNA sequence of an individual
Genotype	The combination of two alleles at one gene locus, one from each chromosome
Hardy-Weinberg equilibrium	When the alleles of a certain gene locus are distributed in proportion to the frequencies for the alleles in a population and remain constant from generation to generation. For two alleles with frequencies p and q , $(p+q)^2 = 1$
Linkage disequilibrium	The tendency of alleles at linked loci on the same chromosome to occur together more frequently than expected by chance
Locus	The position of a gene or specific DNA sequence on a chromosome
Phenotype	The observed physiological, morphological and biochemical characteristics of an individual, determined by the genome and environment
Single nucleotide polymorphism	Variation in a single nucleotide (A, T, G or C) on a certain position in the chromosome

1 INTRODUCTION

Coronary heart disease (CHD) has a complex and multifactorial background including both genetic and lifestyle factors, and often their interactions. Characterization of such interactions will contribute to an increased understanding of the mechanisms causing CHD.

1.1 Coronary heart disease

Cardiovascular disease including CHD, cerebrovascular disease or stroke, peripheral vascular disease and hypertension, is the leading cause of death in high- and middle-income countries (1). CHD is the most common cardiovascular disease, and presents clinically mostly as myocardial infarction (MI) or as angina pectoris. An MI occurs when the blood supply to part of the myocardium is interrupted due to an occlusion of a coronary artery, causing ischemia and irreversible damage to the heart. Angina pectoris occurs when there is a reduced blood flow in the coronary arteries which leads to oxygen depletion, typically causing chest pain. Stable angina pectoris appears during physical activity when the oxygen requirement is increased, while unstable angina pectoris appears also at rest.

Death rates from CHD have fallen the last decades in all Western European regions. There is a large gender difference in CHD death rates, with an age-standardized death rate in people aged 0-64 in the European Union of 38 per 100 000 in men, and 9 per 100 000 in women (2). In Sweden, incidence rates of MI as well as mortality from MI have declined since the 1980s. In 2012, the age-standardized incidence of acute MI was 415 per 100 000 Swedish inhabitants aged 20 and above, with the incidence increasing with age and being higher in men than in women (3).

CHD is also associated with obesity, dyslipidemia, insulin resistance (components of the metabolic syndrome) and diabetes. In the global INTERHEART study, nine modifiable risk factors were found to predict over 90% of the risk of MI (4). These were the apolipoprotein B/apolipoprotein A1 ratio (ApoB/ApoA1), hypertension, diabetes, abdominal obesity, smoking, physical inactivity, a high risk diet (i.e. high intake of meat, fried or salty food and low intake of fruit and vegetables), alcohol intake and psychosocial factors (e.g. stress or depression).

1.1.1 Etiology of CHD

CHD is the ultimate manifestation of a long-term metabolic imbalance and inflammation of the endothelial wall of the arteries, called atherosclerosis. The process generally starts early in life but symptoms may not appear until after decades. Initially, there is an accumulation of lipids that are deposited in the endothelium forming fatty streaks, with attraction of monocytes and leukocytes to the site. Oxidation of lipids increases the inflammatory response and macrophages take up oxidized low density lipoprotein (LDL) particles and form so-called foam cells. Smooth muscle cells grow and a fibrotic/calcified layer, or plaque, is formed, causing stiffness and narrowing the arteries. The atherosclerotic plaques may become vulnerable and rupture which may lead to complete occlusion of the vessel (5).

Blood lipids including mainly cholesterol and triglycerides (TG) play a central role in atherosclerosis. Cholesterol is an essential building block of cell membranes, and is necessary for the production of steroidal hormones and vitamin D. Most cholesterol in the body is synthesized in the liver, but part is also absorbed from food, although absorption is highly regulated. Lipids are transported in blood in lipoproteins that contain apolipoproteins (Apo) and lipids. In the exogenous pathway, cholesterol and TGs are absorbed from the gastrointestinal tract and transported in chylomicrons to cells, where the lipids are taken up as free fatty acids. The chylomicrons are also taken up by receptors in the liver, where cholesterol is stored, metabolized or transported further to peripheral tissues. In the endogenous pathway, cholesterol and TG in the liver are packaged in very low density lipoproteins (VLDL) containing ApoB, which are secreted into the blood. VLDL releases TGs as free fatty acids into muscles or adipose tissue. The VLDL particles become smaller LDL particles, which provide the source of cholesterol uptake in peripheral tissues. ApoE mediates cellular cholesterol uptake by interaction with cell surface receptors including the LDL-receptor. LDL particles are easily oxidized, thereby contributing to atherosclerosis. Cholesterol is transported back from cells to the circulation and liver by high density lipoproteins (HDL) containing ApoA1, which can also capture cholesterol from atherosclerotic lesions, thus preventing atherosclerosis. Overall, the balance between the amount of LDL and HDL particles determines whether there is excess cholesterol transported from the liver to peripheral tissues, or the opposite, cholesterol transported from peripheral tissues back to the liver (6). This is why the ratio ApoB/ApoA1 provides a predictive measure of CHD risk (4).

1.1.2 Genetic factors

Evidence suggest that 40-60% of the risk of CHD is heritable, and being of greater importance in early onset (7, 8). There are multiple genetic variants contributing to CHD, each of them with typically minor effect (7).

Genetic variation originates from mutations of the nucleotide sequence of deoxyribonucleic acid (DNA) that constitutes the genetic material. DNA is built up of the four nucleotides adenine (A), cytosine (C), guanine (G) and thymine (T). The double-stranded DNA helix is formed by pairs of nucleotides, where A on one strand is complementary to T on the other strand and C is complementary to G. Genetic variation commonly appears as single nucleotide polymorphism (SNP), where a nucleotide is replaced by another. Other forms of genetic variation are deletions or insertions of one or several nucleotides, inversions or translocations of nucleotide sequences, or repetitions of DNA-segments. SNPs occur at an average of one per 1000 nucleotides, which sums up to about 3,000,000 SNPs for the entire human genome (9). If the SNP appears in a coding region of a gene, i.e. a DNA sequence that determines the amino acid sequence of a protein, it may alter the function of the protein. Interestingly, most genetic variants are located in non-coding regions e.g. promoter regions or introns (7), but they can still affect expression or splicing of genes and thereby indirectly alter the function of a protein.

Up to year 2007, most studies of the genetics behind CHD were conducted with a candidate gene approach, i.e. studying variation in genes with known mechanism e.g. related to hyperlipidemia or inflammation, thus hypothesized to be involved in the pathogenesis of CHD. Nearly 200 variants in more than 100 genes have been identified by this approach (10). However, many of the studies were performed in small samples with a higher risk of false positive findings, and accordingly, many of the findings have been poorly reproduced.

Since 2007, genome-wide association studies (GWAS) have been conducted that use a hypothesis-free approach where up to 1,000,000 genetic markers or more across the genome are compared in cases with disease and control subjects without disease in large samples, to identify SNPs that are associated with disease (9). These studies, mostly conducted in subjects of European descent, have identified 45 common SNPs (i.e. occurring at a frequency of at least 1%) reaching genome-wide significance for association with CHD (11, 12). The individual effect size of each SNP is typically a per allele odds ratio of <1.2. It was estimated that around 10% of the total heritability of CHD is explained by these common variants (12). Future GWAS in even larger

samples will most likely identify additional SNPs with even smaller effect size. There are also probably many rare genetic variants (i.e. occurring at a frequency below 1%) that contribute to disease, but the currently available standard chips used in GWAS have too low density to be able to capture rare variants, and even larger sample sizes would be required (7).

An analysis of the functional areas of the SNPs identified in GWAS revealed that lipid metabolism and inflammation were the key biological pathways to CHD, where 12 SNPs were directly associated with blood lipids (total cholesterol, LDL cholesterol [LDL-C], HDL cholesterol or TG levels). There was also association with blood pressure and obesity-related traits (body mass index [BMI] and waist-hip ratio), while the association with type 2 diabetes and glucometabolic traits, e.g. fasting insulin and glucose concentrations, was less clear (12).

One of the well-known genes involved in lipid metabolism with relation to CHD is *APOE* (13). *APOE* codes for apolipoprotein E (apoE), which is involved in clearance of circulating cholesterol by acting as ligand for cellular lipoprotein receptors. It also influences platelet aggregation, inhibition of smooth muscle cell proliferation, protects from oxidation and participates in neuronal repair (14). The epsilon (ϵ) polymorphism of *APOE* results in three major protein isoforms E2, E3 and E4 coded by the co-dominant common alleles $\epsilon 2$, $\epsilon 3$ and $\epsilon 4$, of which $\epsilon 3$ is the most common (75-80% allele frequency in most populations). The alleles differ at two SNPs, giving rise to different amino acids at protein positions 112 and 158; cysteine at 112 and arginine at 158 for E3, two cysteine for E2 and two arginine for E4 (15). The three isoforms differ in their LDL-receptor affinity, antioxidant properties ($E2 > E3 > E4$), and inflammation modulatory properties. *APOE* genotypes have an approximately linear relationship with LDL cholesterol and CHD risk when ordered $\epsilon 2\epsilon 2$, $\epsilon 2\epsilon 3$, $\epsilon 2\epsilon 4$, $\epsilon 3\epsilon 3$, $\epsilon 3\epsilon 4$, $\epsilon 4\epsilon 4$ (16). The association with HDL cholesterol is weaker, but indicates higher levels in $\epsilon 2$ carriers and lower in $\epsilon 4$ carriers. A more recent meta-analysis showed that CHD risk is decreased in $\epsilon 2$ carriers but only marginally increased in $\epsilon 4$ carriers compared to $\epsilon 3$ homozygotes (16), which contrasted with previous meta-analyses where $\epsilon 4$ risk increase was more prominent (17, 18).

A gene that has been identified in the more recent GWAS, with relevance for CHD, is the fat mass and obesity associated gene (*FTO*) (19-23). The gene was first identified in 2007 for variation predisposing to obesity and type 2 diabetes, mainly through increased fat mass (24-26). The SNP rs9939609 minor A allele has also been associated with an atherogenic lipid profile,

elevated plasma CRP levels and hypertension (19, 27, 28). A study in >17 000 Europeans showed that the *FTO* association with a range of metabolic traits (including LDL and HDL cholesterol, TG, glucose and insulin) was entirely due to its effect on BMI (29). The exact mechanism of *FTO* is still unclear, but it has a role in regulation of energy balance, possibly both energy intake and expenditure (30-33). *FTO* is expressed in neurons, especially in the arcuate nucleus in the hypothalamus, in sites where appetite and energy expenditure are regulated (34). The gene is also expressed in peripheral nerves and may have additional, as yet unknown, physiological effects (35). Experimental mouse models have shown that *FTO* knockout mice had a reduction in both fat mass and lean mass, increased metabolic rate and food intake, and decreased propensity to gain weight on a high-fat diet (36). A mouse model overexpressing *FTO* had increased body and fat mass, especially on a high-fat diet, increased food intake but no change in energy expenditure or physical activity (37). The risk allele has been linked to overexpression of the gene, which leads to increased body and fat mass, which was even augmented on a high-fat diet (38).

A recent study has demonstrated a functional link between *FTO* and the orexigenic hormone ghrelin, coded by the ghrelin/obestatin prepropeptide gene (*GHRL*), in appetite and food reward mechanisms, where *FTO* rs9939609 AA subjects had attenuated suppression of hunger and higher circulating acyl-ghrelin levels in the postprandial state (39). The study also showed that *FTO* AA compared to TT subjects had increased *FTO* expression in peripheral blood cells, with increased demethylation of *GHRL* mRNA resulting in increased *GHRL* expression.

Despite the advances in identifying genetic variants contributing to CHD, especially since the introduction of GWAS, much research is still required to explain the 'missing' heritability of CHD and the function of the genetic variants identified (7). Part of the 'missing' heritability is probably explained by gene-environment interactions. Most common variants associated with CHD may be regarded as susceptibility variants that only contribute to disease in certain contexts of other genetic and environmental stressors such as smoking (10, 40). This has implications for the detection of genetic risk variants. Depending on the lifestyle risk exposure in the studied population, an effect of certain gene variants may not be visible, or the effect may vary between populations (13). Further, if the genetic associations with CHD are adjusted for lifestyle factors that interact with the genetic factor, then the genetic effect may be diluted (41). There is also heterogeneity between different populations in genetic risk variants. Thus, there are many difficulties in the identification of genetic factors contributing to CHD.

1.1.3 Lifestyle factors

Lifestyle factors are important predictors of the risk of CHD. Partly, this is a consequence of modern lifestyle that is generally too sedentary, with unhealthy dietary habits, smoking and stress. Although many of the lifestyle factors may be altered at the individual level, societal policies that promote and facilitate physical activity, easy access to healthy food and a less stressful work life are of great importance. Collectively, lifestyle risk factors including smoking, physical inactivity and high risk diet account for about 60% of the population attributable risk of MI (4).

Smoking

Smoking increases the risk of CHD in a dose-dependent fashion. Current smoking increased the risk of MI about 3-fold in the INTERHEART study, with a similar effect in men and women, while former smoking increased the risk by roughly 50%, but appeared to have a larger impact in men (4). Smoking partly explains the lower CHD incidence in women than in men especially at younger age, as in most societies women smoke less than men. However, smoking in younger women has increased in some regions, which potentially will increase the risk of CHD in women (42).

Physical activity

A large body of evidence shows that regular physical activity reduces the risk of CHD incidence and mortality (43-46). The physiological effects of physical activity are many, such as improvement of blood lipid profile with increased HDL cholesterol levels, decreased peripheral resistance and reduced blood pressure, improved endothelial function in arteries, improved glucose metabolism and insulin sensitivity, and reduction of inflammation (47-49). The protective effect of physical activity appears to be nearly dose-dependent, at least up to a certain threshold of intense physical activity. The relative risk reduction is greater when comparing a very low to a moderate physical activity level, whereas the beneficial effect of physical activity appears to reach a plateau above which an increased CHD protection is hardly gained. Although vigorous physical exercise improves cardiorespiratory fitness to a larger extent than moderate exercise, moderate activity improves blood lipids and blood pressure equally well (45). Therefore, and also to accomplish greater adherence in the general population, the public health recommendation is regular and moderately intense physical activity of an accumulated 30 minutes per day on most days of the week or preferably every day. The recommended activity intensity is equivalent to brisk walking for most healthy adults (45).

In more recent years, a sedentary behavior typically in the form of prolonged sitting has been shown to be a risk factor on its own, independent of other physical activity (50). This means that even short interruptions of prolonged sitting with intermittent activity e.g. walking of as short duration as 2 minutes has beneficial health effects (51). In the last 50 years, the physical activity pattern in many regions has changed. Strenuous work is much less frequent and an increasing proportion of the population is sitting most of their working time, but also during leisure time e.g. in front of a TV or a computer (52). In Europe, more than 60% of the adult population spends at least 4 hours per day sitting. At the same time, especially in Northern Europe, leisure time physical activity has increased (53).

Studies on physical activity and cardiovascular disease have traditionally focused on leisure time physical activity, while the role of occupational or other activity is less clear (44). The INTERHEART study found that the protective effect of leisure time activity was greater than that of occupational activity, and also that strenuous occupational activity was not associated with reduced risk of MI, and this was not explained by socioeconomic factors (44).

Diet

Dietary patterns with high intake of natural fibre-rich plant foods (vegetables, pulses, fruits, berries, whole grains, nuts and seeds), fish, vegetable oils and low-fat dairy products are associated with lower risk of CHD and other chronic diseases (54, 55). The Mediterranean-type diet is one example of such dietary pattern. In contrast, the so-called Western-type dietary pattern containing high intake of red and processed meat and energy-dense (i.e. high in sugar and fat) food products with low content of micronutrients is associated with chronic diseases and adverse health effects. There is strong evidence that high consumption of processed meat increases the risk of CHD, obesity and type 2 diabetes. High consumption of sugar-rich drinks is linked to increased risk of type 2 diabetes and obesity (54). In general, more recent dietary guidelines focus more on dietary patterns instead of single food items and nutrients (54).

Regarding fat, a high proportion of unsaturated fats is beneficial for health and certain subgroups of essential fatty acids, especially long-chain n-3 fatty acids in fish are important (56). The evidence for total intake of fat in relation to CHD is weaker, and the balance between unsaturated and saturated fats appears more important, so that replacing saturated fatty acids and trans-fatty acids with poly-unsaturated fatty acids reduces the risk of CHD (54). Similarly, the quality of carbohydrates is more important for health than the

total intake. Carbohydrates of low glycaemic index e.g. from whole-grain, vegetables, fruit, pulses, nuts and seeds are more beneficial than sugars or other high-glycaemic load carbohydrates (54, 57).

1.1.4 Gene-lifestyle interaction

Genes and lifestyle are intimately related in the causal pathway to CHD and therefore, investigating these factors in isolation does not provide a full understanding of the pathogenesis of CHD; rather, it is necessary to characterize interactions between these factors (40). Although many gene-environment interactions on CHD have been reported, some of these have not been replicated and it is probable that there are false positive findings (58). There are much fewer examples of replicated, clinically relevant interactions where a plausible mechanism for interaction is identified (58), and therefore there is clearly a need for more studies on gene-lifestyle interactions, both of new candidates for interactions and replication of previous results.

Gene-lifestyle interactions involving the *APOE* polymorphism have been reported in relation to CHD risk, mainly on blood lipids (59). An *APOE*-smoking interaction has been shown in relation to CHD, where $\epsilon 4$ carriers showed higher risk due to smoking than $\epsilon 3$ homozygotes (60-62), but there is little evidence of an interaction in women (63), and some studies do not confirm this interaction (64, 65). There are also reports of *APOE* interaction with physical activity or overweight on blood lipids, but with conflicting results (66-69), and no publications directly on CHD risk. Therefore, there is need for more evidence on *APOE* -lifestyle interaction on CHD risk.

Interaction studies of *FTO* and lifestyle have so far mainly focused on obesity. Increased physical activity has been shown to reduce the *FTO* association with obesity (70). There are two studies on *FTO* -physical activity interaction in relation to cardiovascular outcomes, but with conflicting results. One study in women in the US suggested that the *FTO* risk allele is associated with increased cardiovascular disease risk only in less physically active women (71), while a study in a Swedish cohort demonstrated a stronger association between the risk allele and cardiovascular mortality in more physically active people (72).

Regarding *FTO*-diet interaction, there are conflicting results. Two studies reported that the macronutrient composition of diet modifies the *FTO* effect on obesity, with high fat or SFA and low carbohydrate potentiating the *FTO* effect (73, 74), while a large-scale meta-analysis found no influence either by energy-adjusted intakes of fat, carbohydrates or protein or by total energy

intake on the *FTO* association with BMI (75). Other studies demonstrated that the *FTO* effect on obesity and type 2 diabetes is counteracted by a Mediterranean-type diet (76, 77). Interaction studies of *FTO* -diet in relation to CHD are sparse. Given the suggested interactions on obesity and type 2 diabetes, it is of interest to investigate whether interactions between *FTO* and lifestyle factors are present in relation to CHD.

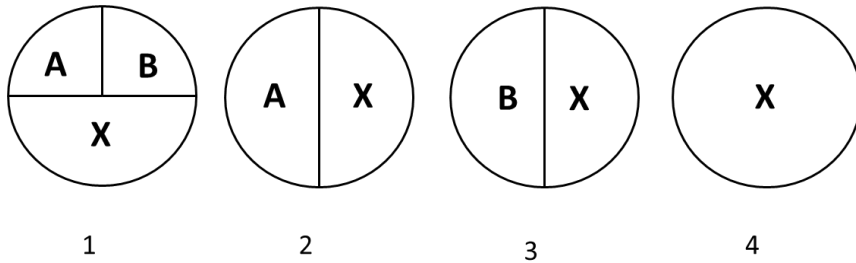
Identifying robust gene-lifestyle (or gene-gene) interactions will contribute to enhanced understanding of mechanistic pathways to disease. In addition, it may explain why certain risk factors are not associated with equal risk of CHD in different populations or geographical regions, as genetic and lifestyle exposures sometimes vary. Ultimately, this may have applications in targeting prevention and treatment to subgroups that are genetically highly susceptible to CHD from a particular exposure (63, 78).

1.2 Definitions of interaction

The term interaction may be used with a general meaning that two or more factors affect one another in the pathway to disease. The notion of interaction between two (or more) risk factors generally describes the situation where the effect of one risk factor on a given outcome depends on the value of the other risk factor(s) (79). However, the term requires a more distinct definition when subject to analysis and mainly two different definitions are used in epidemiological research (80, 81).

1.2.1 Biological interaction

So-called biological interaction, also referred to as causal, mechanistic or sufficient cause interaction (79, 81), means that two or more causal factors are involved in the same causal mechanism (or sufficient cause) for disease (80). This can be illustrated with the causal pie model by Rothman et al (80, 82). Any factors that are component causes (pie slices) of the same causal mechanism (full pie) interact, as shown in Figure 1, pie 1 for factors A and B (82). This means that some cases of disease only occur if factors A and B are jointly present in some individuals. This does not exclude that there are other causal mechanisms for disease, where A and B do not interact. One of the factors A or B may alone be part of one or more other causal mechanisms (pies 2-3, Figure 1), or not part of a causal mechanism at all (pie 4, Figure 1). The opposite of causal interaction is that the two factors are independent, i.e. always part of different causal mechanisms.



*Figure 1. Causal pie model. Four possible causal mechanisms with respect to causal factors (component causes) A and B, with all other unknown component causes denoted as X. In causal pie 1, both A and B are component causes and therefore interact. In pies 2 and 3, either factor A or B alone is a component cause, i.e. no interaction between them. In pie 4, neither factor A or B is a component cause, i.e. no interaction. From Rothman: *Epidemiology -An Introduction*. pp 173-174. New York: Oxford University Press, 2002.*

An important note is that the interacting factors do not necessarily need to directly, or physically, interact for instance in an enzymatic system. A gene codes for a protein, which has a biological effect that may interact with some other biological effect caused by a lifestyle factor somewhere along the causal pathway to disease.

For multifactorial diseases such as CHD, some causal interaction can be assumed to occur for every case of disease (82). In addition, there are most likely many different causal mechanisms, involving different combinations of both genetic and environmental factors.

The empirical criterion for biological interaction is departure from additivity of risks from the interacting factors. Consequently, if two causal factors A and B interact, the risk difference (RD) in those with combined exposure to A and B (R_{AB}) compared with those who have neither exposure, i.e. background risk, (R_0), deviates from the sum of the risk differences for those exposed to A only (R_A) and those exposed to B only (R_B), compared to the background risk (80). The following equation holds in the absence of interaction between two causal factors A and B:

$$R_{AB} - R_A - R_B + R_0 = 0 \quad (1)$$

If the factors in expression (1) are divided by the background risk (R_0), the following equation for relative risks (RR) holds in the absence of interaction:

$$RR_{AB} - RR_A - RR_B + 1 = 0 \quad (2)$$

The expression to the left in equation (2) is called relative excess risk due to interaction (RERI). If $RERI > 0$, then there is more than additive effects on the RD scale, also called synergism. The RERI is intended to be used with risk (or adverse) factors. However, preventive factors can also be studied by reversing the exposure categories so that the high-risk category is the exposure category (i.e. lack or lower level of the preventive factor) (82).

The empirical criterion for biological interaction by Rothman has been formalized further by VanderWeele and Robins. They have shown that the $RERI > 0$ criterion for biological interaction only holds when the factors are monotonic, i.e. never preventive of the outcome (81, 83). Also, as the definition of biological interaction involves causal factors, certain criteria for causality must hold, such as no unmeasured confounding between the exposure and outcome. There are no universal criteria to determine causality although there are many suggestions (82).

1.2.2 Statistical interaction

Another definition of interaction is statistical interaction which refers to departure from the underlying form of a statistical model, i.e. the need to include an interaction term (product) to improve the fit of a statistical model to the data (84). In epidemiology, this type of interaction is also called effect modification, i.e. the effect of one variable changes over values of some other variable (82). The nature and interpretation of statistical interaction depends on the statistical model used and therefore, there is need to specify the effect measure or scale when estimating statistical interaction. In a statistical model where the relationship between the effects of the independent variables is additive, as with linear regression, the interaction term measures departure from additivity of effects (i.e. additive interaction). When analysing risk of disease in case-control studies, the common statistical method is logistic regression, which involves a logarithmic transformation from the original scale of the odds (reflecting probability) of an outcome. Therefore, the interaction term in logistic regression measures departure from multiplicativity of effects (i.e. multiplicative interaction).

In logistic regression with an interaction term between two exposures X_1 and X_2 , the probability of the outcome D is given by:

$$\log \text{odds}(D=1) = \text{logit } P(D=1) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 \quad (3)$$

In the expression, the coefficient β_3 describes how the effect of X_1 differs when $X_2=0$ and $X_2=1$, and vice versa. If $\beta_3=0$, it is equivalent to β_3 exponentiated ($e^{\beta_3}=1$), which is called the interaction odds ratio (OR)

expressing the ratio $OR_{11}/(OR_{10}OR_{01})$. This ratio measures how much the combined effect of the two exposures (OR_{11}) exceeds the product of the separate effects of the two exposures ($OR_{10}OR_{01}$). If the interaction odds ratio equals 1, there is no interaction on the multiplicative scale for odds ratios (80). Thus, a deviation from additivity on the logit scale is translated into a deviation from multiplicativity on the risk scale. The same applies for interaction for any measure of RR so that

$$RR_{11}/(RR_{10}RR_{01}) = 1 \quad (4)$$

in the absence of interaction on the multiplicative (or RR) scale.

An interaction term can be defined for different types of exposure variables such as dichotomous, categorical or continuous. For a dichotomous exposure X_1 in expression (3), $X_1=1$ represents the presence of the risk exposure (e.g. obesity). If X_1 is a continuous exposure e.g. BMI, $X_1=1$ typically equals one unit increase in BMI. If instead a categorical exposure is studied, e.g. level of physical activity, a reference category must be chosen, for instance the lowest risk category, and interaction for all other categories may be tested one by one by comparing with the reference category.

1.2.3 Notes on interaction analysis

If both exposures under study have an effect on the outcome, then there must be effect modification on some scale. Depending on the relation between the exposure effects on e.g. disease risk, interaction may be present on the additive (RD) scale but absent on the multiplicative (RR) scale, or vice versa. Likewise, interaction may be present on both scales, but never absent on both scales (80).

In epidemiological literature, analyses of interaction have often been performed by comparing relative risks of one exposure across strata of another exposure, including a statistical test of multiplicative interaction since a commonly used statistical model is logistic regression (85). However, such analyses do not regard that the baseline risk may vary across the strata, and so the absolute increase in risk from the exposures cannot be determined (86). In fact, such presentations of interaction results may even lead to misleading conclusions if the exposure is associated with a lower RR of the outcome in one stratum but at the same time carries a larger increase in absolute risk in the same stratum. An example of this is given in Table 1. Here, the RR of hypertension from being obese is higher in the high social class ($RR=5.4$) than in the low social class ($RR=3.1$), indicating interaction between obesity and social class on the RR scale so that high social class

potentiates the effect of obesity on hypertension. However, the absolute increase in hypertension risk from being obese is slightly lower in the high social class (RD=4.4) than in the low social class (RD=4.8), indicating weak interaction on the RD scale in the opposite direction, i.e. low social class slightly potentiates the effect of obesity on hypertension. The correct overall interpretation is that the low social class group has a higher baseline risk for hypertension, and the relative effect of adding another risk factor (obesity) is smaller than in the high social class, while the absolute effect of adding the risk factor obesity is at least similar or slightly larger in the low social class. This conclusion can only be made if both the separate and combined effects of both risk factors compared to a common reference of doubly unexposed are presented.

Table 1. Relative risk for hypertension from separate and combined exposure to low social class and obesity compared to common reference group. Table modified from Hallqvist et al 1996 (86).

Social class	Non-obese	Obese	RR	RD
High	1 (Ref)	5.4	5.4	4.4
Low	2.3	7.1	3.1	4.8

RR=relative risk, RD=risk difference

It has been argued that assessing additive interaction is important from a public health perspective, if there is need to target intervention for subgroups of the population when resources are limited (80, 87). Heterogeneity in risk differences from a certain exposure between subgroups gives the information needed for such intervention purposes, without the need to know the baseline risk. For this reason and also because of the scale-dependent interpretation of statistical interaction, it has been recommended to present statistical interaction analyses on both the additive and multiplicative scales when analyzing disease risk (87). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines recommend to present both the separate and joint effects of the two risk factors compared to a common reference group that is unexposed to both risk factors, which gives sufficient information to calculate interaction on both scales (88).

2 AIMS

The overall aim of this thesis was to investigate interactions between genetic and lifestyle factors on CHD risk, as effect modification both on the relative risk scale and the risk difference scale. The specific aims of the constituent studies of this thesis were:

I. to study interaction between *APOE* variation and smoking, physical activity and overweight, respectively, on CHD risk and LDL cholesterol

II. to study interaction between *FTO* variation and physical activity on CHD risk and BMI

III. to study interaction between *FTO* variation and dietary intake of macronutrients on CHD risk and BMI

IV. to study interaction between *FTO* and *GHRL* variation on BMI

3 METHODS

3.1 Study sample

The papers in this thesis were based on data from two independent Swedish population-based studies including cases with CHD; SHEEP (Stockholm Heart Epidemiology Program and INTERGENE (INTERplay between GENetic susceptibility and environmental factors for the risk of chronic diseases in West Sweden), both described briefly below. An overview of the number of subjects included in each paper is given in Figure 2.

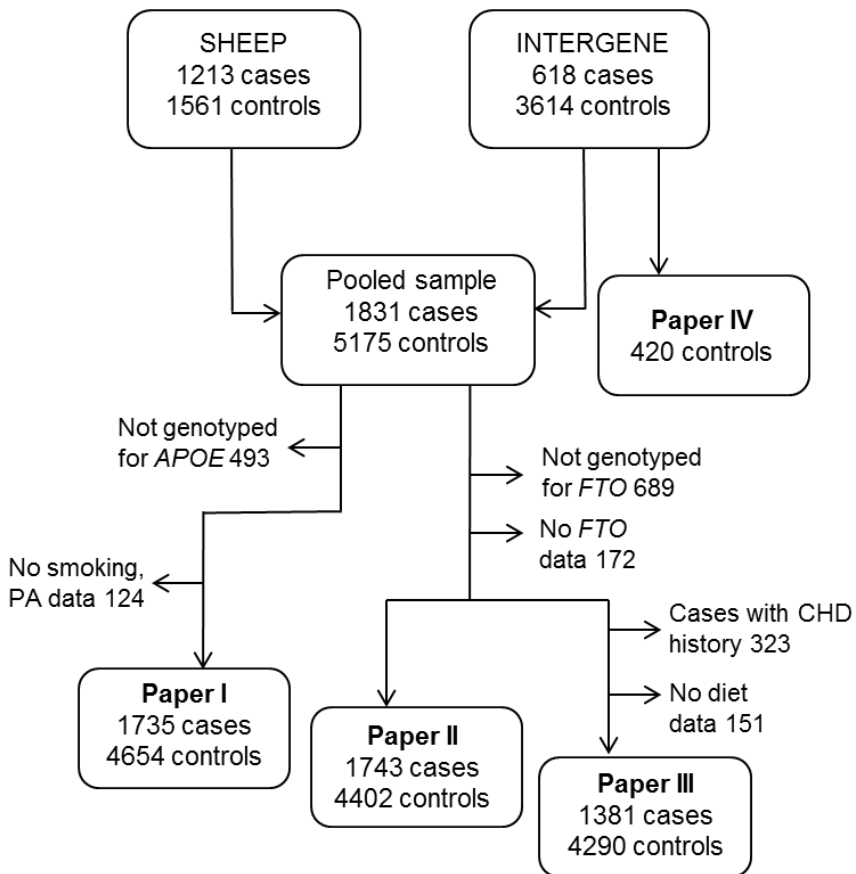


Figure 2. Flowchart of number of subjects included in each paper

3.1.1 The SHEEP study

The Stockholm Heart Epidemiology Program (SHEEP) is a population-based case-control study where the study base consisted of Swedish citizens living in the Stockholm county between 1992 and 1994. The study has been described in detail previously (89). As cases, all patients aged 45-70 years with first-time acute MI were identified from the study base. MI diagnosis criteria specified by the Swedish Association of Cardiologists in 1991 were followed and at least two of the following criteria had to be met: (1) symptoms of acute MI, (2) changes in blood levels of enzymes creatine kinase and lactate dehydrogenase, (3) specified ECG changes, and (4) autopsy findings. Patients who survived at least 28 days after the diagnosis were identified as non-fatal cases.

Control subjects without previous clinically diagnosed MI and matched for sex, age (within a 5-year interval) and residential area were randomly sampled from the study base. Controls were sampled continually during the study, within 2 days of each case occurrence, so called density sampling.

The participation rates were 83% for non-fatal cases and 73% for controls (90). A total of 1643 non-fatal cases and 2339 controls to non-fatal cases were initially included. Patients with re-infarction before study assessments were excluded, leaving 1213 non-fatal cases with exposure data. Of control subjects, 1561 had exposure data (91).

3.1.2 The INTERGENE study

INTERGENE has a case-cohort design with a study base of inhabitants aged 25-74 years in the greater Gothenburg region. As cases in the study, patients surviving acute CHD and discharged from hospital with a diagnosis of CHD (MI, unstable angina pectoris or chronic CHD) were identified during the study period from April 2001 to April 2004. Initially, 818 patients were identified and invited to the study. Of these, 664 patients accepted to participate (participation rate 81%), of which 623 patients were eligible for the study and had a validated CHD diagnosis, and finally 618 had blood samples for DNA analysis and constitute the cases in the study. Of these, 295 had a first-time episode of CHD: 192 with MI, 79 with unstable angina pectoris, and 24 with chronic angina and a positive angiogram; the remaining 323 cases had a previous history of CHD (43 with MI, 91 with unstable angina pectoris and 189 with chronic CHD) and had sought emergency care for cardiac symptoms.

A cohort consisted of subjects randomly sampled from the study base on April 1, 2001 and aged 25-74 years at inclusion. The cohort served as controls, which were sampled at beginning of follow-up of cases from the population at risk of CHD. Initially, 8820 subjects were selected, of which 194 were not eligible for the study leaving 8626 subjects who were invited to the study. Of these, 3614 subjects participated in the study, corresponding to a participation rate of 42%. These subjects were examined during the study period from April 2001 to April 2004.

3.1.3 Pooling of SHEEP and INTERGENE

To obtain a larger sample and increase statistical power for interaction analyses, the two studies were pooled, which requires a reasonable degree of homogeneity between the studies. The studies are similar with respect to design and outcome, and both include control subjects of similar age in larger city-areas in Sweden. The collection of exposure data was done similarly, and the definitions of many variables agree well. Description of how exposure variables were defined in the pooled dataset is given below. In papers II-III, analyses were also made separately in SHEEP and INTERGENE to investigate consistency between the studies.

3.2 Exposure assessment and definitions

3.2.1 Data collection and clinical examination

All participants underwent clinical examinations including measurement of height to the nearest 1 cm, weight to the nearest 0.1 kg, and measurement of waist and hip circumference to the nearest 1 cm while wearing light clothing and no shoes. Blood samples for standard laboratory tests and genotyping were collected after a 4-hour fast. Serum cholesterol and TG concentrations were determined using enzymatic assays. LDL-C levels were estimated according to the Friedewald formula. Lifestyle data including smoking habits, physical activity and dietary habits were collected in self-administered questionnaires. Missing information was checked and solicited by the study personnel.

The clinical examination of non-fatal cases in SHEEP was done about 3 months after the CHD onset (and inclusion in the study), to allow that the cases regained a metabolically stable state. In INTERGENE, a majority of the cases were examined clinically and given questionnaires within 5 months after hospitalization for CHD (and inclusion in the study).

3.2.2 Smoking

Smoking was categorised either as *ever* smoking (including current or past regular smoking) or *never* smoking (neither past nor current regular smoking); or in more detail as *current* smoking (including those who had quit regular smoking <1 year before recording), *past* smoking (including those who had quit regular smoking ≥ 1 year before recording) and *never* smoking (as above).

The ever/never smoking variable was used in Paper I in the primary analyses, but in sensitivity analyses current smoking was compared to never smoking. In papers II-III the current, past, never smoking variable was used, and in paper IV a current/non-smoking variable was used, where non-smoking included both past and never smokers.

3.2.3 Physical activity

Both the SHEEP and INTERGENE questionnaires included questions on leisure-time, occupational and household physical activity. In all papers, leisure time physical activity data were used. There were 4 categories in both studies in response to the question “*How much have you exercised or been physically active during leisure time?*” (Table 2). The SHEEP question referred to average weekly activity during different age intervals (15-24, 25-34, 35-44, 45-54, 55-64 and 65-69). The data used in this thesis were based on the latest reported age interval, which could include up to 10 years preceding the age at reporting. The INTERGENE question referred to the average weekly activity during the year before reporting.

The INTERGENE physical activity questions are based on a validated questionnaire (92), modified from an original version developed by Saltin and Grimby (93), and widely used in epidemiological studies. The SHEEP questions were developed for the study.

Table 2. Leisure time physical activity categories in SHEEP and INTERGENE, and frequencies in control subjects

Level	SHEEP	N (%) controls	INTERGENE	N (%) controls
1	<i>very little activity</i>	117 (8%)	<i>mainly sitting e.g. reading, watching TV, computer</i>	295 (10%)
2	<i>occasional walks, including to/from work</i>	502 (35%)	<i>moderate exercise at least 4 hrs/week e.g. walking, cycling (including to/from work), gardening</i>	1835 (62%)
3	<i>at least 30 min exercise now and then (involving breathlessness)</i>	285 (20%)	<i>regular exercise 2-3 hrs/week e.g. running, swimming, tennis</i>	748 (26%)
4	<i>regular exercise (at least once/week)</i>	517 (36%)	<i>hard training, including competitive sports, several times/week</i>	60 (2%)

In paper I, a dichotomous variable was used with physically *inactive* (levels 1+2 from SHEEP and level 1 from INTERGENE) and *active* (all other levels combined).

In paper II, a categorical variable was used: *low* (level 1 from both SHEEP and INTERGENE), *medium* (level 2 from INTERGENE and 2+3 from SHEEP), and *high* (level 4 from SHEEP and 3+4 from INTERGENE). A dichotomous variable was also used for sensitivity analyses, where the *inactive* group corresponded to the low level and the *active* group corresponded to the medium/high levels combined.

In papers III and IV, a categorical variable for low, medium, high physical activity was used.

3.2.4 Dietary data

Habitual diet during the past 12 month-period was assessed in semi-quantitative food frequency questionnaires (FFQ) that were originally based on an FFQ developed and validated in US by Willett (94, 95), which was translated into Swedish conditions and further developed and validated at Karolinska Institutet (96).

The SHEEP FFQ included 88 food items, while the INTERGENE FFQ was somewhat extended and included 92 food items. For commonly consumed foods, such as milk, bread, cheese and fat on sandwiches, open questions about number of servings per day or week were used. For other food items, participants were asked to estimate their average intake during the past year by choosing between frequencies ranging from “0 times/month” to “3 or more times/day”. The participants were also asked if they had made changes in their dietary habits during the last five years (yes/no).

All frequencies were converted to times/day and multiplied with age and sex-dependent standard portion sizes/servings to estimate the average daily intakes of energy (in kcal) and nutrients (in gram) using the nutrient composition data from the 1997 Swedish National Food Administration database. The dietary variables used in the current study were total intake of energy including alcohol (kcal/day), total intake of non-alcohol energy (kcal/day), total intake (g/day) and percentage of non-alcohol energy (E%) from macronutrients: fat, saturated fatty acids (SFA), mono-unsaturated fatty acids (MUFA), polyunsaturated fatty acids (PUFA), carbohydrate, sucrose, protein and ethanol (g/day). Fibres were not included in carbohydrates. Macronutrient intakes expressed as E% were dichotomized into low and high intake based on the median E% in control subjects.

To judge the validity of the reported total energy intakes, the ratio between the mean food intake level (FIL) and physical activity level (PAL) was calculated. FIL is defined as total energy intake (including alcohol) divided by basal metabolic rate (BMR). BMR for each individual was predicted from age, sex, weight and height according to the Mifflin equation (97). PAL is the total energy expenditure divided by BMR. Individual PAL values were assigned and based on self-reported leisure time physical activity (3 levels) and occupational activity (6 levels), resulting in 18 levels, according to WHO reference values (98). The PAL values ranged from 1.55 for individuals reporting both a sedentary leisure time and the lowest occupational activity level, to 2.3 for those reporting both a vigorous leisure time physical activity and a very physically demanding occupation.

3.2.5 Genetic data

Genotyping of *APOE* SNPs rs429358 and rs7412 was performed by solid-phase mini-sequencing in INTERGENE (99), and by the LightCycler-*APOE* mutation Detection Kit (Roche) in SHEEP.

Genotyping at *FTO* SNP rs9939609 was performed in 4080 subjects from INTERGENE with a success rate of 99% and in 2713 subjects from SHEEP with a success rate of 95% using a Sequenom MassARRAY platform (Sequenom San Diego, CA, USA). The allele frequencies were in Hardy-Weinberg equilibrium in both genotyping assays.

Genotyping of *GHRL* used TaqMan Pre-Designed SNP Genotyping Assays® as previously described (100).

Deviation from Hardy-Weinberg equilibrium for allele frequencies was tested by chi-square.

For *APOE*, two categorical variables were used, one with 6 categories for all genotypes $\epsilon 2\epsilon 2$, $\epsilon 2\epsilon 3$, $\epsilon 2\epsilon 4$, $\epsilon 3\epsilon 3$, $\epsilon 3\epsilon 4$, $\epsilon 4\epsilon 4$, and another with 3 categories: $\epsilon 2$ carriers ($\epsilon 2+$) including genotypes $\epsilon 2\epsilon 2$, $\epsilon 2\epsilon 3$ and $\epsilon 2\epsilon 4$, $\epsilon 4$ carriers ($\epsilon 4+$) including genotypes $\epsilon 3\epsilon 4$ and $\epsilon 4\epsilon 4$, and $\epsilon 3\epsilon 3$ homozygotes ($\epsilon 3$).

For *FTO* SNP rs9939609 T/A, where A is the risk allele, a categorical variable for TT, TA, AA was used to study the genotype-specific associations. Also, both additive (0=TT, 1=TA, 2=AA) and dominant (TT=0, TA/AA=1) genetic models were used.

For *GHRL* SNPs rs35680 A/G, rs26802 T/G, rs42451 G/A, rs696217 G/T and rs4684677 T/A, additive genetic models were defined (0=no minor allele, 1=one minor allele, 2=two minor alleles). In addition, recessive (0=no or 1 minor allele, 1=2 minor alleles) and dominant (0=no minor allele, 1=1-2 minor alleles) genetic models were defined in sensitivity analyses.

3.2.6 Other variables

BMI was defined as weight divided by height squared and is given in units of kg/m^2 . Overweight (including obesity) and obesity were defined as $\text{BMI} \geq 25 \text{ kg/m}^2$ and $\geq 30 \text{ kg/m}^2$, respectively.

The waist-hip-ratio (WHR) was defined as the ratio between the waist and hip circumference.

A dichotomous variable (yes/no) was constructed for lipid-lowering medication based on self-report, where 'yes' corresponded to regular use, at least once weekly, of lipid-lowering medication.

3.3 Statistical analyses

Descriptives of data were given as mean values and standard deviations (SD) for continuous variables, and proportions for categorical variables. Skewed data were log-transformed. Mean values between groups were compared using ANOVA, and differences in proportions were tested using the chi-square test. The correlation between E% of different macronutrients was estimated by the Pearson correlation coefficient.

Association between exposures and odds of CHD were estimated by multiple logistic regression. Effect estimates are given as odds ratios (OR) with 95% confidence intervals (95% CI). With certain sampling methods of controls, such as longitudinally from the study base (person-time sampling), as in SHEEP, or from the population at risk at the beginning of follow-up, as in INTERGENE, the OR is an estimate of the incidence rate ratio or the risk ratio (101).

Association between exposures and continuous outcome variables LDL-C and BMI were estimated by multiple linear regression.

In all regression models, basic adjustment was made for age, sex and study (INTERGENE or SHEEP) in the pooled sample, and for age and sex in the study-specific samples. Additional adjustment was made for important co-variables and potential confounders.

In sensitivity analyses in paper II, cases with history of CHD from INTERGENE were excluded to evaluate the influence of previous CHD diagnosis on the associations.

In additional exploration (not in paper III) of which food items that had a positive association with SFA E%, a step-wise linear regression was performed in the pooled sample, separately in cases and controls. A selection of food items from the FFQ assumed to be associated with SFA intake were made: whole milk, whole yoghurt/sour milk, crème fraiche, full-fat cheese, butter, substitute butter, pancakes, minced meat, sausage, meat (beef/calf/lamb), eggs/omelet, fried potatoes, buns, cake, chocolate, ice-cream and creamy sauce. Rapeseed oil was also included as control. In the first step, bivariable regression was made and food items with a positive association with SFA E% (p -value <0.25) were selected and included in a multivariable model. Then, food items that had significant ($p<0.05$) positive association with SFA E% in the multivariable model were identified.

An additional analysis was also performed of predictors of high E% fat in controls using multivariable logistic regression with high E% fat as dependent variable and the independent variables age, sex, physical activity, *FTO* genotype, BMI, smoking, change in dietary habits, alcohol intake, sucrose E%, protein E%, high serum cholesterol and use of lipid-lowering medication. Factors with significant ($p < 0.05$) association with fat E% were identified.

3.3.1 Interaction analyses

Analyses of interaction between genetic and lifestyle factors on CHD risk was performed both as effect modification on the RR scale (multiplicative interaction) by introducing an interaction term in the logistic regression model, and as effect modification on the RD scale (additive interaction) by calculation of RERI (defined in Section 1.2.1) with 95% CI estimates (87).

For RERI calculation, each exposure was turned into a risk factor, by using the category with lowest risk as reference category. In order to estimate CHD risk from both separate and combined exposures to genetic and lifestyle risk factor, categories were defined by presence of either risk factor alone or combined exposure to both risk factors, and ORs were estimated for these categories compared to a reference group unexposed to both risk factors.

Analyses of interaction in relation to the continuous variables LDL-C and BMI was made by introducing the corresponding interaction term in the linear regression model.

For interaction models with two binary or continuous factors, there is only one interaction term. For interaction between a binary variable (or continuous variable) and a categorical variable with r categories ($r > 2$), represented by $(r-1)$ binary indicator variables in the regression model, there are also $(r-1)$ interaction terms. P-values for each interaction term individually were obtained from the model. Significance for the overall interaction (i.e. comparison of models with and without overall interaction term) was also obtained from the model.

In exploration of the main associations between macronutrients and CHD, we used substitution models where E% of all macronutrients except the substituted macronutrient were included (e.g. SFA substituting carbohydrate), adjusted for total non-alcohol energy intake. Since all macronutrient E% add up to 100%, the effects of macronutrients included in the model can be interpreted as effects of these when replacing the excluded macronutrient (102).

All analyses were conducted with SAS (Statistical Analysis System, Inc) software (Version 9.2 for Windows), except the meta-analysis included in Paper II, which was conducted with STATA software. Point estimates were supplemented with 95% confidence intervals (CI) where relevant, and a p-value <0.05 (two-sided test) was considered significant when assessing statistical significance.

3.3.2 Power considerations

The required sample size for detecting gene-environment interaction is often larger than for detecting main effects. Factors that affect the required sample size are the strength of the interaction, exposure and allele frequencies, and study design (58). In this thesis, power calculations for the interaction analyses could not be made with precision *a priori*, because the analyses were mainly exploratory as the size of interaction effects was unknown. However, below are examples of power calculation using the Quanto software (103), with different assumptions about the size of interaction (i.e. interaction OR), given allele and exposure frequencies, and main effects of genetic and lifestyle exposures on CHD as observed in the pooled sample.

In Example A in Table 3, for *APOE*-smoking interaction, the prevalence of ever smoking was 0.50 and frequency of $\epsilon 2$ carriers=0.08 in the population controls, assuming dominant genetic model, main genetic effect of 1.35 (CHD OR in non-carriers versus carriers of $\epsilon 2$) and main effect of smoking of 2.1 (OR from ever versus never smoking), the pooled sample size with case:control ratio of 1:2.7 (1700 cases), and significance level=0.05.

In Example B in Table 3, for *FTO*-physical activity interaction, the prevalence of physical inactivity was 0.10 and *FTO* minor A allele frequency 0.40 in the population controls, assuming dominant genetic model, main genetic effect of 1.25 (OR from genotype TA/AA versus TT), main effect of physical inactivity of 1.5, and otherwise the same parameters as in Example A.

Table 3. Statistical power for detection of gene-lifestyle interaction of different size, with two examples

Example A)		Example B)	
Interaction OR	Power	Interaction OR	Power
1.2	30%	1.2	17%
1.5	76%	1.5	62%
1.6	87%	1.6	75%
1.7	94%	1.7	85%
2.0	99%	2.0	99%

Case:control ratio 1:2.7 and 1700 cases

Thus, the studies were adequately powered for detection of an interaction effect of roughly 1.5 and above, which can be regarded as a medium-size interaction effect, and arguably, these may be of most interest (58).

Regarding paper III, diet effect sizes are typically modest, so the study of gene-diet interactions in epidemiology is a methodological challenge. Therefore, the interaction analysis in paper III should be regarded as exploratory.

In paper IV, the gene-gene interaction analysis should also be considered as exploratory given that the sample size was small, as the subsampling with subjects genotyped for the *GHRL* SNPs was designed for another study.

4 RESULTS

A summary of the results in paper I-IV is given below, and more detailed descriptions of the results are given in each paper. Any additional data or analyses that were not included in the papers are shown below, and referred to as additional analyses.

4.1 Characteristics of the study population

The CHD cases had higher prevalence of smoking, overweight/obesity and low leisure time physically activity than population controls (Table 4). Cases were also older and included more women (although in SHEEP these factors were matched in controls).

Table 4. Characteristics by case status in the pooled sample

	CHD Cases N=1831	Population controls N=5175
Women, %	29%	47% ***
Age, years mean (range)	60 (31-74)	54 (24-77)***
BMI, kg/m ² mean (SD)	27.2 (4.1)	26.0 (4.0)***
Overweight (BMI ≥25)	69%	56% ***
Obese (BMI ≥30)	21%	14% ***
LDL cholesterol, mean (SD)	3.7 (1.3)	3.5 (1.0) ***
HDL cholesterol, mean (SD)	1.2 (0.4)	1.5 (0.5) ***
Smoking, %		***
Current	37%	21%
Former	37%	32%
Never	26%	47%
Physical activity, %		***
Low	22%	13%
Medium	57%	57%
High	22%	30%
<i>APOE</i> genotype		***
ε2+ (ε2ε2, ε2ε3, ε2ε4)	11%	15%
ε3ε3	60%	58%
ε4+ (ε3ε4, ε4ε4)	29%	27%
<i>FTO</i> genotype		***
TT	32%	37%
TA	49%	47%
AA	19%	16%

*Test for difference between cases and controls, *** p<0.001.*

4.2 Main effects of lifestyle factors on CHD

Smoking, physical inactivity and overweight/obesity were associated with increased risk of CHD, as expected (Table 5). Regarding dietary macronutrients, high E% of fat or SFA were associated with lower CHD risk and high E% protein was associated with increased CHD risk, while the other macronutrients (carbohydrate, sucrose and PUFA) were not associated with CHD.

Table 5. CHD risk associated with genetic and lifestyle factors in the pooled sample, OR (95% CI)

Risk factor	Contrast	OR	95% CI
Smoking ^a	Ever vs never	2.14	1.88-2.45
	Current vs never	2.56	2.19-2.99
	Former vs never	1.78	1.53-2.06
Physical activity ^b	Inactive vs active	1.49	1.30-1.71
	Low vs medium activity	1.28	1.05-1.55
	High vs medium activity	0.69	0.59-0.80
Overweight/obesity ^c	Overweight, including obesity vs normal-weight	1.68	1.47-1.91
	Obesity vs normal-weight	2.09	1.75-2.49
Macronutrients ^{d,e}	High vs low E% fat	0.70	0.61-0.81
	High vs low E% SFA	0.72	0.62-0.83
	High vs low E% PUFA	0.95	0.83-1.10
	High vs low E% carbohydrate	1.12	0.97-1.29
	High vs low E% sucrose	0.94	0.82-1.08
	High vs low E% protein	1.31	1.14-1.51

^aAdjusted for age, sex, study, BMI and physical activity.

^bAdjusted for age, sex, study, BMI and smoking.

^cAdjusted for age, sex, study, smoking, and physical activity.

^dAdjusted for age, sex, study, smoking, physical activity and total energy intake.

^eHigh=above median, low=below median E% in population controls.

4.3 APOE and lifestyle factors

APOE genotype was associated with both LDL-C and CHD risk, so that carriers of the $\epsilon 2$ allele, compared to the common genotype $\epsilon 3\epsilon 3$, had lower LDL-C levels and lower CHD risk, with OR 0.63 (95% CI 0.52-0.77), and $\epsilon 4$ carriers had higher LDL-C levels but no further increase in CHD risk, with OR 1.03 (95% CI 0.90-1.18) (see paper I, Table 2).

An interaction between *APOE* and smoking was found on the RR scale, so that the smoking-related CHD risk was lower in $\epsilon 2$ carriers, with OR 1.35 (95% CI 0.92-1.97), than in $\epsilon 3\epsilon 3$ subjects with OR 2.17 (95% CI 1.82-2.59) and $\epsilon 4$ carriers with OR 2.43 (95% CI 1.88-3.14) (Figure 3). The overall *APOE*-smoking interaction was significant ($p=0.04$), when adjusted for age, sex, study, physical activity, BMI and LDL-C. Also, the interaction was stronger in women ($p=0.006$) and with partly differing pattern. Women carrying $\epsilon 4$ appeared to have higher smoking-related risk, with OR 3.69 (95% CI 2.33-5.83), than women with the common $\epsilon 3\epsilon 3$ genotype, while among men the risk in these two genotypes was more similar (Figure 3).

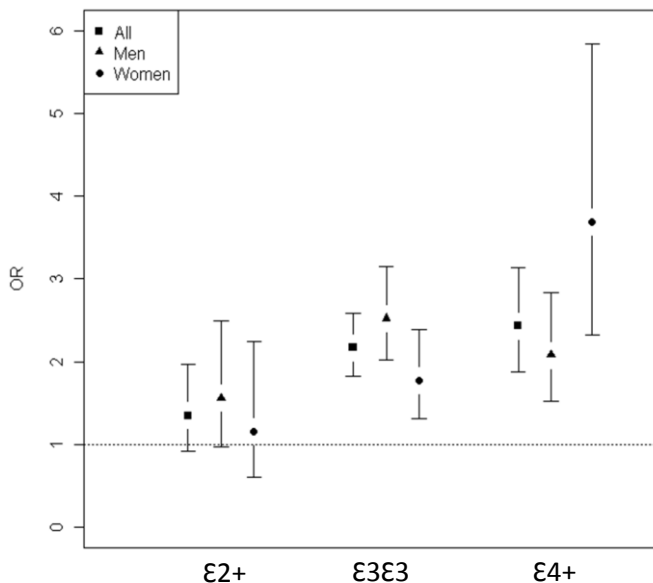


Figure 3. Smoking-related risk of CHD (OR, 95% CI) by *APOE* genotype, adjusted for age, sex, study, physical activity, BMI and LDL-C. $P=0.04$ for interaction. From Paper I, Table 3.

$\epsilon 2+$ includes genotypes $\epsilon 2\epsilon 2$, $\epsilon 2\epsilon 3$, $\epsilon 2\epsilon 4$

$\epsilon 4+$ includes genotypes $\epsilon 3\epsilon 4$, $\epsilon 4\epsilon 4$

In additional analyses (not in paper I), interaction on the additive scale was evaluated by calculation of RERI. The separate and combined effects of *APOE* genotype and smoking on CHD were compared to a common reference category of lowest CHD risk ($\epsilon 2$ carriers and never smoking), see Table 6. The RERI for both $\epsilon 3\epsilon 3$ and $\epsilon 4+$ and smoking were significantly above 0, i.e. indicating more than additive effects.

Table 6. Risk of CHD, OR (95% CI), from separate and combined exposures to APOE genotype and smoking, and additive interaction

<i>APOE</i> genotype	Smoking	
	Never	Ever
$\epsilon 2+$	1 Reference	1.45 (1.00-2.10)
$\epsilon 3\epsilon 3$	1.16 (0.83-1.63)	2.62 (1.90-3.60)
$\epsilon 4+$	1.15 (0.80-1.67)	2.73 (1.95-3.82)

Additive interaction (reference $\epsilon 2+$ and never smoking)

RERI (95% CI)	$\epsilon 3\epsilon 3$ and smoking 1.02 (0.59-1.44)
	$\epsilon 4+$ and smoking 1.15 (0.61-1.70)

Adjusted for age, sex, study, physical activity and BMI.

$\epsilon 2+$ includes genotypes $\epsilon 2\epsilon 2$, $\epsilon 2\epsilon 3$, $\epsilon 2\epsilon 4$

$\epsilon 4+$ includes genotypes $\epsilon 3\epsilon 4$, $\epsilon 4\epsilon 4$

A suggested interaction on the RR scale was found for *APOE* and physical inactivity, where $\epsilon 2$ carriers had no increased CHD risk from physical inactivity with OR 1.04 (95% CI 0.69-1.56), while $\epsilon 4$ carriers had more pronounced CHD risk from physical inactivity with OR 1.77 (95% CI 1.36-2.31) than subjects with $\epsilon 3\epsilon 3$ genotype who had OR 1.36 (95% CI 1.13-1.64), and a similar pattern in both men and women (Figure 4). The suggested effect modification was borderline statistical significant ($p=0.07$), when adjusted for age, sex, study, smoking, BMI and LDL-C.

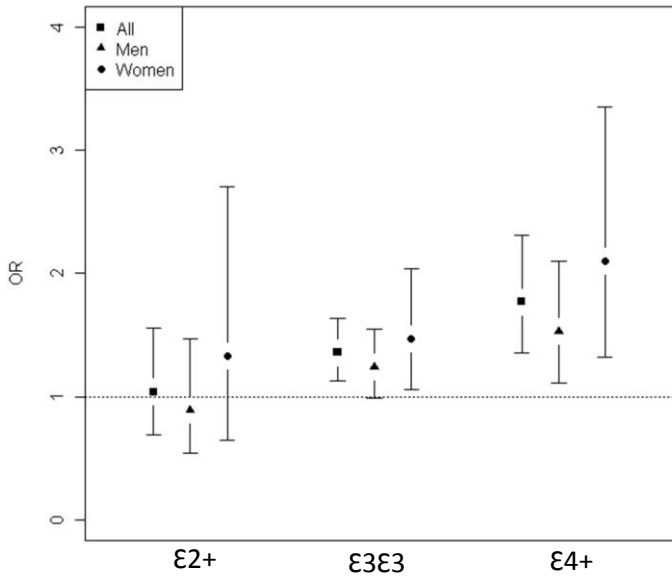


Figure 4. CHD risk (OR, 95% CI) from physical inactivity by APOE genotype, adjusted for age, sex, study, smoking, BMI and LDL-C. $P=0.07$ for interaction. From Paper I, Table 3.

ε2+ includes genotypes ε2ε2, ε2ε3, ε2ε4

ε4+ includes genotypes ε3ε4, ε4ε4

RERI was also calculated for the APOE-physical activity interaction (not in paper I). When contrasting to a common reference category of ε2+ and physically active, the RERI for ε3ε3 and physical inactivity was 0.44 (95% CI -0.068, 0.94), and for ε4+ and physical inactivity it was 1.03 (95% CI 0.35, 1.71), suggesting a synergistic interaction between APOE risk genotype and physical inactivity (Table 7).

Table 7. Risk of CHD, OR (95% CI), from separate and combined exposures to APOE genotype and physical inactivity, and additive interaction

APOE genotype	Physical activity	
	Active	Inactive
ε2+	1 Reference	1.09 (0.73-1.61)
ε3ε3	1.50 (1.20-1.87)	2.01 (1.56-2.58)
ε4+	1.42 (1.11-1.81)	2.53 (1.88-3.40)

Interaction on additive scale, reference ε2+, active

RERI (95% CI)	ε3ε3 and inactive	0.44 (-0.068, 0.94)
	ε4+ and inactive	1.03 (0.35, 1.71)

Adjusted for age, sex, study, smoking and BMI.

ε2+ includes genotypes ε2ε2, ε2ε3, ε2ε4

ε4+ includes genotypes ε3ε4, ε4ε4

No interaction on the RR scale was found for APOE and overweight on CHD risk, or between APOE and any of the studied lifestyle factors in relation to LDL-C levels (paper I).

4.4 FTO and physical activity

Subjects carrying the *FTO* SNP rs9939609 minor allele A had an increased risk of CHD with OR 1.25 (95% CI 1.10-1.42) in the pooled sample, adjusted for age, sex and study. This association was only slightly attenuated when additionally adjusting for BMI, with OR 1.20 (95% CI 1.06-1.37). The *FTO* association with CHD was similar in SHEEP and INTERGENE analysed separately, and also in men (OR 1.21, 95% CI 1.04-1.40) and women (OR 1.36, 95% CI 1.07-1.71). The *FTO* A allele was also associated with increased BMI, with per allele effect of +0.35 (95% CI 0.19-0.52) kg/m², which was similar in both SHEEP and INTERGENE (paper II).

No significant interaction between *FTO* (TA/AA versus TT genotype) and leisure time physical activity (low, medium or high) on CHD risk was found either on the RR or RD scale (Table 8, C). However, a trend of increased CHD risk attributable to the *FTO* A allele was observed with increasing physical activity level, although the confidence intervals were wide and largely overlapping (Table 8, B).

The RERI for *FTO* TA/AA and low or medium physical activity compared to TT and high activity were close to 0, indicating the absence of interaction on the RD scale and thus additive effects of risk genotype and lower physical activity on CHD risk (Table 8, C). Compared to low-risk subjects with both TT genotype and high physical activity level, those with combined exposure to *FTO* genotype TA/AA and low physical activity had approximately 3-fold increased CHD risk (OR 3.30, 95% CI 2.44-4.46).

Table 8. Risk of CHD, OR (95% CI), by *FTO* genotype and physical activity

A. Separate and combined effects with common reference		
PA level	TT	TA/AA
High	1 (Ref)	1.38 (1.06-1.80)
Medium	1.82 (1.42-2.34)	2.23 (1.76-2.82)
Low	2.98 (2.06-4.31)	3.30 (2.44-4.46)
B. PA stratum-specific effects of <i>FTO</i> (TA/AA versus AA)		
High	1.38 (1.06-1.80)	
Medium	1.22 (1.04-1.44)	
Low	1.11 (0.77-1.60)	
C. Interaction measures (Reference=TT, high PA)		
Relative risk scale ^a	TA/AA x medium PA p=0.4	
	TA/AA x low PA p=0.3	
Risk difference scale, RERI (95% CI)	TA/AA + medium PA 0.03 (-0.5, 0.4)	
	TA/AA + low PA -0.07 (-1.4, 1.1)	

PA=physical activity

^aFrom logistic regression model with interaction term *FTO* (TA/AA versus TT) x PA (low or medium with high PA as reference), adjusted for age, sex and study.

Interaction between *FTO* and physical activity in relation to BMI showed a tendency of stronger *FTO* effect in subjects with lower physical activity level in the pooled sample, but without statistically significant interaction ($p=0.12$), Table 2 in paper II.

4.5 *FTO* and dietary macronutrients

In age- and sex-adjusted comparisons of macronutrient intakes in the pooled sample, cases reported slightly higher intake of total energy (excluding alcohol) and fat, but no difference in fat E%. Cases also had lower intake of PUFA (total and E%) and slightly higher total carbohydrate and protein E% (Table 1, Paper III).

However, these patterns differed somewhat in SHEEP and INTERGENE. In INTERGENE, fat and SFA E% were significantly lower in cases than in controls while PUFA E% was similar in cases and controls. Another difference between the studies was the prevalence of subjects reporting recently changed dietary habits, which was overall considerably higher in INTERGENE, but in both studies more common in cases than in population controls (Appendix Table A1). The subjects reporting diet change had lower fat and SFA E%, but higher PUFA, protein and carbohydrate E%, and had higher proportion of obese than subjects with stable dietary habits (Supplement Table 1, paper III).

In population controls, E% of fat was strongly positively correlated with both SFA and MUFA, but not with PUFA. Fat, SFA and MUFA, respectively, were also strongly negatively correlated with carbohydrate (Table 9).

Table 9. Pearson correlation coefficients between macronutrient E% in population controls

	Fat	SFA	MUFA	PUFA	Carbo- hydrate	Sucrose	Protein
Fat	1	0.92	0.92	0.36	-0.93	-0.38	0.26
SFA		1	0.74	0.01	-0.88	-0.31	0.29
MUFA			1	0.45	-0.83	-0.33	0.16
PUFA				1	-0.30	-0.25	-0.01
Carbohydrate					1	0.54	-0.59
Sucrose						1	-0.58
Protein							1

Additional analyses (not in paper III) showed that food items positively associated with SFA E% in control subjects were full-fat cheese, substitute

butter, sausage, buns, chocolate, creamy sauce and crème fraîche. The pattern was similar in cases with positive association for full-fat cheese, sausage, chocolate, creamy sauce and crème fraîche. Other factors significantly associated with high fat E% among control subjects were younger age, female gender, low leisure time physical activity, lower BMI, smoking, no reported change in dietary habits, higher protein E% and non-use of lipid-lowering medication.

When exploring the main association between macronutrient E% and CHD in substitution models, increased fat substituting either carbohydrate or protein was associated with lower CHD risk. SFA was associated with significantly decreased CHD risk only when substituting carbohydrate, but not when replacing PUFA (Supplement Table 3, Paper III).

In interaction analyses using the dichotomous macronutrient E% variables, no significant interactions were found for any macronutrient and *FTO* in relation to CHD, either on the RR or RD scale. However, association between *FTO* TA/AA genotype and CHD was only observed in strata of low fat intake or low SFA intake, and in strata of high carbohydrate intake or high protein intake. There were positive RERI estimates for *FTO* TA/AA and low fat E%, *FTO* and low SFA E% and *FTO* and high protein E%, but none was statistically significant. When using a continuous macronutrient E% variable, the *FTO*-related CHD risk was stronger with lower SFA E%, with borderline significant ($p=0.05$) effect modification on the RR scale (Table 10).

After excluding subjects reporting changed dietary habits during the last 5 years (leaving 3589 subjects), the interaction on the RR scale was significant for both *FTO*-fat and *FTO*-SFA when using the continuous E% variable (Supplement Table 4, paper III). Similarly, the RERI was somewhat strengthened for *FTO* and low fat (0.35, 95% CI -0.04 , 0.74) and for *FTO* and low SFA (0.33, 95% CI -0.08 , 0.74), but neither statistically significant. The RERI for *FTO* and high protein was weakened.

No interaction was found between *FTO* and macronutrients on BMI (Table 3, paper III).

In summary, observations of weak interactions on the RR scale between *FTO* and fat or SFA on CHD risk were made, suggesting that the *FTO*-attributable risk of CHD was strengthened with lower energy-related intakes of fat/SFA.

Table 10. Risk of CHD by FTO genotype and macronutrient energy percentage (E%), and interaction measures

Macronutrient E%	OR (95% CI) for FTO TA/AA vs TT	P interaction ^a	OR (95% CI) with common reference		RERI ^b (95% CI)
			FTO genotype		
			TT	TA/AA	
Fat					
Low	1.36 (1.11-1.66)	0.44 (0.16)	1.18 (0.92-1.52)	1.60 (1.28-2.01)	0.22 (-0.13, 0.57)
High	1.20 (0.95-1.51)		1.0	1.20 (0.95-1.51)	
SFA					
Low	1.36 (1.10-1.69)	0.45 (0.05)	1.15 (0.89-1.48)	1.57 (1.25-1.96)	0.21 (-0.14, 0.55)
High	1.21 (0.97-1.51)		1.0	1.21 (0.97-1.51)	
PUFA					
Low	1.26 (1.04-1.53)	0.70 (0.60)	1.08 (0.83-1.41)	1.37 (1.07-1.74)	0.02 (-0.38, 0.34)
High	1.31 (1.02-1.68)		1.0	1.31 (1.02-1.68)	
Carbohydrate					
Low	1.23 (0.98-1.55)	0.68 (0.38)	1.0	1.23 (0.98-1.55)	0.08 (-0.24, 0.41)
High	1.32 (1.07-1.61)		0.99 (0.77-1.28)	1.31 (1.04-1.64)	
Sucrose					
Low	1.24 (1.00-1.54)	0.75 (0.53)	1.0	1.24 (1.00-1.54)	0.04 (-0.28, 0.35)
High	1.31 (1.05-1.62)		0.91 (0.71-1.17)	1.19 (0.96-1.48)	
Protein					
Low	1.14 (0.92-1.41)	0.18 (0.49)	1.0	1.14 (0.92-1.41)	0.29 (-0.04, 0.63)
High	1.41 (1.13-1.75)		1.09 (0.85-1.40)	1.53 (1.24-1.89)	

Adjusted for age, sex, study, physical activity, smoking, lipid-lowering medication, non-alcohol energy and alcohol intake (g/day).

^aFrom logistic regression model with interaction term between FTO (TT=0, TA/AA=1) and macronutrient E% (low risk=0, high risk=1). (from an interaction model with continuous macronutrient E%).

^bRERI= Relative excess risk due to interaction, estimated measure of deviation from additive effects.

4.6 FTO and GHRL interaction on BMI

In paper IV, interaction between *FTO* and GHRL on BMI was investigated in a subset (n=420) of the INTERGENE cohort.

There were no statistically significant main effects between any *GHRL* SNP and BMI (Table 11). The *FTO* stratum-specific (TT and TA/AA) associations, showed nearly opposite effects of the *GHRL* rs35680 minor allele G on BMI in the two *FTO* strata, with a negative association with BMI in the *FTO* TT stratum, and positive association with BMI in the *FTO* TA/AA stratum (p for interaction=0.006). There was also a borderline significant interaction between *FTO* and *GHRL* SNP rs42451 (p = 0.06). There was no evidence of interactions with the other *GHRL* SNPs. When adjusting additionally for physical activity, smoking and ethanol consumption, the interaction was still present for *FTO* and *GHRL* rs35680, but weakened for *FTO* and GHRL rs42451 (Table 1, paper IV).

Table 11. *GHRL* SNP (additive model) associations with BMI (kg/m^2 , Beta, 95% CI) in all subjects and by *FTO* genotype stratum (TT and TA/AA)

<i>GHRL</i> SNP	All subjects	<i>FTO</i> genotype strata		<i>P</i> ^a
	n=420	TT n=148	TA/AA n=272	
rs35680 A/G	0.09 (-0.50, 0.69)	-0.93 (-1.92, 0.05)	0.81 (0.05, 1.56)	0.006
rs26802 T/G	0.20 (-0.39, 0.80)	0.70 (-0.27, 1.68)	-0.09 (-0.83, 0.66)	0.2
rs42451 G/A	0.13 (-0.52, 0.77)	0.88 (-0.18, 1.94)	-0.40 (-1.21, 0.42)	0.06
rs696217 G/T	0.31 (-0.56, 1.18)	0.20 (-1.17, 1.57)	0.46 (-0.67, 1.59)	0.8
rs4684677 T/A	-0.71 (-1.74, 0.33)	-0.14 (-2.05, 1.78)	-1.08 (-2.32, 0.15)	0.4

Estimates from linear model assuming additive effect of *GHRL*, adjusted for age and sex.

^a*P* for interaction in a linear model including *FTO*, *GHRL* SNP (assuming additive effect), an interaction term *FTO*×*GHRL*, adjusted for age and sex.

In sensitivity analyses using a recessive genetic model for *GHRL* rs35680, the interaction was strengthened in the same direction (p=0.0009), adjusted for age and sex. Mean BMI by genotype strata of *FTO* and *GHRL* rs35680 are displayed in Figure 5, where subjects homozygous for rs35680 G who were also in the *FTO* risk stratum TA/AA had the highest mean BMI of

28.8 kg/m², compared to subjects homozygous for rs35680 G in the *FTO* TT stratum who had a mean BMI of 24.8 kg/m².

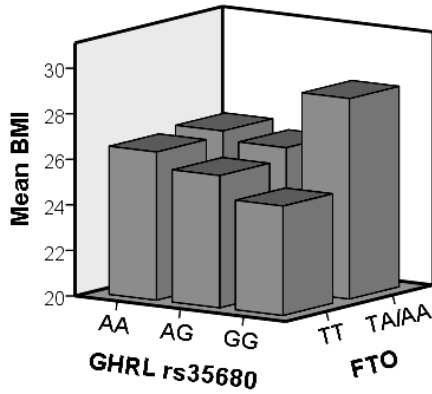


Figure 5. Mean BMI (kg/m²) by GHRL rs35680 and FTO rs9939609 genotype

5 DISCUSSION

5.1 APOE and lifestyle factors

Paper I confirms previous reports of *APOE*-smoking interaction in relation to CHD (104). This study extends knowledge in that the *APOE*-smoking interaction was demonstrated in women also, whereas previous studies have included mainly men (63). Moreover, a somewhat differing interaction pattern was found in men and women, where women carrying the $\epsilon 4$ allele appeared to be at particularly high risk from smoking, while this was not observed in men. This observation might be an important concern for women and deserves further investigation in future studies. An explanation for the suggested gender difference is not obvious, especially as the mechanism of the *APOE*-smoking interaction is unclear. Biological gender differences are known, e.g. in cholesterol turnover (105), which may affect the LDL-C levels influenced by *APOE* genotype. The finding may also be due to chance, which is why replication studies are needed.

An additional finding that differs from most previous *APOE*-smoking interaction studies is that the $\epsilon 2$ allele rather than the $\epsilon 4$ allele was driving the interaction, since the smoking-related CHD risk was similar in the common $\epsilon 3\epsilon 3$ genotype and $\epsilon 4$ carriers (at least in men). In some previous studies, $\epsilon 4$ carriers have been compared to all other genotypes combined (i.e. $\epsilon 2$ and $\epsilon 3$ carriers), so it is possible that a protective effect of the $\epsilon 2$ allele has been interpreted as a risk from the $\epsilon 4$ allele (16). Also, since *APOE* had no association with CHD risk in the absence of the environmental stressor smoking, the CHD risk associated with *APOE* may differ in populations with different smoking prevalence (41). It should be noted in this context that the partly protective effect of the $\epsilon 2$ allele is only of relevance to a minority, as $\epsilon 2$ carriers constitute about 15% of the population, and smoking is still a risk factor also in this group.

A novel finding was the suggested *APOE*-physical inactivity interaction on CHD risk, both on the RR and RD scales. In $\epsilon 2$ carriers, physical inactivity was not associated with CHD risk, whereas in $\epsilon 3\epsilon 3$ genotype and $\epsilon 4$ carriers, physical inactivity was associated with increased CHD risk. There were more than additive effects of the $\epsilon 4$ allele and physical inactivity in relation to the $\epsilon 2$ allele and physical activity on CHD risk, suggesting the presence of biological interaction.

Interestingly, there was no interaction between either *APOE* and smoking, or *APOE* and physical inactivity, on LDL-C levels, and the interactions for CHD were not weakened after adjusting for LDL-C, which suggests that the interactions are explained by some other mechanism than a direct effect on LDL-C. A plausible mechanistic explanation is the gradient of antioxidant status of the apoE2>apoE3>apoE4 isoforms due to the amino acids at positions 112 and 158 differing so that apoE2 has two cysteine, apoE3 has one cysteine, while apoE4 has no cysteine but arginine at both positions instead (41). Cysteine has a thiol (-SH) group acting as a scavenger of oxidative particles, thus with antioxidant effect. Therefore, the apoE2 isoform may reduce the oxidative stress caused by smoking, which is of central importance for atherosclerosis underlying CHD. Physical inactivity is also known to increase oxidative stress (106). The interaction finding thus also lends indirect support to that *APOE*-attributable CHD risk may not only be due to increased LDL-C, but also to oxidative status or other effects (41, 104).

Considering the significant *APOE*-smoking interaction on the additive scale, replicated data on *APOE* variation associated with CHD, and the known causal relation between smoking and risk of CHD, the *APOE*-smoking interaction can with high certainty be regarded as a biological interaction. For similar reasons, although the additive interaction was only borderline significant, the *APOE*-physical inactivity interaction is a strong candidate for biological interaction, although this finding needs to be confirmed in further studies.

5.2 *FTO* and physical activity

The *FTO* minor allele association with increased risk of CHD appears robust as it was observed in both SHEEP and INTERGENE, and in men and women separately, and agrees with previous reports (21-23, 71). The association was largely independent of BMI, also confirming previous observations (23), which suggests that the increased CHD risk is at least partly due to some other effect of *FTO* than increased body/fat mass or obesity-related traits. Apart from the role of *FTO* in hypothalamic control of energy homeostasis and appetite, it may have other effects, and *FTO* is expressed not only in hypothalamic neurons but also widely in neurons throughout the body (35, 107). *FTO* is also suggested to have a role in adipogenesis, lipogenesis and mitochondrial function in skeletal muscles (35).

The combined effects of *FTO* risk genotype and lower physical activity level on CHD risk were approximately additive, indicating the absence of additive

interaction, suggesting independent effects of these factors on CHD according to the biological interaction definition. This finding certainly needs replication, but if reflecting true conditions, it may add some clues to the understanding of the mechanisms connecting *FTO* variants with risk of CHD.

This study found no evidence of *FTO*-physical activity interaction on CHD on the multiplicative scale either, although a trend of increased relative CHD risk attributable to the A allele was observed with increased physical activity. Interestingly, this pattern is in opposite direction to the pattern observed in relation to BMI, where the *FTO* effect was suggested to be weaker in more physically active subjects, which agrees with several previous reports demonstrating that increased physical activity attenuates the *FTO* effect on obesity (70).

There are only few studies on *FTO*-physical activity interaction in relation to CHD. The current findings are in line with one previous study in a cohort of over 22,000 subjects in Sweden, showing that subjects with TT genotype had a more pronounced effect on cardiovascular mortality from decreased physical activity than subjects carrying the A allele (72). In contrast, another prospective study in over 21,000 women in USA found that only physically inactive women had an increased risk of cardiovascular disease associated with an *FTO* risk allele (71). These contrasting results are not easily explained, but one potential explanation is that populations in different geographical regions have different co-variation between physical activity and other lifestyle factors (e.g. diet), which in turn might explain the differing interaction patterns.

Overall, we found no evidence that increased physical activity would counteract the *FTO* association with CHD risk, implying that although physical activity protects from CHD, it does not necessarily protect from the relatively small increase in CHD risk associated with *FTO* gene variation.

5.3 *FTO* and dietary macronutrients

A suggested interaction was observed between *FTO* and fat or SFA intake on CHD on the RR scale, but it was rather weak and significant only in sensitivity analyses after excluding those with unstable dietary habits, and when using a continuous variable for macronutrient E%. Therefore, it is difficult to draw any firm conclusions on *FTO*-macronutrient interaction on CHD risk. It may be that the dichotomous E% variable did not sufficiently capture the variation in macronutrient intakes, as many subjects have an intake around the median, so the contrast between intakes slightly below or

above the median is not so large. In addition, it is recognised that the magnitude of diet-related interaction may have been too small for detection in this study. However, the observation of potential weak interaction between *FTO* and fat/SFA deserves some consideration.

Given the *FTO* influence on appetite and satiety, it is reasonable to hypothesize that diet could modify the effect of *FTO*. Previous studies on *FTO*-diet interaction have mainly focused on obesity and diabetes, of which two studies suggest that a healthy diet pattern (higher adherence to a Mediterranean diet) reduces the *FTO* risk allele effect on obesity and diabetes (76, 77). Other studies proposed that a high-fat or high-SFA diet increases the effect of the *FTO* risk allele on obesity (73, 74). However, there are also studies that found no interaction between *FTO* and diet in relation to obesity (77, 108), and a large meta-analysis demonstrated no interaction between *FTO* variants and energy-adjusted intakes of fat, carbohydrate or protein in relation to BMI (75). With respect to CHD, we have found one publication on interaction between *FTO* and dietary fat on total and cardiovascular mortality, which reported no interaction (72).

Considering the previous reports of *FTO*-diet interaction on obesity, and that SFA is expected to be a risk factor for CHD (54), our finding that the *FTO* association with CHD risk was, if anything, slightly strengthened with lower energy-related intake of fat and SFA was unexpected. However, as the *FTO* association with cardiovascular disease appears to be only partly explained by an increased body/fat mass and obesity-related traits (23), there may well be some other effect of the *FTO* variant that is responsible for the increased CHD risk. The *FTO* gene has pleiotropic effects, of which all are not fully understood (35). Consequently, it is perfectly possible that a *FTO*-diet interaction could have different patterns in relation to obesity or CHD. *FTO* variation is also associated with diabetes and insulin resistance (109), so *FTO* might influence glucose metabolism. Since fat/SFA intake were both inversely correlated with carbohydrate intake, an explanation for the observed interaction pattern could have been given by a higher carbohydrate intake in subjects with low fat/SFA intake, especially if most of the carbohydrates gave high glycaemic load. However, as we found no evidence of *FTO* interaction with carbohydrate or sucrose intake, this explanation appears less likely. The underlying mechanism for an *FTO*-fat/SFA interaction, if it would prove to be robust, needs further investigation.

The inverse association between fat or SFA E% and CHD risk in our study population was puzzling. This was consistent in both SHEEP and INTERGENE (although stronger in INTERGENE), in men and women

(stronger in women), and also both in normal-weight and overweight subjects, with a stronger association in normal-weight subjects. Whether the inverse association reflects a true effect or is due to some bias in diet reporting, we can only speculate about (see further discussion of information bias in section 5.6.2). The substitution analyses go some way in helping to explain this finding, as increased fat or SFA were only associated with lower CHD risk when substituting carbohydrate, or when fat substituted protein, but not when SFA substituted PUFA. These findings partly replicate previous studies, which show that replacing SFA with carbohydrates of high glycaemic index increases CHD risk, while replacing SFA with PUFA decreases CHD risk (57, 102). Nevertheless, if the main association of fat/SFA with CHD risk in our study population is due to bias in dietary data, it may also influence the observed *FTO*-fat/SFA interaction tendency on CHD.

The understanding of what characterizes a healthy diet is subject to many studies. Some common dietary factors that are widely replicated as protective of cardiovascular disease are higher intake of fruit and vegetables, higher intake of fibre and lower intake of red and processed meat (54, 55). Macronutrients represent merely one dimension of a complex dietary pattern, and each macronutrient includes a variety of nutrients. Even subgroups of fat are heterogenous, for instance PUFA that contains both $\omega 3$ and $\omega 6$ fatty acids of which $\omega 3$ have more beneficial health effects (56). Consequently, there may be no clear effects of macronutrients on CHD. The role of SFA in cardiovascular disease is also being reevaluated, as more recent analyses suggest that SFA originating from dairy products and coconut oil may have beneficial health effects (56, 110, 111).

Finally, no firm interactions were found between *FTO* and macronutrient composition of diet on CHD risk. However, the suggested weak interaction between *FTO* and fat/SFA in an unexpected direction is an interesting finding, which deserves to be investigated more in future studies.

5.4 *FTO* and *GHRL* interaction on BMI

The observed *FTO*-*GHRL* interaction supports an underlying mechanistic link between effects of the *FTO* protein and ghrelin in appetite-regulating pathways to obesity, as has been reported (39). Associations between *GHRL* polymorphisms and obesity have been demonstrated previously (112). A recent study of 338 common SNPs in 30 genes related to energy homeostasis observed that variation in *GHRL* rs35683 was associated with higher BMI in European Americans, albeit with a small effect size: +0.04 kg/m² per allele

(113). Interestingly, the rs35680 (which is in complete linkage disequilibrium with rs35683) association with BMI in our study had notably higher effect size in the *FTO* TA/AA stratum (+0.81 kg/m², 95% CI 0.05-1.56 per allele), and a nearly opposite effect in the *FTO* TT stratum, resulting in absence of BMI association in the whole sample. The suggested interaction could be of relevance for obesity in a considerable subgroup of the population (roughly 40%), as *FTO* TA/AA genotypes constitute a majority of populations of Caucasian origin (25). Both SNPs rs35680 and the previously studied rs35683 are intronic variants, but their function is yet unknown. Future functional studies of *GHRL* polymorphisms would be of interest and may explain the underlying mechanism for the suggested interaction.

Both *FTO* and *GHRL* polymorphisms have been associated with preference for energy-dense foods (35, 39). In addition, a previous analysis in the current study sample suggested an association between *GHRL* variation and sucrose intake (114). Future studies of possible diet influence on the suggested *FTO*-*GHRL* interaction would be of interest.

A weakness of this study is the small sample. Notably, the observed *FTO*-*GHRL* interaction was found for the *GHRL* SNP with highest minor allele frequency, with corresponding higher power. This does not exclude that the other SNPs under study, or other *GHRL* polymorphisms in linkage disequilibrium with rs35680, may be functionally responsible for the suggested gene-gene interaction.

Given the previously demonstrated mechanistic link of *FTO* and ghrelin, the limited problem with confounding for genetic variants, and the finding of *FTO*-*GHRL* interaction on the additive scale, this finding has potential to be a causal interaction. However, it is also possible that this is a chance finding, and therefore replication in future studies is definitely desirable.

5.5 General discussion of interaction findings

As described in the Introduction, there are two separate definitions of interaction: biological/causal and statistical interaction/effect modification. Effect modification on disease risk can be assessed either on the RD scale or the RR scale. In this thesis, effect modification was assessed on both scales to provide a more complete picture of interaction patterns. Interaction models were also adjusted for possible confounding, where data was available, between each of the interacting factors and CHD, with the aim to find causal

relations and not just associations. Undoubtedly, most investigations are not able to fully meet the strict criteria required to draw firm conclusions about causality. For an understanding of interaction findings, a global view should be taken considering the statistical significance, the scale on which the interaction finding is observed, the biological plausibility of the interaction finding, and to what extent it can be assumed that there is causal interaction. Referring only to the statistical significance of the interaction would not give a complete interpretation of the interaction results. Of note, non-statistically significant interaction results do not prove the absence of interactions, but may be due to limited power, or chance, and if both factors are associated with disease there must be interaction on at least one scale. Therefore, even trends of interactions may be of interest, even though firm conclusions cannot be drawn.

The interaction effects found in this thesis are not so considerable that they would motivate changes in intervention measures in any subgroup with certain genetic variants. However, what the examples of gene-lifestyle interaction may tell us is perhaps above all that the exact effects of risk factors cannot be generalized to all people, and that some subgroups of the population are more susceptible to certain risk factors than others. The interaction findings may also provide added insights into disease mechanisms, as some clues to which gene variants and lifestyle factors participate in common mechanisms are given when biological interaction is present.

5.6 Methodological considerations

5.6.1 Study design

The case-control study measures exposure in actual cases when they have occurred, and in a sample of the study base free of disease at the time of sampling (i.e. the controls), which is an efficient sampling method especially when the disease is uncommon (82, 115). Therefore, the order of measurement in a case-control study is typically from outcome to exposure (retrospective exposure assessment) as opposed to from exposure to outcome (prospective outcome assessment) in a cohort study. However, neither the “backward” directionality of exposure assessment nor the sampling of controls from the study base *per se* make the case-control design less valid than the cohort design (115). But, as any type of epidemiological study, the case-control study design is subject to potential biases, some of which are specific to retrospective exposure assessment.

Causality requires that the exposure occurs before the outcome. Although technically all assessments of exposure were made in the cases and controls at the time of inclusion in the studies, the directionality or timing of the actual exposure in relation to disease status differs somewhat depending on the exposure. Regarding genotypes, they are unchanged from conception with a life-long longitudinal exposure and therefore, the directionality of a specific genotype exposure may be regarded as from exposure to outcome. For variables collected at physical examinations, such as weight, BMI and blood lipids, the exposure assessment is in principle cross-sectional. However, it can be assumed that variables such as BMI have not changed substantially in most participants in the recent past, so at least some degree of directionality from exposure to outcome can be assumed. For smoking, physical activity and diet, retrospectively self-reported in questionnaires, the participants were asked to report their lifestyle for the past year (diet), or the past year (INTERGENE) or 5-10 year interval (SHEEP) for physical activity. Here, at least 1 year of exposure before outcome can be assumed, provided that the reporting is unbiased. For smoking, information about timing of smoking cessation was obtained, so that current smoking included cessation within 1 year before reporting. For lifestyle exposures, the more recent exposure may even be more relevant in relation to risk of CHD, compared to a baseline exposure collected for instance 10 years ago.

5.6.2 Potential systematic bias

Systematic error, or bias, in exposure measurements may produce a tendency of effects in a specific direction that does not reflect the true effect, thus affecting the validity of the results. Systematic bias does not depend on sample size but on the sampling method and quality of exposure assessments (80). Sources of relevant potential systematic biases are discussed below.

Selection bias

Despite the random sampling of controls in SHEEP and INTERGENE, and the intention to include all incident MI cases occurring in the study base (SHEEP) or a pre-specified number of consecutive cases (INTERGENE), selection bias may arise if the exposure pattern differs in participants and non-participants. Such selection bias may influence the exposure associations with risk of CHD.

Selection bias in the SHEEP and INTERGENE studies has been investigated previously. In SHEEP, participation rates were quite high (73% among controls and 83% in cases), and non-participation was not systematically different between age groups or geographical areas, reducing the risk of

selection bias due to these factors. Also, nearly all cases (97%) with first-time MI from the study base were identified, after checking with registries (89).

In INTERGENE, which was conducted about 10 years later than SHEEP, participation was only 42% in the population controls. In general, participation rates in epidemiological studies have declined since the 1990's (116). Participants in the INTERGENE cohort were more likely to be women, older, with higher education and income, married and of Nordic origin than non-participants (117). It was also found that smoking, lower physical activity, obesity and hypertension were related to lower education. However, there are no data on non-participation in INTERGENE cases and therefore the influence of this source of bias on associations is uncertain. It is reasonable to assume that at least partly, similar reasons for non-participation were present in cases, which would reduce distortion of exposure-CHD associations. However, the higher participation rate (81%) in cases should produce somewhat less selection than in controls due to e.g. education level and therefore, CHD associations for smoking, physical inactivity and overweight/obesity may be somewhat overestimated in INTERGENE due to the higher education level in controls. Most of the INTERGENE cases were identified and recruited from the Östra hospital in Gothenburg. However, the hospital has a large catchment area including areas of different socio-economy, and therefore the exposures under study are not likely to be affected by this.

An important aspect is that only non-fatal cases were included in the analyses, which potentially affects the investigated associations between exposures and CHD. It may be that the disease mechanism(s) leading to either a severe CHD event with fatal outcome, or a non-fatal CHD event, are at least partly different. It is probable that the fatal compared to non-fatal CHD cases had more accumulation of risk factors, in which case associations with CHD of any outcome may be under-estimated. In the SHEEP study, where fatal CHD cases were also included, the risk exposure profiles were compared between fatal and non-fatal cases based on questionnaires to close relatives to the fatal cases. It was found that fatal cases had higher frequency of physical inactivity, diabetes, job strain and current smoking (the latter only among women), whereas family history of CHD (i.e. one or both biological parents with CHD before age 65) was less frequent (118). The lower frequency of family history of CHD among fatal cases could imply that the genetic component of CHD has less importance in relation to lifestyle risk factors in fatal cases. Another likely explanation for this observation is that people who are aware of a family history of CHD are more prone to take precautions and adjust their lifestyle, e.g. avoid smoking.

On balance, the known and potential effects from selection bias in cases and controls are not likely to severely distort the associations for the lifestyle risk factors under study with CHD.

Information bias

If the information about exposure or outcome is misclassified, then information bias can arise. If information bias regarding exposure is similar in cases and controls (non-differential), the misclassification tends to dilute a true effect from an exposure towards the null. However, if the information bias is differential in cases and controls, it may either under- or overestimate an effect depending on the direction of bias.

Physical activity was based on retrospective self-report. Some degree of differential reporting is possible, e.g. if cases tended to over-report a harmful exposure such as physical inactivity. Such recall bias is more likely to occur when there is common knowledge or concern about the exposure being harmful (80). A study of recall bias of retrospective self-report of physical activity has been made in Swedish MI patients. The study found no differential recall of leisure-time physical activity between MI cases and controls, while there was some differential recall for occupational activity, with cases tending to over-report heavy occupational activity (119). Thus, it seems reasonable to assume that the associations for leisure time physical activity in this thesis were not severely influenced by information bias.

Assessment of diet is accompanied with many difficulties, and self-reported dietary data is often subject to information bias (120, 121). This may be in the form of misclassification due to the complexity of the questionnaires with many food items and the possibility to misunderstand the questions, or recall bias so that subjects do not recall which food items or how often/how much of some food items they ate, recall some food items better than others, or do not wish to report their true intake.

Underestimation of energy intake is common in self-reported diet, especially among obese (120, 122-124). Underreporting of energy is related to dieting, trying to lose weight and social desirability, and is more common in women than in men (123). Particular food items and nutrients are also more underreported, as has been shown for fat and high-fat foods, especially in obese (124, 125). Also, there is evidence that foods that are considered unhealthy e.g. between-meal snacks, candy, cookies and fat are selectively underreported (123). Since we studied energy-related intakes of macronutrients, a non-specific underreporting of energy is taken care of. But

if underreporting is specific for certain food items and nutrients, then the macronutrient E% will not reflect true conditions.

The FFQ is intended to assess habitual diet of individuals, and includes common food items that are selected to discriminate intakes most between individuals and not necessarily contributing most to absolute intakes. Therefore, the FFQ is not expected to capture absolute intakes of either total energy or nutrients, but rather to rank the individuals in a study with respect to intake of energy, nutrients or food items (94). Due to their relative simplicity, FFQs are the current dominating method for dietary assessment in large-scale studies (126). The FFQ versions used in this thesis are based on an original FFQ developed in the US and validation showed that the overall mean of correlation coefficients comparing intakes of 18 nutrients measured by the FFQ and by a 1-year diet record was 0.60, and the FFQ was able to reasonably categorize individuals by nutrient intake (94, 95). The Swedish version has been additionally validated in part, showing that the energy-related intakes of fatty acids agreed well between the FFQ and the reference of repeated 1-week weighed dietary records (96).

One possible differential recall bias of diet may have arisen due to higher frequency of overweight/obese among cases. If overweight/obese are more prone to selectively underreport fat/SFA intake, this could produce an inverse association between fat/SFA E% and risk of CHD, and thus possibly also influence the interaction results. It appears more unlikely that control subjects would selectively over-report their fat/SFA intake, which could also produce an inverse association. In addition, as the *FTO* risk genotype is associated with overweight/obesity it might add to spurious interaction results. However, the inverse association between fat/SFA and CHD risk was present even after excluding obese subjects, in fact it was strengthened, suggesting that this specific type of recall bias does not explain the inverse association.

Another differential recall bias of diet data could have arisen if CHD cases tended to misreport food items considered unhealthy. Such tendencies have been observed in breast cancer patients, who over-reported intake of fat/SFA, producing a positive association between fat/SFA intake and cancer risk from retrospective report, which was not present in prospective report of diet (127). However, if fat/SFA were over-reported among the cases, this source of bias would not explain the observed inverse association between fat/SFA and risk of CHD. On the contrary, it would rather under-estimate a true inverse association (i.e. produce an OR closer to 1). If, instead, cases tended to under-report their fat/SFA intake (or over-report carbohydrate intake), this could produce an inverse association between fat/SFA and risk of CHD.

However, the literature is scarce regarding possible selective misreporting of diet in CHD patients, and therefore, these biases are hypothetical.

Genotyping technology has a high level of precision although errors do occur, but this error is not expected to be differential regarding exposure or outcome.

Regarding misclassification of the CHD outcomes, the criteria for case selection in both studies were strict and validation of all initially included cases was performed where diagnosis criteria were checked and non-verified cases excluded to ensure high specificity, which should avoid serious misclassification of CHD. In addition, crosschecks with registries of SHEEP MI cases showed that case ascertainment was high (i.e. high sensitivity) with <3% unidentified cases (89), while in INTERGENE, case ascertainment was less complete due to design. For association analyses, high specificity of the outcome is of greater importance than sensitivity, to ensure that misclassification bias of estimated associations remains low.

5.6.3 Residual confounding

Confounding arises when an association between the exposure under study and outcome is affected by some other exposure correlated to the exposure under study and the outcome. Genetic factors are generally not confounded by environmental exposures, but more possibly by other genetic variants in linkage disequilibrium with the genetic variants under study, but as these were unmeasured such confounding could not be controlled for. Also, as both *APOE* and *FTO* have replicated association with CHD in different populations, this type of confounding is likely to be minimal.

Regarding the lifestyle exposures under study, we sought to control for the major confounding factors for which data were available, and different statistical models with a number of possible confounders were compared. For the established and major risk factors smoking and physical inactivity, expected effects on outcomes were observed, suggesting that confounding was not a major concern for these exposures, although some residual confounding may be present. For dietary macronutrients, the risk of residual confounding is probably somewhat more pronounced, regarding that these are not as clear risk factors, and due to the complexity of dietary patterns. However, we controlled for some of the most obvious confounders such as physical activity and smoking in these associations, reducing the risk of residual confounding for diet associations. Overall, there is always a risk for residual confounding, but since we adjusted for the most important lifestyle

risk factors, and these may also take care of some residual confounding if they are correlated with other risk factors (e.g. socio-economy), and sensitivity analyses were made with additional adjustment factors, a severe influence of residual confounding on the results is unlikely.

5.6.4 Reverse causation

As the CHD cases most probably had an accumulation of risk factors before they had an event of CHD (for first-time cases), including manifestations such as hypertension, elevated blood lipids, glucose and insulin, overweight/obesity or type 2 diabetes, there is risk of reverse causation. For example, cases may have decreased their physical activity level due to e.g. chest discomfort, or stopped smoking or changed dietary habits after receiving lifestyle advice from health professionals. The lifestyle data used in the current analyses were reported for the last year or 5-10 years before inclusion in the study. In INTERGENE, about half of the cases had a previous diagnosis of CHD, increasing potential reverse causation when including these cases in analyses. To account for this source of bias, we made sensitivity analyses excluding cases with previous diagnosis of CHD in paper II. In Paper III, they were excluded entirely, since the dietary associations were assumed to be at greater risk of reverse causation. Further, smoking was categorized both as ever/never smoking and as current/former/never smoking, to capture effects of past smoking.

Regarding diet associations with CHD, they may have been affected by reverse causation if cases had changed their dietary habits towards a more healthy diet in the more recent past. The general guidance has often been to reduce intake of SFA, sugar-rich food and increase intake of PUFA and fruit/vegetables. This could lead to cases being more likely to have e.g. reduced SFA intake more recently. Interestingly, we observed that subjects reporting diet change had lower fat and SFA E% (and also higher PUFA, protein and carbohydrate E%) than subjects reporting stable dietary habits, which supports this hypothesis. In addition, INTERGENE had higher frequency of cases reporting diet change than SHEEP, and the inverse fat/SFA association with CHD risk was also stronger in INTERGENE. Therefore, a reverse causation bias may have contributed to the inverse fat/SFA association with CHD. However, when subjects reporting a change in dietary habits were excluded in sensitivity analyses, the interaction pattern for *FTO* and fat/SFA was slightly strengthened, which suggests that a reverse causation bias (at least not with respect to fat/SFA intake) did not explain the observed interaction tendency.

6 CONCLUSIONS

An interaction between *APOE* and smoking on CHD risk was observed as effect modification both on the relative risk and risk difference scales, such that subjects carrying the *APOE* $\epsilon 2$ allele had lower smoking-related CHD risk than subjects without the $\epsilon 2$ allele ($\epsilon 3\epsilon 3$, $\epsilon 3\epsilon 4$ and $\epsilon 4\epsilon 4$ genotypes). This finding largely confirms previous reports, but also clearly demonstrates more than additive effects of *APOE* and smoking on CHD risk, which indicates presence of biological interaction. Another novel finding is that women carrying the $\epsilon 4$ allele appeared particularly susceptible to the smoking-related CHD risk.

Potential interaction between *APOE* and physical activity on CHD risk was also found, both on the relative risk and risk difference scales, which is a novel finding. The CHD risk from physical inactivity was less pronounced in $\epsilon 2$ allele carriers, while it was stronger in $\epsilon 4$ allele carriers, compared to subjects with the common genotype $\epsilon 3\epsilon 3$. There were more than additive effects of the $\epsilon 4$ allele and physical inactivity in relation to the $\epsilon 2$ allele and physical activity on CHD risk, suggesting the presence of biological interaction. No interaction was found for *APOE*-overweight on CHD risk, or between *APOE* and any lifestyle factor on LDL-C levels. This suggests another mechanism for the *APOE* interactions with smoking or physical inactivity, with one possible mechanism being the antioxidant gradient apoE2>apoE3>apoE4 acting in the oxidative stress from smoking and physical inactivity.

The *FTO* variant rs9939609 minor allele A was associated with about 20% increased CHD risk, independent of BMI. No evidence of interaction was found between the *FTO* gene variant and physical activity on CHD risk either on the relative risk or risk difference scale, indicating that the *FTO*-related CHD risk is not counteracted by increased physical activity. The combined effects of *FTO* risk allele and lower physical activity level on CHD risk were approximately additive, suggesting independent effects of these factors on CHD risk. There was a tendency of weaker *FTO* effect on BMI with increasing physical activity, but without statistically significant interaction.

No clear interactions were found between *FTO* gene variants and dietary intake of macronutrients on CHD risk. However, in subjects reporting stable dietary habits in recent years, a weak interaction was found on the relative risk scale between *FTO* and fat or SFA, with slightly stronger *FTO*-attributable CHD risk with lower energy-related intake of fat or SFA. No

interactions were found between *FTO* gene variation and macronutrients on BMI.

Finally, a gene-gene interaction was demonstrated for the first time between *FTO* and *GHRL* variant rs35680 in relation to BMI, where the minor alleles of these gene variants had synergistic effects on BMI. This supports a mechanistic interaction between the *FTO* and *GHRL* genes in the development of obesity.

7 FUTURE PERSPECTIVES

The novel suggested interaction findings in this thesis should be replicated in future studies to determine if they are robust interactions and not just chance findings. Further, follow-up of epidemiological interaction findings with experimental studies investigating the function of genetic variants could provide information on the biological mechanisms underlying an observed epidemiological interaction. Regarding the *FTO* gene and its polymorphism, the exact mechanisms are still unknown (35), especially in relation to CHD. Since *FTO* probably has pleiotropic effects, it would be of interest to investigate what other effects in addition to the regulation of energy balance and appetite could be responsible for the association with CHD.

Analysis of interactions is still an emerging field with several challenges. Given the multifactorial background of CHD, the number of possible interactions appears huge. Studying only two-way interactions may be too simplistic as in a more realistic scenario several factors probably interact in the same disease mechanism. However, including more factors requires additional statistical power and increases the complexity of interpreting interaction patterns. Today, reports of 3-way interactions and beyond are still rare. Large study samples would be required, and with high-quality lifestyle data. Such large-scale studies have been implemented in more recent years, e.g. the Swedish Lifegene study aiming to include 500,000 individuals with detailed and longitudinal exposure assessments (128), which provides promising future opportunities for interaction analyses. Another approach of studying gene-environment interaction is by hypothesis-free screening of interactions in a large-scale GWAS type of analysis. However, hypotheses based on biologically plausible mechanisms are also needed for understanding interaction patterns, so different approaches for studying interactions complement one another.

Another complexity of interaction analyses is the choice of definition of interaction and analysis method to be used. Surprisingly, many reports of interaction in relation to disease risk are still often based solely on effect modification of relative risks, despite some available guidelines to present sufficient data to assess effect modification on both the absolute and relative scales (85-88), and the risk of inconsistent interpretation from analyses on the different scales (86). Clearly, the understanding of interaction findings would benefit from using both scales in analyses, and therefore meta-analyses of previous interaction studies could be done with application of both scales, provided the appropriate data from each study is available.

There appears to be great expectations on application of gene-environment interaction findings in personalized medicine, i.e. targeted prevention such as lifestyle advice and treatment to genetically susceptible individuals. This is predicted to be the future of public health (58). However, the road to applying interaction findings in clinical practice seems long, as many of the findings still need replication in a large scale and in different populations. Most findings so far point towards the well-known lifestyle risk factors really being risk factors in individuals of different genotypes, so it is more a question of to what degree and in which ways they increase the risk. Also, due to the complex, multifactorial pathophysiology of CHD, there are probably additional factors (a third, fourth factor and beyond) that also influence specific two-way interactions, which may need to be taken into account for applications in personalized medicine. A more immediate use of interaction findings is to generate more specific hypotheses of disease mechanisms, and use this information in further studies, both experimental and epidemiological. This will lead to more detailed characterization of the causal mechanisms leading to CHD, which will also eventually have potential applications in personalized medicine.

ACKNOWLEDGEMENT

I would like to express my gratitude to several people who have contributed in different ways to the completion of this thesis:

Fredrik Nyberg, my main supervisor, for offering me the opportunity to do this interesting PhD project, for excellent guidance in epidemiological research and for giving me such freedom to go my own paths in research.

My co-supervisors: *Kirsten Mehlig* for all knowledgeable and immediate advice in statistical methods, interpretation and programming, *Karin Leander* for moral support (especially in interaction analyses) and good advice. It was always a pleasure visiting you at Karolinska Institutet! *Elisabeth Strandhagen* for good advice, support and your cheerful attitude, *Lauren Lissner* for providing constructive advice and encouragement to follow my own ideas. It has been a true privilege having you all on board!

My co-authors *Dag Thelle*, *Lena Björck*, *Annika Rosengren*, *Henrik Zetterberg* and *Kaj Blennow* for valuable input. Special thanks to *Christina Berg* and *Gianluca Tognon* for advice in the complex but fascinating field of diet. Thanks also to *Lars Barregård* and *Oscar Hammar* for preparing wise questions on my articles, and *Gösta Axelsson* for valuable advice in epidemiology.

Kjell Torén, for support and aiding a friendly research atmosphere at Occupational and Environmental Medicine (AMM), and all colleagues at AMM and Public Health Epidemiology, for making work fun and stimulating. My fellow PhD students *Ebba Brann*, *Santosh Dahgam*, *Sofie Fredriksson*, *Lotta Nybergh*, *Hanne Krage Carlsen* and *Petra Brembeck* for friendship and good company. Special thanks to *Sara Gustavsson*, also my room-mate, for sharing the ups and downs of PhD work, and being an enthusiastic discussion partner, and *Anna Levinsson* for finding the course on interaction in Halle (Spitze!) and for nice travelling company.

Jag vill även tacka alla vänner och hela min familj för all uppmuntran under doktorandtiden. Särskilt tack till alla fina vänner som erbjöd sovplats då jag pendlade mellan Värnamo och Göteborg, det betydde väldigt mycket. Till sist vill jag tacka *Fredrik*, den coolaste mannen i universum, för att du alltid finns där som ett orubbligt stöd, samt våra barn *Elina* och *Ville*, för att ni lånat ut av er tid med mamma till forskningen. Ni är mitt allt. *Punkt.*

REFERENCES

1. The global burden of disease update. World Health Organisation (WHO), 2004.
2. Health in the European Union- Trends and analyses. World Health Organisation (WHO), 2009.
3. www.socialstyrelsen.se/statistics.
4. Anand SS, Islam S, Rosengren A, Franzosi MG, Steyn K, Yusufali AH, et al. Risk factors for myocardial infarction in women and men: insights from the INTERHEART study. *European heart journal*. 2008;29(7):932-40.
5. Thygesen K, Alpert JS, Jaffe AS, Simoons ML, Chaitman BR, White HD, et al. Third universal definition of myocardial infarction. *Circulation*. 2012;126(16):2020-35.
6. Steinberg D. Thematic review series: the pathogenesis of atherosclerosis: an interpretive history of the cholesterol controversy, part III: mechanistically defining the role of hyperlipidemia. *Journal of lipid research*. 2005;46(10):2037-51.
7. Roberts R, Stewart AF. Genes and coronary artery disease: where are we? *Journal of the American College of Cardiology*. 2012;60(18):1715-21.
8. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med*. 2002;252(3):247-54.
9. Roberts R, Stewart AF, Wells GA, Williams KA, Kavaslar N, McPherson R. Identifying genes for coronary artery disease: An idea whose time has come. *The Canadian journal of cardiology*. 2007;23 Suppl A:7A-15A.
10. Mayer B, Erdmann J, Schunkert H. Genetics and heritability of coronary artery disease and myocardial infarction. *Clinical research in cardiology : official journal of the German Cardiac Society*. 2007;96(1):1-7.
11. Peden JF, Farrall M. Thirty-five common variants for coronary artery disease: the fruits of much collaborative labour. *Hum Mol Genet*. 2011;20(R2):R198-205.
12. Consortium CAD, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45(1):25-33.

13. Visvikis-Siest S, Marteau JB. Genetic variants predisposing to cardiovascular disease. *Current opinion in lipidology*. 2006;17(2):139-51.
14. Mahley RW. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science*. 1988;240(4852):622-30.
15. Weisgraber KH, Rall SC, Jr., Mahley RW. Human E apoprotein heterogeneity. Cysteine-arginine interchanges in the amino acid sequence of the apo-E isoforms. *The Journal of biological chemistry*. 1981;256(17):9077-83.
16. Bennet AM, Di Angelantonio E, Ye Z, Wensley F, Dahlin A, Ahlbom A, et al. Association of apolipoprotein E genotypes with lipid levels and coronary risk. *JAMA*. 2007;298(11):1300-11.
17. Song Y, Stampfer MJ, Liu S. Meta-analysis: apolipoprotein E genotypes and risk for coronary heart disease. *Ann Intern Med*. 2004;141(2):137-47.
18. Wilson PW, Schaefer EJ, Larson MG, Ordovas JM. Apolipoprotein E alleles and risk of coronary disease. A meta-analysis. *Arterioscler Thromb Vasc Biol*. 1996;16(10):1250-5.
19. Doney AS, Dannfald J, Kimber CH, Donnelly LA, Pearson E, Morris AD, et al. The FTO gene is associated with an atherogenic lipid profile and myocardial infarction in patients with type 2 diabetes: a Genetics of Diabetes Audit and Research Study in Tayside Scotland (Go-DARTS) study. *Circ Cardiovasc Genet*. 2009;2(3):255-9.
20. He M, Cornelis MC, Franks PW, Zhang C, Hu FB, Qi L. Obesity genotype score and cardiovascular risk in women with type 2 diabetes mellitus. *Arterioscler Thromb Vasc Biol*. 2010;30(2):327-32.
21. Hubacek JA, Stanek V, Gebauerova M, Pilipcincova A, Dlouha D, Poledne R, et al. A FTO variant and risk of acute coronary syndrome. *Clin Chim Acta*. 2010;411(15-16):1069-72.
22. Lappalainen T, Kolehmainen M, Schwab US, Tolppanen AM, Stancakova A, Lindstrom J, et al. Association of the FTO gene variant (rs9939609) with cardiovascular disease in men with abnormal glucose metabolism--the Finnish Diabetes Prevention Study. *Nutr Metab Cardiovasc Dis*. 2011;21(9):691-8.
23. Liu C, Mou S, Pan C. The FTO gene rs9939609 polymorphism predicts risk of cardiovascular disease: a systematic review and meta-analysis. *PLoS One*. 2013;8(8):e71901.

24. Dina C, Meyre D, Gallina S, Durand E, Korner A, Jacobson P, et al. Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet.* 2007;39(6):724-6.
25. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science.* 2007;316(5826):889-94.
26. Scuteri A, Sanna S, Chen WM, Uda M, Albai G, Strait J, et al. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet.* 2007;3(7):e115.
27. Fisher E, Schulze MB, Stefan N, Haring HU, Doring F, Joost HG, et al. Association of the FTO rs9939609 single nucleotide polymorphism with C-reactive protein levels. *Obesity (Silver Spring).* 2009;17(2):330-4.
28. Pausova Z, Syme C, Abrahamowicz M, Xiao Y, Leonard GT, Perron M, et al. A common variant of the FTO gene is associated with not only increased adiposity but also elevated blood pressure in French Canadians. *Circ Cardiovasc Genet.* 2009;2(3):260-9.
29. Freathy RM, Timpson NJ, Lawlor DA, Pouta A, Ben-Shlomo Y, Ruokonen A, et al. Common variation in the FTO gene alters diabetes-related metabolic traits to the extent expected given its effect on BMI. *Diabetes.* 2008;57(5):1419-26.
30. Berentzen T, Kring SI, Holst C, Zimmermann E, Jess T, Hansen T, et al. Lack of association of fatness-related FTO gene variants with energy expenditure or physical activity. *J Clin Endocrinol Metab.* 2008;93(7):2904-8.
31. Fredriksson R, Hagglund M, Olszewski PK, Stephansson O, Jacobsson JA, Olszewska AM, et al. The obesity gene, FTO, is of ancient origin, up-regulated during food deprivation and expressed in neurons of feeding-related nuclei of the brain. *Endocrinology.* 2008;149(5):2062-71.
32. Olszewski PK, Fredriksson R, Olszewska AM, Stephansson O, Alsio J, Radomska KJ, et al. Hypothalamic FTO is associated with the regulation of energy intake not feeding reward. *BMC Neurosci.* 2009;10:129.
33. Speakman JR. FTO effect on energy demand versus food intake. *Nature.* 2010;464(7289):E1; discussion E2.
34. Stratigopoulos G, Padilla SL, LeDuc CA, Watson E, Hattersley AT, McCarthy MI, et al. Regulation of Fto/Ftm gene expression in mice and humans. *American journal of physiology Regulatory, integrative and comparative physiology.* 2008;294(4):R1185-96.

35. Larder R, Cheung MK, Tung YC, Yeo GS, Coll AP. Where to go with FTO? Trends in endocrinology and metabolism: TEM. 2011;22(2):53-9.
36. Fischer J, Koch L, Emmerling C, Vierkotten J, Peters T, Bruning JC, et al. Inactivation of the Fto gene protects from obesity. Nature. 2009;458(7240):894-8.
37. Church C, Lee S, Bagg EA, McTaggart JS, Deacon R, Gerken T, et al. A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. PLoS Genet. 2009;5(8):e1000599.
38. Church C, Moir L, McMurray F, Girard C, Banks GT, Teboul L, et al. Overexpression of Fto leads to increased food intake and results in obesity. Nat Genet. 2010;42(12):1086-92.
39. Karra E, O'Daly OG, Choudhury AI, Yousseif A, Millership S, Neary MT, et al. A link between FTO, ghrelin, and impaired brain food-cue responsivity. The Journal of clinical investigation. 2013;123(8):3539-51.
40. Tired L. Gene-environment interaction: a central concept in multifactorial diseases. The Proceedings of the Nutrition Society. 2002;61(4):457-63.
41. Talmud PJ. Commentary on the paper by Gustavsson et al. entitled 'Interaction of apolipoprotein E genotype with smoking and physical inactivity on coronary heart disease risk in men and women'. Atherosclerosis. 2012;220(2):323-4.
42. Jha P, Peto R, Zatonski W, Boreham J, Jarvis MJ, Lopez AD. Social inequalities in male mortality, and in male mortality from smoking: indirect estimation from national death rates in England and Wales, Poland, and North America. Lancet. 2006;368(9533):367-70.
43. Williams PT. Physical fitness and activity as separate heart disease risk factors: a meta-analysis. Medicine and science in sports and exercise. 2001;33(5):754-61.
44. Held C, Iqbal R, Lear SA, Rosengren A, Islam S, Mathew J, et al. Physical activity levels, ownership of goods promoting sedentary behaviour and risk of myocardial infarction: results of the INTERHEART study. European heart journal. 2012;33(4):452-66.
45. Erlichman J, Kerbey AL, James WP. Physical activity and its impact on health outcomes. Paper 1: The impact of physical activity on cardiovascular disease and all-cause mortality: an historical perspective. Obes Rev. 2002;3(4):257-71.
46. Lee IM, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden

- of disease and life expectancy. *Lancet*. 2012;380(9838):219-29.
47. Leon AS, Sanchez OA. Response of blood lipids to exercise training alone or combined with dietary intervention. *Medicine and science in sports and exercise*. 2001;33(6 Suppl):S502-15; discussion S28-9.
48. Pescatello LS, Franklin BA, Fagard R, Farquhar WB, Kelley GA, Ray CA, et al. American College of Sports Medicine position stand. Exercise and hypertension. *Medicine and science in sports and exercise*. 2004;36(3):533-53.
49. Borghouts LB, Keizer HA. Exercise and insulin sensitivity: a review. *International journal of sports medicine*. 2000;21(1):1-12.
50. Wilmot EG, Edwardson CL, Achana FA, Davies MJ, Gorely T, Gray LJ, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. *Diabetologia*. 2012;55(11):2895-905.
51. Howard BJ, Fraser SF, Sethi P, Cerin E, Hamilton MT, Owen N, et al. Impact on hemostatic parameters of interrupting sitting with intermittent activity. *Medicine and science in sports and exercise*. 2013;45(7):1285-91.
52. Church TS, Thomas DM, Tudor-Locke C, Katzmarzyk PT, Earnest CP, Rodarte RQ, et al. Trends over 5 decades in U.S. occupation-related physical activity and their associations with obesity. *PLoS One*. 2011;6(5):e19657.
53. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, et al. Global physical activity levels: surveillance progress, pitfalls, and prospects. *Lancet*. 2012;380(9838):247-57.
54. Nordic Nutrition Recommendations 2012. 2014.
55. Iqbal R, Anand S, Ounpuu S, Islam S, Zhang X, Rangarajan S, et al. Dietary patterns and the risk of acute myocardial infarction in 52 countries: results of the INTERHEART study. *Circulation*. 2008;118(19):1929-37.
56. Lawrence GD. Dietary fats and health: dietary recommendations in the context of scientific evidence. *Advances in nutrition*. 2013;4(3):294-302.
57. Jakobsen MU, Dethlefsen C, Joensen AM, Stegger J, Tjonneland A, Schmidt EB, et al. Intake of carbohydrates compared with intake of saturated fatty acids and risk of myocardial infarction: importance of the glycemic index. *Am J Clin Nutr*. 2010;91(6):1764-8.
58. Dempfle A, Scherag A, Hein R, Beckmann L, Chang-Claude J, Schafer H. Gene-environment interactions for complex

- traits: definitions, methodological requirements and challenges. *European journal of human genetics : EJHG*. 2008;16(10):1164-72.
59. Corella D, Guillen M, Saiz C, Portoles O, Sabater A, Cortina S, et al. Environmental factors modulate the effect of the APOE genetic polymorphism on plasma lipid concentrations: ecogenetic studies in a Mediterranean Spanish population. *Metabolism: clinical and experimental*. 2001;50(8):936-44.
 60. Humphries SE, Talmud PJ, Hawe E, Bolla M, Day IN, Miller GJ. Apolipoprotein E4 and coronary heart disease in middle-aged men who smoke: a prospective study. *Lancet*. 2001;358(9276):115-9.
 61. Talmud PJ, Stephens JW, Hawe E, Demissie S, Cupples LA, Hurel SJ, et al. The significant increase in cardiovascular disease risk in APOEepsilon4 carriers is evident only in men who smoke: potential relationship between reduced antioxidant status and ApoE4. *Annals of human genetics*. 2005;69(Pt 6):613-22.
 62. Pezzini A, Grassi M, Del Zotto E, Bazzoli E, Archetti S, Assanelli D, et al. Synergistic effect of apolipoprotein E polymorphisms and cigarette smoking on risk of ischemic stroke in young adults. *Stroke*. 2004;35(2):438-42.
 63. Talmud PJ. Gene-environment interaction and its impact on coronary heart disease risk. *Nutr Metab Cardiovasc Dis*. 2007;17(2):148-52.
 64. Keavney B, Parish S, Palmer A, Clark S, Youngman L, Danesh J, et al. Large-scale evidence that the cardiotoxicity of smoking is not significantly modified by the apolipoprotein E epsilon2/epsilon3/epsilon4 genotype. *Lancet*. 2003;361(9355):396-8.
 65. Ward H, Mitrou PN, Bowman R, Luben R, Wareham NJ, Khaw KT, et al. APOE genotype, lipids, and coronary heart disease risk: a prospective population study. *Arch Intern Med*. 2009;169(15):1424-9.
 66. Taimela S, Lehtimaki T, Porkka KV, Rasanen L, Viikari JS. The effect of physical activity on serum total and low-density lipoprotein cholesterol concentrations varies with apolipoprotein E phenotype in male children and young adults: The Cardiovascular Risk in Young Finns Study. *Metabolism: clinical and experimental*. 1996;45(7):797-803.
 67. Hagberg JM, Ferrell RE, Katzel LI, Dengel DR, Sorkin JD, Goldberg AP. Apolipoprotein E genotype and exercise training-induced increases in plasma high-density lipoprotein (HDL)- and HDL2-cholesterol levels in overweight men. *Metabolism: clinical and experimental*. 1999;48(8):943-5.

68. Bernstein MS, Costanza MC, James RW, Morris MA, Cambien F, Raoux S, et al. Physical activity may modulate effects of ApoE genotype on lipid profile. *Arterioscler Thromb Vasc Biol.* 2002;22(1):133-40.
69. Boer JM, Ehnholm C, Menzel HJ, Havekes LM, Rosseneu M, O'Reilly DS, et al. Interactions between lifestyle-related factors and the ApoE polymorphism on plasma lipids and apolipoproteins. The EARS Study. *European Atherosclerosis Research Study. Arterioscler Thromb Vasc Biol.* 1997;17(9):1675-81.
70. Kilpelainen TO, Qi L, Brage S, Sharp SJ, Sonestedt E, Demerath E, et al. Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. *PLoS Med.* 2011;8(11):e1001116.
71. Ahmad T, Chasman DI, Mora S, Pare G, Cook NR, Buring JE, et al. The fat-mass and obesity-associated (FTO) gene, physical activity, and risk of incident cardiovascular events in white women. *Am Heart J.* 2010;160(6):1163-9.
72. Sonestedt E, Gullberg B, Ericson U, Wirfalt E, Hedblad B, Orho-Melander M. Association between fat intake, physical activity and mortality depending on genetic variation in FTO. *Int J Obes (Lond).* 2011;35(8):1041-9.
73. Sonestedt E, Roos C, Gullberg B, Ericson U, Wirfalt E, Orho-Melander M. Fat and carbohydrate intake modify the association between genetic variation in the FTO genotype and obesity. *Am J Clin Nutr.* 2009;90(5):1418-25.
74. Corella D, Arnett DK, Tucker KL, Kabagambe EK, Tsai M, Parnell LD, et al. A high intake of saturated fatty acids strengthens the association between the fat mass and obesity-associated gene and BMI. *The Journal of nutrition.* 2011;141(12):2219-25.
75. Qi Q, Kilpelainen TO, Downer MK, Tanaka T, Smith CE, Sluijs I, et al. FTO genetic variants, dietary intake and body mass index: insights from 177 330 individuals. *Hum Mol Genet.* 2014.
76. Ortega-Azorin C, Sorli JV, Asensio EM, Coltell O, Martinez-Gonzalez MA, Salas-Salvado J, et al. Associations of the FTO rs9939609 and the MC4R rs17782313 polymorphisms with type 2 diabetes are modulated by diet, being higher when adherence to the Mediterranean diet pattern is low. *Cardiovascular diabetology.* 2012;11:137.
77. Corella D, Ortega-Azorin C, Sorli JV, Covas MI, Carrasco P, Salas-Salvado J, et al. Statistical and biological gene-lifestyle interactions of MC4R and FTO with diet and physical activity

- on obesity: new effects on alcohol consumption. *PLoS One*. 2012;7(12):e52344.
78. Stephens JW, Bain SC, Humphries SE. Gene-environment interaction and oxidative stress in cardiovascular disease. *Atherosclerosis*. 2008;200(2):229-38.
79. de Mutsert R, Jager KJ, Zoccali C, Dekker FW. The effect of joint exposures: examining the presence of interaction. *Kidney international*. 2009;75(7):677-81.
80. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Philadelphia: Lippincott Williams & Wilkins; 2008.
81. VanderWeele TJ. Sufficient cause interactions and statistical interactions. *Epidemiology*. 2009;20(1):6-13.
82. KJ. R. *Epidemiology An introduction*. New York: Oxford University Press Inc.; 2012.
83. VanderWeele TJ, Robins JM. The identification of synergism in the sufficient-component-cause framework. *Epidemiology*. 2007;18(3):329-39.
84. Ahlbom A, Alfredsson L. Interaction: A word with two meanings creates confusion. *Eur J Epidemiol*. 2005;20(7):563-4.
85. Knol MJ, Egger M, Scott P, Geerlings MI, Vandembroucke JP. When one depends on the other: reporting of interaction in case-control and cohort studies. *Epidemiology*. 2009;20(2):161-6.
86. Hallqvist J, Ahlbom A, Diderichsen F, Reuterwall C. How to evaluate interaction between causes: a review of practices in cardiovascular epidemiology. *J Intern Med*. 1996;239(5):377-82.
87. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. *Int J Epidemiol*. 2012;41(2):514-20.
88. Vandembroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Epidemiology*. 2007;18(6):805-35.
89. Reuterwall C, Hallqvist J, Ahlbom A, De Faire U, Diderichsen F, Hogstedt C, et al. Higher relative, but lower absolute risks of myocardial infarction in women than in men: analysis of some major risk factors in the SHEEP study. The SHEEP Study Group. *J Intern Med*. 1999;246(2):161-74.
90. Leander K, Hallqvist J, Reuterwall C, Ahlbom A, de Faire U. Family history of coronary heart disease, a strong risk factor for myocardial infarction interacting with other cardiovascular

- risk factors: results from the Stockholm Heart Epidemiology Program (SHEEP). *Epidemiology*. 2001;12(2):215-21.
91. Leander K, Wiman B, Hallqvist J, Sten-Linder M, de Faire U, Stockholm Heart Epidemiology P. PAI-1 level and the PAI-1 4G/5G polymorphism in relation to risk of non-fatal myocardial infarction: results from the Stockholm Heart Epidemiology Program (SHEEP). *Thromb Haemost*. 2003;89(6):1064-71.
92. Aires N, Selmer R, Thelle D. The validity of self-reported leisure time physical activity, and its relationship to serum cholesterol, blood pressure and body mass index. A population based study of 332,182 men and women aged 40-42 years. *Eur J Epidemiol*. 2003;18(6):479-85.
93. Saltin B, Grimby G. Physiological analysis of middle-aged and old former athletes. Comparison with still active athletes of the same ages. *Circulation*. 1968;38(6):1104-15.
94. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol*. 1985;122(1):51-65.
95. Willett WC, Reynolds RD, Cottrell-Hoehner S, Sampson L, Browne ML. Validation of a semi-quantitative food frequency questionnaire: comparison with a 1-year diet record. *Journal of the American Dietetic Association*. 1987;87(1):43-7.
96. Wolk A, Ljung H, Vessby B, Hunter D, Willett WC. Effect of additional questions about fat on the validity of fat estimates from a food frequency questionnaire. Study Group of MRS SWEA. *European journal of clinical nutrition*. 1998;52(3):186-92.
97. Mifflin MD, St Jeor ST, Hill LA, Scott BJ, Daugherty SA, Koh YO. A new predictive equation for resting energy expenditure in healthy individuals. *Am J Clin Nutr*. 1990;51(2):241-7.
98. Energy and protein requirements. Report of a joint FAO/WHO/UNU Expert Consultation. World Health Organization technical report series. 1985;724:1-206.
99. Blennow K, Ricksten A, Prince JA, Brookes AJ, Emahazion T, Wasslavik C, et al. No association between the alpha2-macroglobulin (A2M) deletion and Alzheimer's disease, and no change in A2M mRNA, protein, or protein expression. *Journal of neural transmission*. 2000;107(8-9):1065-79.
100. Landgren S, Jerlhag E, Zetterberg H, Gonzalez-Quintela A, Campos J, Olofsson U, et al. Association of pro-ghrelin and GHS-R1A gene polymorphisms and haplotypes with heavy

- alcohol use and body mass. *Alcoholism, clinical and experimental research*. 2008;32(12):2054-61.
101. Pearce N. What does the odds ratio estimate in a case-control study? *Int J Epidemiol*. 1993;22(6):1189-92.
102. Jakobsen MU, O'Reilly EJ, Heitmann BL, Pereira MA, Balter K, Fraser GE, et al. Major types of dietary fat and risk of coronary heart disease: a pooled analysis of 11 cohort studies. *Am J Clin Nutr*. 2009;89(5):1425-32.
103. Gauderman WJ. Sample size requirements for matched case-control studies of gene-environment interaction. *Stat Med*. 2002;21(1):35-50.
104. Corella D, Ordovas JM. Aging and cardiovascular diseases: the role of gene-diet interactions. *Ageing research reviews*. 2014;18:53-73.
105. Ordovas JM. Gender, a significant factor in the cross talk between genes, environment, and health. *Gender medicine*. 2007;4 Suppl B:S111-22.
106. Laufs U, Wassmann S, Czech T, Munzel T, Eisenhauer M, Bohm M, et al. Physical inactivity increases oxidative stress, endothelial dysfunction, and atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2005;25(4):809-14.
107. Gerken T, Girard CA, Tung YC, Webby CJ, Saudek V, Hewitson KS, et al. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science*. 2007;318(5855):1469-72.
108. Vimalaewaran KS, Angquist L, Hansen RD, van der AD, Bouatia-Naji N, Holst C, et al. Association between FTO variant and change in body weight and its interaction with dietary factors: the DiOGenes study. *Obesity (Silver Spring)*. 2012;20(8):1669-74.
109. Tschritter O, Preissl H, Yokoyama Y, Machicao F, Haring HU, Fritsche A. Variation in the FTO gene locus is associated with cerebrocortical insulin resistance in humans. *Diabetologia*. 2007;50(12):2602-3.
110. Siri-Tarino PW, Sun Q, Hu FB, Krauss RM. Saturated fat, carbohydrate, and cardiovascular disease. *Am J Clin Nutr*. 2010;91(3):502-9.
111. Siri-Tarino PW, Sun Q, Hu FB, Krauss RM. Meta-analysis of prospective cohort studies evaluating the association of saturated fat with cardiovascular disease. *Am J Clin Nutr*. 2010;91(3):535-46.
112. Ukkola O, Ravussin E, Jacobson P, Perusse L, Rankinen T, Tschop M, et al. Role of ghrelin polymorphisms in obesity based on three different studies. *Obesity research*. 2002;10(8):782-91.

113. Li P, Tiwari HK, Lin WY, Allison DB, Chung WK, Leibel RL, et al. Genetic association analysis of 30 genes related to obesity in a European American population. *Int J Obes (Lond)*. 2014;38(5):724-9.
114. Landgren S, Simms JA, Thelle DS, Strandhagen E, Bartlett SE, Engel JA, et al. The ghrelin signalling system is involved in the consumption of sweets. *PLoS One*. 2011;6(3):e18170.
115. Kopec JA, Esdaile JM. Bias in case-control studies. A review. *Journal of epidemiology and community health*. 1990;44(3):179-86.
116. Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. *Am J Epidemiol*. 2006;163(3):197-203.
117. Strandhagen E, Berg C, Lissner L, Nunez L, Rosengren A, Toren K, et al. Selection bias in a population survey with registry linkage: potential effect on socioeconomic gradient in cardiovascular risk. *Eur J Epidemiol*. 2010;25(3):163-72.
118. Leander K. Family history in relation to myocardial infarction, and analyses of gene-environment interactions involving factors of haemostasis. Stockholm 2005.
119. Fransson E, Knutsson A, Westerholm P, Alfredsson L. Indications of recall bias found in a retrospective study of physical activity and myocardial infarction. *J Clin Epidemiol*. 2008;61(8):840-7.
120. Hill RJ, Davies PS. The validity of self-reported energy intake as determined using the doubly labelled water technique. *The British journal of nutrition*. 2001;85(4):415-30.
121. Westerterp KR, Goris AH. Validity of the assessment of dietary intake: problems of misreporting. *Current opinion in clinical nutrition and metabolic care*. 2002;5(5):489-93.
122. Black AE, Prentice AM, Goldberg GR, Jebb SA, Bingham SA, Livingstone MB, et al. Measurements of total energy expenditure provide insights into the validity of dietary measurements of energy intake. *Journal of the American Dietetic Association*. 1993;93(5):572-9.
123. Livingstone MB, Black AE. Markers of the validity of reported energy intake. *The Journal of nutrition*. 2003;133 Suppl 3:895S-920S.
124. Lissner L. Measuring food intake in studies of obesity. *Public health nutrition*. 2002;5(6A):889-92.
125. Heitmann BL, Lissner L. Dietary underreporting by obese individuals--is it specific or non-specific? *Bmj*. 1995;311(7011):986-9.

126. Willett WC, Hu FB. Not the time to abandon the food frequency questionnaire: point. *Cancer Epidemiol Biomarkers Prev.* 2006;15(10):1757-8.
127. Giovannucci E, Stampfer MJ, Colditz GA, Manson JE, Rosner, Longnecker M, et al. A comparison of prospective and retrospective assessments of diet in the study of breast cancer. *Am J Epidemiol.* 1993;137(5):502-11.
128. Almqvist C, Adami HO, Franks PW, Groop L, Ingelsson E, Kere J, et al. LifeGene--a large prospective population-based study of global relevance. *Eur J Epidemiol.* 2011;26(1):67-77.

APPENDIX

Table A1. Dietary intake of macronutrients in SHEEP and INTERGENE by case status

Macronutrients, daily intake	Population controls		Cases		
	SHEEP n=1422	INTERGENE n=2868	SHEEP n=1092	INTERGENE n=289	
Non-alcohol energy, kcal	2075	2157	2156	2263	**
Fat, g	74.4	77.9	77.1	78.2	
Fat, E%	32.0	32.3	31.8	30.8	*
SFA, g	35.1	35.0	36.5	34.1	
SFA, E%	15.0	14.4	15.0	13.4	***
MUFA, g	24.5	26.4	25.5	26.7	
MUFA, E%	10.6	11.0	10.5	10.6	
PUFA, g	9.1	10.6	9.3	11.4	***
PUFA, E%	4.0	4.5	3.9	4.5	***
Protein, g	87.8	92.5	92.6	100.0	***
Protein, E%	17.0	17.3	17.2	17.8	***
Carbohydrates, g	264	272	273	290	**
Carbohydrates, E%	51.0	50.4	51.0	51.4	
Sucrose, g	42.9	45.9	45.0	49.5	
Sucrose, E%	8.2	8.4	8.3	8.4	
Alcohol, g	12.3	8.1	11.1	9.6	**
Diet change last 5 years	12%	51%	18%	69%	***

Test for difference between SHEEP and INTERGENE adjusted for age and sex, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.