Genetic Risk Factors for Type 2 Diabetes and Related Traits

Jens Kristoffer Hertel

Dissertation for the degree Philosophiae Doctor (PhD)

at the University of Bergen

2012

Contents

SCIENTIFIC ENVIRONMENT **5**

ACKNOWLEDGEMENTS **6**

ABBREVIATIONS **8**

GENE NAME ABBREVIATIONS **10**

LIST OF PUBLICATIONS **13**

1 INTRODUCTION **14**

1.1 DIABETES MELLITUS TYPE 2 AND RELATED TRAITS 15

1.1.1 Definition, description and classification of diabetes mellitus 15

1.1.2 Type 2 diabetes – etiology, pathophysiology and long-term complications 21

1.1.3 Type 2 diabetes-related traits 24

1.2 GENETIC MAPPING IN HUMAN DISEASES 26

1.2.1 Classification of genetic variants 26

1.2.2 Mapping of genetic variants underlying human traits 28

1.3 GENETIC PREDISPOSITION AND SUSCEPTIBILITY GENES FOR TYPE 2 DIABETES AND DIABETES RELATED TRAITS 34

1.3.1 Heritability of type 2 diabetes 34

1.3.2 Genetics of type 2 diabetes and intermediary phenotypes 35

1.3.3 Genetics of common obesity 46

2 AIMS **49**

3 RESEARCH DESIGN AND METHODS **50**

3.1 DEFINITION OF COHORTS 50

3.1.1 The Nord-Trøndelag Health Study 50

3.1.2 Malmö Diet and Cancer Cohort 51

3.1.3 Malmö Preventive Project 52

3.2 STUDY SAMPLES 53

3.3 GENETIC ANALYSES 55

3.3.1 SNP selection 55

3.3.2 Genotyping 57

3.3.3 Quality control 58

3.4 ETHICS 58

3.4.1 Ethical issues and approvals 58

3.4.2 Funding 59

4 MAIN RESULTS 61

4.1 PAPER I 61

4.2 PAPER II 62

4.3 PAPER III 63

4.4 PAPER IV 63

5 GENERAL DISCUSSION 65

5.1 REPLICATION OF GENOTYPE-PHENOTYPE ASSOCIATIONS (PAPER I) 65

5.2 LARGE-SCALE META-ANALYSES (PAPER II) 73

5.3 GENETIC FINE-MAPPING FOR IDENTIFICATION OF FURTHER PHENOTYPIC AND GENETIC COMPLEXITY (PAPER III) 82

5.4 DISCOVERY OF DISEASE RISK GENES BY DEFINING THE GENETIC BACKGROUND OF INTERMEDIARY PHENOTYPES (PAPER IV) 88

5.5 CONCLUDING REMARKS AND FUTURE PERSPECTIVES 94

REFERENCES 99

PAPERS I – IV

Scientific environment

Section for Pediatrics,
Department of Clinical Medicine,
University of Bergen,
Bergen, Norway

Center for Diabetes Genetics,
Department of Pediatrics,
Haukeland University Hospital,
Bergen, Norway

Center for Medical Genetics and Molecular Medicine,
Haukeland University Hospital,
Bergen, Norway

Department of Cardiology,
Haukeland University Hospital,
Bergen, Norway

Acknowledgements

This work was carried out at the Section for Pediatrics, Department of Clinical Medicine, University of Bergen, Bergen, Norway, and Center for Medical Genetics and Molecular Medicine at Haukeland University Hospital, Bergen, Norway. The work was mainly funded by the Research Council of Norway, the University of Bergen, Haukeland University Hospital, Helse Vest and Innovest. Additional funding from a Helse Vest-supported project (CARDIAC-GEN) led by Professor Ottar Nygård gave me the opportunity to accomplish this thesis.

First of all, I would like to extend my sincere gratitude towards my excellent principal supervisors Professor Pål Rasmus Njølstad and Professor Anders Molven, and my co-supervisor Associate Professor Stefan Johansson. They have given me the opportunity to work in the inspiring scientific environment at the Bergen Diabetes Research Group, and introduced me to an exciting and challenging field of research. Your faith in the project has been greatly appreciated, and your knowledge, optimism and dedication to science have inspired me enormously.

I would like to give a special thank to Stefan Johansson for excellent daily training in the field of genetics, and guidance through genetic analyses and statistical issues. Your endless support has been of outmost importance. Thank you for finding time even when you had none! Moreover, I want to thank all members of the Bergen Diabetes Research Group and my colleagues at Center for Medical Genetics and Molecular Medicine for being helpful, encouraging and for collectively creating a friendly atmosphere and high academic standard. Without the scientific and social support and fellowship from all of you, this thesis would not have been possible.

I have immensely appreciated the opportunity to work with both outstanding national and international collaborators. I would like to thank all coauthors on the papers on which this thesis is based for their help and constructive criticism: Helge Ræder, Ottar Nygård, Rolv Terje Lie, Kristian Midthjell, Carl Platou, Kristian Hveem, Emily

Sonestedt, Anna Jonsson, Peter M Nilsson, Gull Rukh, Olle Melander, Valeriya Lyssenko, Leif Groop and Marju Orho-Melander.

Sincere thanks go to my parents Anne Kari and Lasse, and my two brothers Jonas and Joakim for taking interest in my work, being supportive and for always believing in me. I also want to thank the rest of my family and my many, often temporarily ignored, friends.

At last, but not at least, I want to express my deepest gratitude to my dearest Beate and our two fantastic and energetic sons for their endless patience, understanding, love and support. Twenty-four hours a day, you keep me going!

Abbreviations

ADA	American Diabetes Association
ApoB	Apolipoprotein B
BMI	Body mass index
BP	Blood pressure
CD/CV	Common disease – common variant
CEU	U.S. residents with northern and western European ancestry
CHD	Coronary heart disease
CHR	Chromosome
CNV	Copy number variation
CVD	Cardiovascular disease
DCCT	Diabetes Control and Complication Trail
DGI	Diabetes Genetics Initiative
DIAGRAM	Diabetes Genetics Replication And Meta-analysis Consortium
DNA	Deoxyribonucleic acid
EWA	Environment-wide association
FG	Fasting glucose
FPG	Fasting plasma glucose
FUSION	Finland-United States Investigation of NIDDM Genetics
GAD	Glutamic acid decarboxylase
GDM	Gestational diabetes mellitus
GI	Gastrointestinal
GIANT	Genetic Investigation of Anthropometric Traits
GoKinD	Genetics of Kidneys in diabetes study
GWA	Genome-wide association
GWAS	Genome-wide association study
HbA1c	Glycated hemoglobin
HDL	High-density lipoprotein
HLA	Human leukocyte antigen
HOMA-IR	Homeostasis model assessment of insulin resistance
HUNT	The Nord-Trøndelag Health Survey
HUNT1	The first Nord-Trøndelag Health Survey
HUNT2	The second Nord-Trøndelag Health Survey
HWE	Hardy-Weinberg equilibrium
IDF	International Diabetes Federation
IFG	Impaired fasting glycemia
IGT	Impaired glucose tolerance

IR	Insulin resistance
IS	Insulin secretion
Kb	Kilobase
LADA	Latent autoimmune diabetes of adults
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
MAF	Minor allele frequency
MAGIC	The Meta-Analyses of Glucose and Insulin-related traits Consortium
MALDI-TOF	Matrix-assisted laser desorption/ionisation-time of flight
Mb	Megabase
MDC	Malmö Diet and Cancer Cohort
MI	Myocardial infarction
MODY	Maturity-onset diabetes of the young
MPP	Malmö Preventive Project
mRNA	Messenger RNA
NCBI	National Center for Biotechnology Information
NCI-NHGRI	National Cancer Institute and National Human Genome Research Institute
OGTT	Oral glucose tolerance test
OR	Odds ratio
PAR	Population attributable risk
PCOS	Polycystic ovary syndrome
PCR	Polymerase chain reaction
Q1-3	Questionnaires 1-3
QC	Quality control
RAF	Risk allele frequency
RNA	Ribonucleic acid
RR	Relative risk
SD	Standard deviation
SNP	Single-nucleotide polymorphism
T1D	Type 1 diabetes
T2D	Type 2 diabetes
Tg	Triglycerides
TZD	Thiazolidinediones
UKT2D	U.K. Type 2 Diabetes
UTR	Untranslated region
WHO	World Health Organization
WTCCC	Wellcome Trust Case Control Consortium

Gene name abbreviations

ABCB11	ATP-binding cassette, sub-family B (MDR/TAP), member 11
ABCC8	ATP-binding cassette, sub-family C (CFTR/MRP), member 8
ADAMTS9	ADAM metallopeptidase with thrombospondin type 1 motif, 9
ADCY5	Adenylate cyclase 5
ANK1	Ankyrin 1, erythrocytic
AP3S2	Adaptor-related protein complex 3, sigma 2 subunit
ARAP1	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain
ATP11A	ATPase, class VI, type 11A
BDNF	Brain-derived neurotrophic factor
BLK	B lymphoid tyrosine kinase
BNC2	Basonuclin 2
C2CD4A	C2 calcuim-dependent domain containing 4A
CDC123	Cell division cycle 123 homolog
CDKAL1	CDK5 regulatory subunit associated protein 1-like 1
CDKN2A	Cyclin-dependent kinase inhibitor 2A
CDKN2B	Cyclin-dependent kinase inhibitor 2B
CDKN2BAS	CDKN2B antisense RNA 1 (non-protein coding)
CEL	Carboxyl ester lipase
DGKB	Diacylglycerol kinase, beta 90kDa
DUSP9	Dual specificity phosphatase 9
FLJ39370 (C4ORF32)	Chromosome 4 open reading frame 32
FN3K	Fructosamine 3 kinase
FTO	Fat mass and obesity associated
G6PC2	Glucose-6-phosphatase, catalytic, 2
GCK	Glucokinase (hexokinase 4)
GCKR	Glucokinase (hexokinase 4) regulator
GCS	Goosecoid protein
GRB14	Growth factor receptor-bound protein 14
HFE	Henochromatosis
HHEX	Hematopoietically expressed homeobox
HK1	Hexokinase 1
HMG20A	High mobility group 20A
HMGA2	High mobility group AT-hook 2
HNF1A	HNF1 homeobox A
HNF1B	HNF1 homeobox B
HNF4A	Hepatocyte nuclear factor 4, alpha

IGF2BP2	Insulin-like growth factor 2 mRNA binding protein 2
INS	Insulin
IRS1	Insulin receptor substrate 1
JAZF1	JAZF zinc finger 1
KCNJ11	Potassium inwardly-rectifying channel, subfamily J, member 11
KCNQ1	Potassium voltage-gated channel, KQT-like subfamily, member 1
KLF11	Krüppel-like factor 11
KLF14	Krüppel-like factor 14
LARP6	La ribonucleoprotein domain family, member 6
MADD	MAP-kinase activating death domain
MC4R	Melanocortin 4 receptor
MITF	Microphthalmia-associated transcription factor
MTNR1B	Melatonin receptor 1B
NEGR1	Neuronal growth regulator 1
NEUROD1	Neuronal differentiation 1
NOTCH2	Notch 2
PAX4	Paired box 4
PCSK1	Proprotein convertase subtilisin/kexin type 1
PDX1	Pancreatic and duodenal homeobox 1
PKN2	Protein kinase N2
PPARG	Peroxisome proliferator-activated receptor gamma
PRC1	Protein regulator of cytokinesis 1
PROX1	Prospero homeobox 1
PTPRD	Protein tyrosine phosphatase, receptor type, D
RBMS1	RNA binding motif, single stranded interacting protein 1
SGSM2	Small G protein signaling modulator 2
SH2B1	SH2B adaptor protein 1
SLC30A8	Solute carrier family 30 (zinc transporter), member 8
SORCS1	Sortilin-related VPS10 domain containing receptor 1
SPTA1	Spectrin, alpha, erythrocytic 1 (elliptocytosis 2)
SPRY2	Sprouty homolog 2
SRR	Serine racemase
ST6GAL1	ST6 beta-galactosamide alpha-2,6-sialyltranferase
TCF7L2	Transcription factor 7-like 2 (T-cell specific, HMG-box)
THADA	Thyroid adenoma associated
TLE4	Transducin-like enhancer of split 4
TMPRSS6	Transmembrane protease, serine 6
TP53INP1	Tumor protein p53 inducible nuclear protein 1
TSPAN8	Tetraspanin 8

TUBGCP3	Tubulin, gamma complex associated protein 3
VPS13C	Vacuolar protein sorting 13 homolog C
VPS26A	Vacuolar protein sorting 26 homolog A
WDR72	WD repeat domain 72
WFS1	Wolfram syndrome 1 (wolframin)
ZFAND6	Zinc finger, AN1-type domain 6

List of publications

I. Hertel JK, Johansson S, Ræder H, Midthjell K, Lyssenko V, Groop L, Molven A, Njølstad PR: **Genetic analysis of recently identified type 2 diabetes loci in 1,638 unselected patients with type 2 diabetes and 1,858 control participants from a Norwegian population-based cohort (the HUNT study).** *Diabetologia* 51:971-977, 2008

II. Hertel JK, Johansson S, Sonestedt E, Jonsson A, Lie RT, Platou CG, Nilsson PM, Rukh G, Midthjell K, Hveem K, Melander O, Groop L, Lyssenko V, Molven A, Orho-Melander M, Njølstad PR: ***FTO*, Type 2 Diabetes, and Weight Gain Throughout Adult Life: A Meta-Analysis of 41,504 Subjects From the Scandinavian HUNT, MDC, and MPP Studies.** *Diabetes* 60:1637-1644, 2011

III. Hertel JK, Molven A, Ræder H, Platou CG, Midthjell K, Hveem K, Nygård O, Njølstad PR, Johansson S: **Genetic Fine-Mapping of Chromosome 9p21 - a Region Associated with Diabetes and Cardiovascular Disease in a Population-Based Sample (the HUNT2 Survey).** Revised manuscript submitted to PLoS One.

IV. Hertel JK, Johansson S, Ræder H, Platou CG, Midthjell K, Hveem K, Molven A, Njølstad PR: **Evaluation of four novel genetic variants affecting hemoglobin A1c levels in a population-based type 2 diabetes cohort (the HUNT2 study).** *BMC Med Genet* 12:20, 2011

"Implementation of current knowledge will bring some improvements to [non-communicable disease] care and prevention, but further research is essential if we are to truly defeat theses diseases"

- Jean Claude Mbanya,
President of IDF,
Lisbon, 2011.

1. INTRODUCTION

In complex diseases like type 2 diabetes, obesity and cardiovascular diseases, multiple genetic and environmental factors as well as the interaction between these factors determine the phenotype. The worldwide rise in prevalence of type 2 diabetes and other cardiometabolic disorders has led to an intense search for genetic factors influencing the susceptibility for these common disorders. Although environmental influences, such as high-caloric fat- and carbohydrate-enriched diets and a sedentary lifestyle with markedly reduced physical activity, certainly accelerate disease development in those with genetic predisposition, it is nonetheless of great clinical importance, and indeed a formidable challenge, to elucidate the genetic variants that increase the risk of diseases like type 2 diabetes [1]. Even though much research has been conducted, the knowledge of the specific causes of common complex diseases at the genetic level is still somewhat at its infancy. More detailed insight into the genetic risk factors and the underlying molecular mechanisms involved in type 2 diabetes and related traits is expected to improve clinical investigations, advance the prevention of disease development, elucidate the diseases mechanisms and hopefully highlight new pathways relevant for therapeutic intervention.

Thus, the general aim of this PhD project was to contribute to the progressing exploration of genetic risk factors in type 2 diabetes as well as of diabetes-related phenotypes like obesity and cardiovascular disease. The purpose of this first part of the thesis is to present a literature review of past and current findings in the dissection of the genetic background of type 2 diabetes and related traits. In particular, two loci on chromosome 16q12.2 and 9p21.3 (*FTO* and *CDKN2B*) that are shared by two or more conditions or traits (e.g. type 2 diabetes, obesity, cardiovascular disease) were

more comprehensively investigated in this project and are therefore discussed more thoroughly in the literature review.

1.1 DIABETES MELLITUS TYPE 2 AND RELATED TRAITS

1.1.1 Definition, description and classification of diabetes mellitus

Diabetes mellitus is a group of metabolic disorders of heterogeneous etiology characterized by persistent elevated blood glucose levels (hyperglycemia) with disturbances of carbohydrate, fat and protein metabolism as a result of defects in insulin secretion, impaired effectiveness of insulin action, or both [2, 3]. The disease is classified as type 1 diabetes, type 2 diabetes, gestational diabetes and other types of diabetes, including monogenic diabetes [2]. Type 1 and type 2 diabetes are considered the two major types. Type 1 diabetes normally develops before adulthood and is typically caused by an auto-immune destruction of the insulin-producing β-cells leading to an absolute insulin deficiency, whereas type 2 diabetes is normally associated with insulin resistance and relative insulin deficiency.

Diabetes is a major global health problem due to dramatically increasing prevalence in both the western world and in the developing countries. Rising health care costs are a serious problem, and a significant portion of health care spending is incurred by people with diabetes. The number of people with diabetes is increasing due to aging (increase in the proportion of people >65 years of age), general population growth, urbanization, and increasing prevalence of obesity and physical inactivity. The total number of people worldwide with diabetes is projected to rise from 285 million in 2010 to 439 million in 2030 corresponding to a predicted increase in prevalence from 6.4% in 2010 to 7.7% in 2030 [4]. Similar patterns are seen in Norway as well. Data from the Nord-Trøndelag Health surveys (HUNT) indicate an increase in the prevalence of diabetes during the last two decades, with 3.8% of women and 4.9% of men being diagnosed with diabetes in 2006-08 [5]. The prevalence of diabetes is probably underestimated due to the rapid rise in the number of obese individuals. In

Norway, studies have indicated that the total number of individuals with diabetes is twice of what has been diagnosed [6].

Diagnosing diabetes The diagnostic criteria for diabetes and pre-diabetes (intermediate hyperglycemia such as impaired fasting glucose (IFG) and impaired glucose tolerance (IGT)) have been debated for several years and modified numerous times. In 1997 the fasting glucose cut-off level was lowered from 7.8 to 7.0 mmol/l [3, 7] and in 2003 the American Diabetes Association (ADA) changed the threshold for IFG from 6.1 to 5.6 mmol/l [8]. Moreover, since 2010, ADA included the use of glycated hemoglobin (HbA1c) to diagnose diabetes and to identify individuals at "increased risk for future diabetes" [2].

Table 1 Present diagnostic criteria for diabetes, and non-diabetic hyperglycemia (IFG and IGT) according to serum/plasma levels. Adapted from [2, 9, 10].

	WHO 2006	**ADA 2011**
Diabetes Mellitus		
Fasting glucose	≥ 7.0 mmol/l	≥ 7.0 mmol/l
2-hour glucose	≥ 11.1 mmol/l	≥ 11.1 mmol/l
HbA1c	≥ 6.5 %[1]	≥ 6.5 %
Non-diabetic hyperglycemia		
Fasting glucose[2]	6.1 - 6.9 mmol/l	5.6 - 6.9 mmol/l
2-hour glucose[3]	7.8 – 11.0 mmol/l	7.8 - 11.0 mmol/l
HbA1c	-	5.7 - 6.4 %

[1]From a WHO consultation report from 2011 that was an addendum to the diagnostic criteria published in the 2006. [2]Impaired fasting glucose. [3]Impaired glucose tolerance.

HbA1c levels are better predictors than fasting glucose of the development of long-term complications in type 1 and type 2 diabetes [11]. In addition, higher levels in the sub-diabetic range have been shown to predict type 2 diabetes risk and cardiovascular disease [12, 13]. Thus, in a very recent report, the World Health Organization (WHO)

as well recommended the use of HbA1c in the diagnosis of diabetes [10]. The current diagnostic criteria for diabetes and intermediate hyperglycemia according to WHO and ADA are shown in Table 1.

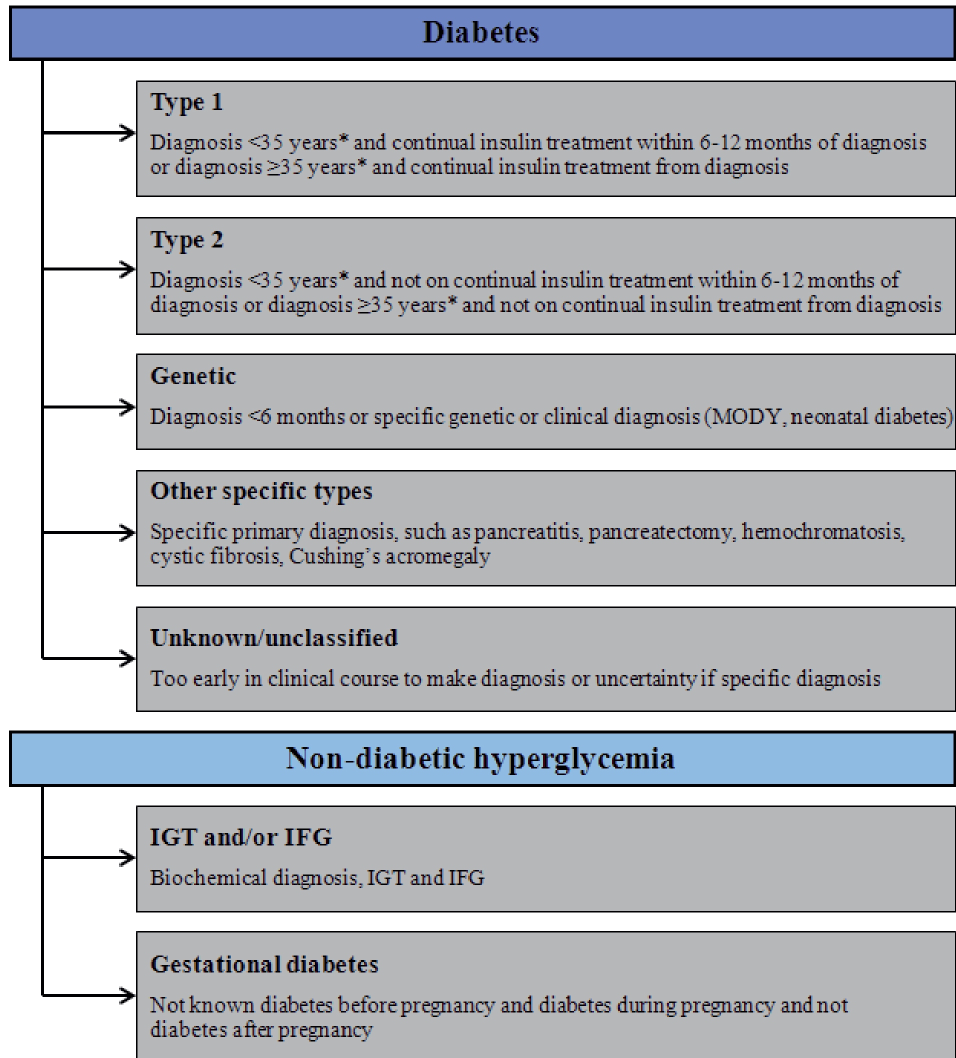

Figure 1 Disorders of glycemia: etiologic types and practical classification guidelines. *For some high risk ethnic groups a cut-off of 30 years should be used. Modified from [14].

Classification The traditional view is that type 1 and type 2 diabetes characterize distinct conditions with widely different pathophysiology. On the other hand, there is a notable increase in the proportion of people having diabetes with the characteristics of both types [15, 16]. Thus, it may not always be possible to assign each patient with diabetes to a precise diagnostic box (Figure 1). As an alternative, it has been suggested to think more in terms of a disease continuum, with the conventional autoimmune type 1 diabetes at one end and the conventional metabolic type 2 diabetes at the other [17, 18]. However, current practical classification guidelines still apply the traditional view that type 1 diabetes and type 2 diabetes are two different conditions attributable to entirely distinct, but as yet somewhat uncertain mechanisms.

Type 1 diabetes is caused by destruction of the insulin-producing cells of the pancreas, primarily due to an autoimmune-mediated reaction, resulting in absolute insulin deficiency. The reason why this occurs is not fully understood. In general, the disease is diagnosed at any age, but most frequently it develops during childhood and puberty. Type 1 diabetes is one of the most common endocrine and metabolic conditions in childhood and progresses rapidly. Latent autoimmune diabetes in the adult (LADA) is a slowly progressive form of autoimmune diabetes, characterized by diabetes-associated autoantibody positivity, and insulin independence at diagnosis, which distinguishes LADA from classic type 1 diabetes [19]. People with type 1 diabetes are usually insulin-dependent from diagnosis and require daily injections of insulin in order to control their blood glucose and to stay alive. Moreover, type 1 diabetes, especially the autoimmune process, is substantially determined by inherited variation [20]. There is a strong association between human leukocyte antigen (HLA) genes and type 1 diabetes. HLA variants confer either high risk of or protection against the disease. Currently, over 40 genetic loci have shown to affect risk of type 1 diabetes [21]. The incidence of type 1 diabetes is increasing, the reasons for which are unclear but may be due to changes in environmental risk factors that could initiate autoimmunity or accelerate already ongoing beta cell destruction.

Type 2 diabetes (see also chapter 1.1.2) is the most common form of diabetes and accounts for over 90 % of all diabetes cases worldwide [22]. Type 2 diabetes is characterized by insulin resistance and relative insulin deficiency, either of which may be present at the time that diabetes becomes clinically manifest. Type 2 diabetes usually occurs after the age of 35-40 years but may be diagnosed earlier, especially in populations with high diabetes prevalence. Type 2 diabetes can remain undetected (asymptomatic), for many years and the diagnosis is often made from associated complications or incidentally through an abnormal blood or urine glucose test. Type 2 diabetes is often, but not always, associated with metabolic abnormalities such as obesity, which itself can cause insulin resistance and lead to elevated blood glucose levels. In contrast to type 1 diabetes, people with type 2 diabetes are not absolutely dependent on exogenous insulin, but may require insulin for control of hyperglycemia if this is not achieved with diet alone or with oral hypoglycemic agents. Type 2 diabetes has a strong familial component, and at least 50 genetic variants have been reported to influence susceptibility to type 2 diabetes (see also chapter 1.3.1) [23].

Whereas type 2 diabetes is thought to be primarily heterogeneous and polygenic with low penetrance for the variants discovered, there exist monogenic types of non-autoimmune diabetes showing a Mendelian dominant pattern of inheritance, of which maturity-onset diabetes of the young (MODY) is the most common type [24]. Monogenic disorders of diabetes accounts for approximately 1-2% of all non-autoimmune diabetes and are largely affecting genes involved in β-cell development and function [25]. The onset of disease usually occurs in childhood or young adulthood, generally before 25 years of age, although the hyperglycemia is mild in some cases and may be missed, as with type 2 diabetes. When hyperglycemia is detected in children, MODY may be misdiagnosed as type 1 diabetes. Genetic studies have defined a number of subtypes of MODY. Mutations in the genes encoding hepatic nuclear factor 4 (*HNF4*), glucokinase (*GCK*), hepatic nuclear factor 1 alpha and 1 beta (*HNF1A* and *HNF1B*), pancreatic and duodenal homeobox 1 (*PDX1*), transcription factor neurogenic differentiation 1 (*NEUROD1*), krüppel-like factor 11 (*KLF11*), transcription factor paired box 4 (*PAX4*), carboxyl ester lipase (*CEL*),

insulin (*INS*) and B-lymphocyte specific tyrosine kinase (*BLK*) are the cause of the 11 known forms of MODY (MODY1-11) [26]. The most frequent forms of MODY results from mutations in the genes: *HNF1A*, *GCK*, *HNF4A* and *HNF1B* [25, 27, 28]. Other monogenic forms of diabetes include mitochondrial diabetes, neonatal diabetes, syndromes of severe insulin resistance and rare genetic syndromes. There are five genes currently known to be associated with non-syndromic permanent neonatal diabetes: potassium channel, inwardly rectifying, subfamily J, member 11 (*KCNJ11*), ATP-binding cassette, subfamily C, member 8 (*ABCC8*), *INS*, *GCK*, and pancreas/duodenum homeobox protein 1 (*PDX1*) [28-30]. Genetic testing and counseling is indicated and highly relevant when monogenic forms of diabetes are suspected, since patients with mutations in *KCNJ11*, *ABCC8*, *HNF1A* and *HNF4A* can be treated with oral antidiabetic agents (sulphonylureas) [31, 32], in contrast to most of those who have mutations in the other genes. Prognosis, treatment and complications may also vary between the various forms of monogenic diabetes, depending on which gene that is affected. The predictive and clinical value of genetic testing is therefore substantial for monogenic forms of diabetes [28].

Gestational diabetes mellitus (GDM) represents glucose intolerance of varying degrees of severity with onset or first detection during pregnancy [2]. Most cases resolve after delivery. Maternal hyperglycemia may lead to complications in the baby, including large size at birth, birth trauma, hypoglycemia and infant respiratory syndrome [33, 34]. Women who have had GDM have an increased risk of developing diabetes later in life [35]. Moreover, children of women with GDM have an increased risk for childhood and adult obesity and an increased risk of glucose intolerance [36].

In addition, there are other specific types of diabetes with a known etiology, such as secondary to other diseases, trauma or surgery, or the effects of drugs. Examples include diabetes caused by hemochromatosis, exocrine pancreatic disease, or certain types of medications (e.g., long-term steroid use) [2].

1.1.2 Type 2 Diabetes – etiology, pathophysiology and long-term complications

The etiology of type 2 diabetes is not fully understood, but presumably, type 2 diabetes develops when a diabetogenic lifestyle (i.e. excessive caloric intake, inadequate caloric expenditure, obesity) acts in conjugation with a susceptible genotype. The majority of patients who develop type 2 diabetes are obese [37]. Energy-dense diet as a risk factor has, however, shown to be independent of baseline obesity for the development of type 2 diabetes [38]. Further, it has been suggested that type 2 diabetes in some cases are caused by environmental pollutants [39]. Even though there is some disparity regarding the reasons for the development of type 2 diabetes, most physicians and scientists agree that the major independent risk factors for developing type 2 diabetes are: obesity [40, 41], family history of type 2 diabetes (first-degree relative) [42], ethnicity (some ethnic groups have higher prevalence of diabetes) [43], history of previous IGT or IFG [44], hypertension or dyslipidemia [45, 46], physical inactivity [47], history of gestational diabetes [48], low birth weight as a result of an *in utero* environment [49], polycystic ovarian syndrome leading to insulin resistance [50], and finally, decline in insulin secretion due to advancing age [51, 52]. Until recently, type 2 diabetes was considered to be a disease confined to adulthood, rarely observed in individuals under the age of 40, but clinically based reports and regional studies suggest that type 2 diabetes in children and adolescents is now more frequently being diagnosed [53]. This reflects the increasing number of children entering adulthood with unprecedented levels of obesity.

Type 2 diabetes is primarily caused by obesity, insulin resistance in liver, skeletal muscle and adipose tissue and a relative insulin secretion defect by the pancreatic β-cell (3,4). Insulin is a hormone produced by the pancreatic β-cells and is the key hormone for the regulation of blood glucose. The hormone stimulates uptake of glucose from the blood in the muscle and fat tissue, storage of glucose as glycogen in the liver and muscle cells, and uptake and esterification of fatty acids in adipocytes. In addition, insulin inhibits the breakdown of proteins, the hydrolysis of triglycerides and the production of glucose from amino acids, lactate and glycerol. Glucagon,

which is also secreted by the endocrine pancreas, has the opposite effects to that of insulin. The hormone causes the liver to convert stored glycogen into glucose, thereby increasing blood glucose. Besides, glucagon stimulates insulin secretion, so that glucose can be used by insulin-dependent tissues. Hence, glucagon and insulin are part of a feedback system that keeps blood glucose at the right level (Figure 2).

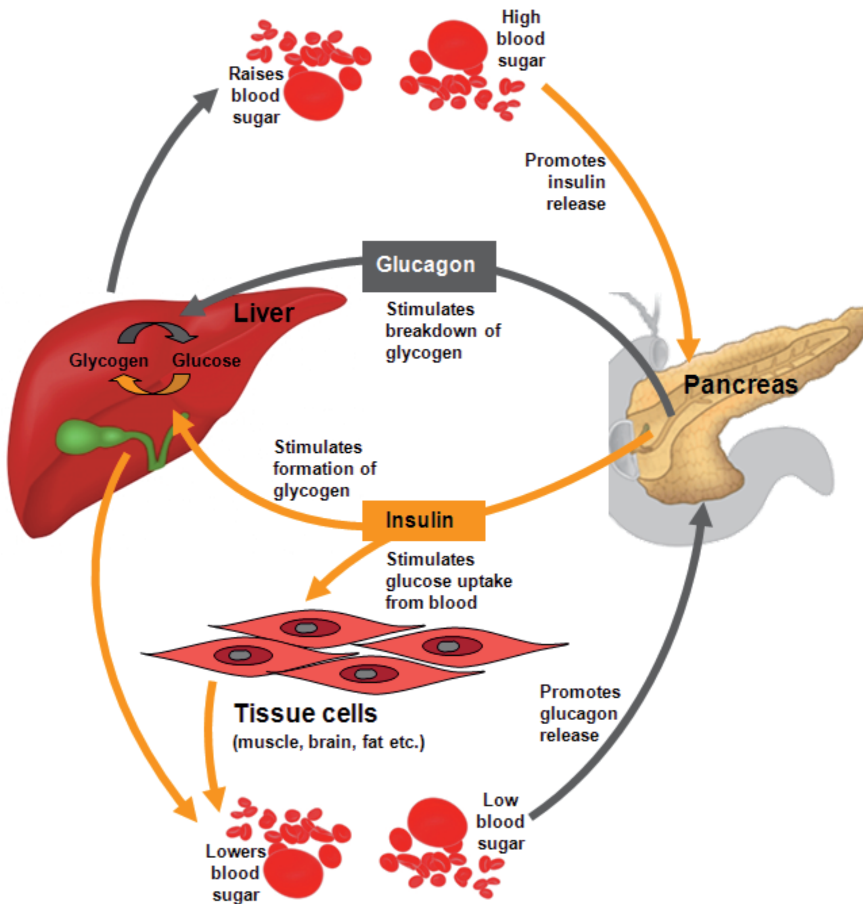

Figure 2 Insulin production and action. Diabetes results from an imbalance between the insulin-producing capacity of the pancreatic β-cells and the requirement for insulin action in insulin target tissues such as liver, adipose tissue and skeletal muscle. Redrawn and modified after the IDF Diabetes Atlas [54].

For type 2 diabetes to occur the balanced relationship between insulin action and release have to be disrupted. In other words, type 2 diabetes develops mainly in those who cannot increase insulin secretion sufficiently to compensate for their insulin resistance. Whereas insulin resistance is an early phenomenon partly related to obesity, pancreas β-cell function declines gradually over time already before the onset of clinical hyperglycemia. Several mechanisms have been proposed for these two defects. Insulin resistance have been ascribed to elevated levels of free fatty acids [55], inflammatory cytokines [56], adipokines [57] and mitochondrial dysfunction [58], while glucose toxicity [59], lipotoxicity [60], and amyloid formation [61] have been proposed as central aspects for β-cell dysfunction (all reviewed in [62]).

The medical and socioeconomic burden of type 2 diabetes is generally caused by the associated complications of the disease. The severe complications accompanying type 2 diabetes are mostly microvascular (e.g. retinopathy, neuropathy and nephropathy) and macrovascular diseases, leading to reduced quality of life and increased morbidity and mortality from end-stage renal failure and cardiovascular disease (CVD). Hyperglycemia plays a central role in the development and progression of the vascular complications, which often persist and progress despite improved glucose control, possibly as a result of prior occurrences of hyperglycemia. Increased cardiovascular risk, however, appears to begin before the development of frank hyperglycemia, presumably because of the effects of insulin resistance. This phenomenon has been described as the "ticking clock" hypothesis of complications [63, 64], where the clock starts ticking for microvascular risk at the onset of hyperglycemia, and for macrovascular risk at some antecedent point, i.e. with the onset of insulin resistance.

It is generally accepted that the long-term complications of diabetes mellitus are far less common and less severe in people who have well-controlled blood sugar levels [65, 66]. However, some recent trails that had great success in lowering blood sugar in type 2 diabetes patients, but no success in reducing deaths from cardiovascular disease, challenges the theory of hyperglycemia as the major cause of diabetic

complications [67]. The familial clustering of the degree and type of diabetic complications indicates that genetics may also play a role in causing diabetic complications [68]. Although not fully understood, the complex mechanisms by which diabetes leads to these complications involves hyperglycemia and both functional and structural abnormalities of small blood vessels along with accelerating factors such as smoking, elevated cholesterol levels, obesity, high blood pressure and lack of regular exercise.

1.1.3 Type 2 diabetes-related traits

Type 2 diabetes clusters with dyslipidemia and hypertension, which together with insulin resistance and glucose intolerance are hallmarks of the metabolic syndrome [69]. The prevalence of these metabolic abnormalities together with the rapid increase in overweight and obesity are leading to a rise in the global cardiometabolic risk, followed by a relentless rise in premature deaths due the increased risk of diseases such as type 2 diabetes, cardiovascular disease and cancer. Cardiometabolic diseases are currently considered as one of the major health and social challenges in the upcoming years.

A wide range of factors are crucial for the development of cardiometabolic disease, some with greater effect than others, but generally they are all considered important. High blood pressure, old age, family history, smoking, elevated levels of glucose and LDL-cholesterol have for a long time been regarded as the classical risk factors. Obesity (in particular abdominal), inflammation, insulin resistance, elevated levels of triglycerides and decreased levels of HDL cholesterol have in recent times become more highlighted. The basis for all the risk factors is an interaction between genetic and environmental influences. Type 2 diabetes is in itself a major risk factor for cardiovascular disease (Figure 3). Cardiovascular morbidity in patients with type 2 diabetes is two to four times greater than that of non-diabetic people [70]. Additionally, patients with type 2 diabetes experience an increased frequency of non-fatal heart attack and stroke. In light of this, it is of great clinical and biological

interest to find the genetic variants that cause increased risk of cardiometabolic disease.

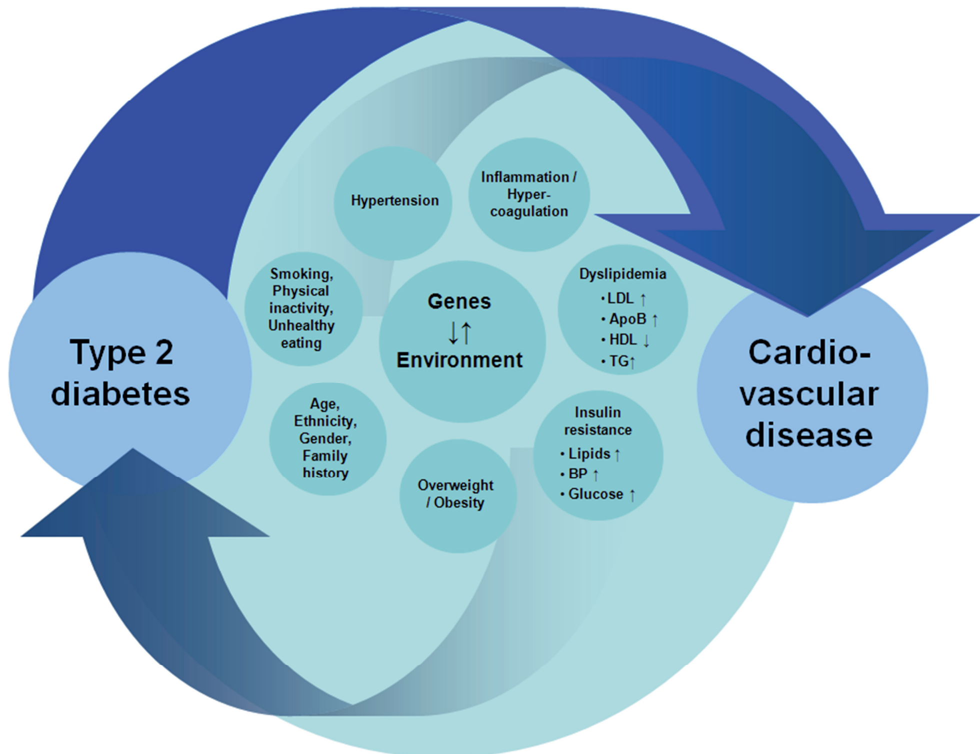

Figure 3 Cardiometabolic risk factors. Obesity, insulin resistance, dysglycemia, dyslipidemia, and hypertension frequently cluster and are major risk factors for both type 2 diabetes and cardiovascular disease (CVD). The impact of these risk factors is amplified by lifestyle (physical inactivity, smoking, and diet) as well as by genetics, gender, and age. Type 2 diabetes alone is a major risk factor for CVD. Abbreviations; ApoB: apolipoprotein B; BP: blood pressure; HDL: high-density lipoprotein cholesterol; LDL: low-density lipoprotein cholesterol; TG: triglycerides.

1.2 GENETIC MAPPING IN HUMAN DISEASES

Genetic mapping is a powerful approach used for identification of genes underlying any trait influenced by inheritance, including human diseases. The methodology is based on the correlation between trait and DNA variation and is carried out without the need for prior hypotheses about biological function [71]. Ever since the re-discovery of Mendel's laws of inheritance in the early 1900s and the subsequent awareness that most naturally occurring phenotype variation involves the action of multiple genes and non-genetic factors, geneticists have searched for practical tools for discovering genes contributing to human diseases. Human genetic variation was termed "breakthrough of the year" by *Science* in 2007 [72], reflecting the recent years' striking progress in understanding the genetic basis underlying normal human phenotypic variation and susceptibility to a wide range of diseases [73].

1.2.1 Classification of genetic variants

Genetic information is contained in the form of DNA. The basic complement of DNA in an organism is called the genome. The human genome is packed in two sets of 23 chromosomes; one set inherited from each parent whose own DNA is a mosaic of preceding ancestors. Consequently, the human genome functions as a diploid unit with phenotypes arising due to the complex interplay of alleles of genes and/or their non-coding functional regulatory elements [73]. The haploid human genome consists of approximately 3 billion nucleotides, in each cell. Among two random individuals the genomes vary by approximately 0.5% [74]. This variation affects the majority of human phenotypic differences, from eye color and height to disease susceptibility and responses to drugs [73].

Phenotypic diversity is determined by genetic variation acting in conjugation with environmental and behavioral factors. The genetic variants are classified by two basic criteria: their frequency in the population and their composition – i.e. sequence variants or structural variants. Sequence variation varies from single nucleotide variants to 1 kilo-base (kb) insertions or deletions (indels) of DNA segments.

Structural variation is a common designation for larger insertions and deletions, as well as duplications, inversions and translocations, ranging in size from 1 kb to more than 5 mega-bases (Mb) (Figure 4). If a DNA segment is present in variable numbers compared to the reference sequence, as in duplications, deletions or insertions, it is termed a copy number variant (CNVs) [73, 75, 76].

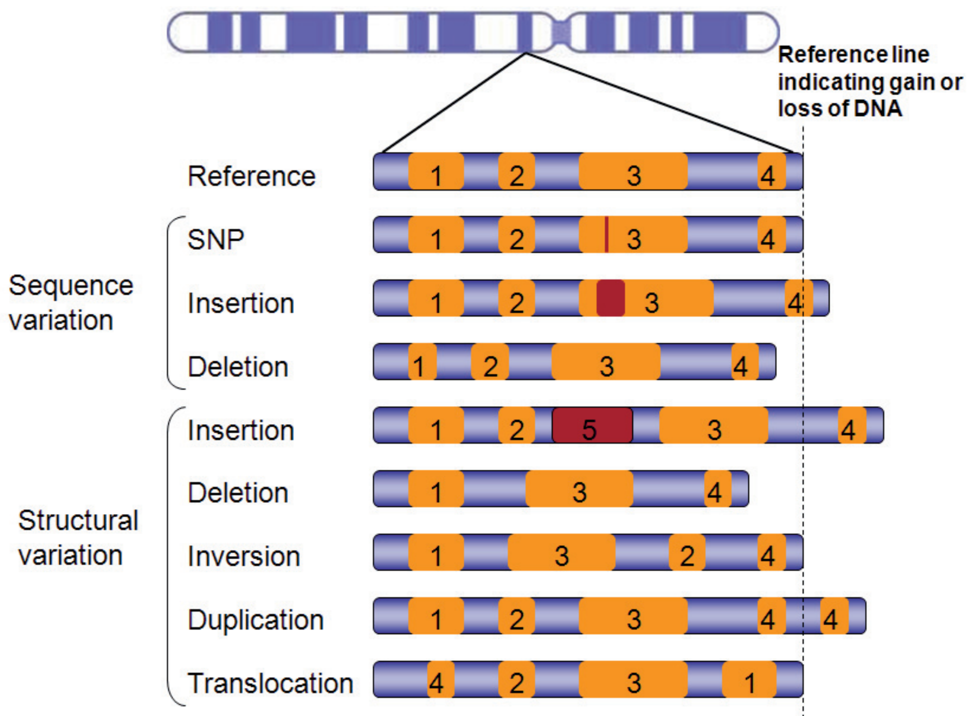

Figure 4 Classification of genetic variants by composition, showing examples of sequence variation and structural variation compared to a reference sequence. Modified from [73, 77]

According to their frequency, genetic variants are referred to as common if their minor allele frequency (MAF) is >5% in the population, while rare variants are present at a frequency <5%. A polymorphism is, in principle, defined as a genetic variant that is present in ≥1% of the population. Thus, a single-nucleotide variant showing a frequency >1% is consequently termed single nucleotide polymorphism (SNP) [73]. It is estimated that the human genome harbors approximately 10 million

SNPs comprising 78% of the human variants. In contrast, structural variants are estimated to account for no more than 22% of all variants, but comprise an estimated 74% of the nucleotides that may differ from person to person [74].

1.2.2 Mapping of genetic variants underlying human traits

Mapping of genetic variation underlying human traits depends on two key concepts: genetic linkage and linkage disequilibrium. Genetic linkage is the phenomenon where recombination between two loci occurs with less than 50% probability in a single generation; resulting in co-segregation more often than if they were independently inherited. In other words, genetic linkage is the tendency of certain alleles to be inherited together. Genetic loci that are physically close to one another on the same chromosome tend to stay together during meiosis, and are thus genetically linked [73].

The second concept, linkage disequilibrium (LD), is a measure of association between alleles at separate but linked loci, usually resulting from a particular ancestral chromosomal segment (haplotype) being common in the population studied. This phenomenon causes polymorphisms to be correlated to the point of being strong proxies for each other [78]. Different statistics have been used to measure the amount of linkage disequilibrium between two variant alleles, one of the most common being the coefficient of correlation r^2 [79]. When $r^2 = 1$, the two variant alleles are in complete linkage disequilibrium, whereas $r^2 < 1$ indicate that the ancestral complete linkage disequilibrium has been eroded. Due to this phenomenon of LD, it is possible to choose a subset of highly informative SNPs, or "tag" SNPs, to represent certain haplotypes, and the number of SNPs to be genotyped in a larger sample can therefore be reduced without losing the ability to capture most of the variation. For example, it is possible to select a set of 300,000 to one million SNPs to represent most of the 10 million common SNPs estimated to be present in the human genome [78].

Because the causal SNP is often not typed within a genetic association study, it is important to cross-examine SNPs that have not been genotyped directly. This can be

done through imputation. Imputation methods predict the alleles of SNPs not directly genotyped in the study (or hidden SNPs) using the correlation structure (LD) between the SNPs in the region. The starting point of any imputation method is a reference data set for which the genotypes of a dense set of SNPs are provided, such as HapMap. The fundamental assumption is that the reference samples, the cases, and the controls are all sampled from the same population. Under this simplifying assumption, the three populations share the same LD structure. Thus, the structure of the LD in the reference population, in combination with the structure of the LD of the observed SNPs within the cases and the controls, may be used to impute the alleles of a hidden SNP [80].

Several different approaches have been used in the exploration of genetic factors involved in complex disease. The progress has generally been guided by technological advances in genotyping and sequencing techniques, statistical handling of data and also by collection of larger cohorts suitable for genetic studies. In general, two methods have been used for studying genetic factors involved in human diseases in the 20th century: the so-called candidate gene approach and the linkage analysis approach [81-83]. The latter is not based on assumptions, but identifies genes through their genomic position and is based on the rationale that family members sharing a specific phenotype will also share chromosomal regions surrounding the gene involved. The linkage approach has proved very effective in the identification of rare variants with a high degree of penetrance, such as those responsible for extreme forms of early-onset diseases segregating as monogenic (Mendelian) disorders – including MODY, mitochondrial diabetes with deafness, neonatal diabetes and rare forms of severe childhood obesity [84-86]. However, because the risk for relatives is lower in complex diseases due to the low penetrance of polygenic risk variants, the statistical power of this method in studies of polygenic traits is limited [87]. Even for loci with considerable effects on susceptibility at the population level, the number of families needed to offer sound power to detect linkage has proven hard to obtain [88]. Very few variants with large phenotypic effect (high-impact risk alleles) appear to be present in common complex diseases, thus most linkage studies have, in retrospect,

been seriously underpowered. This could also explain the inadequate findings and the lack of replication of regions putatively linked to disease. Moreover, even when evidence of linkage is observed, the genomic region linked to the trait of interest is often very large; hence the identification of the causal gene or genetic variant often remained the main challenge.

The candidate gene approach examines specific genes with a plausible role in the disease process. For diabetes, natural candidates are genes involved in glucose homeostasis and metabolism. The approach is biased since it assumes that a specific gene or loci is associated with disease before testing. The genetic variants are identified through focused sequencing and further assessed by genotyping them in a large number of cases and controls. Even though this approach has contributed to the identification of numerous published associations, only a fraction of the associations have been replicated by other studies [89, 90]. There are several reasons for non-replications: a) lack of statistical power in follow-up studies to detect or exclude a previously reported finding, b) false positive findings in the initial report due to incomplete or no correction for multiple testing, c) spurious associations as a consequence of population stratification or by random, d) differences in allele frequencies or LD between the genetic variants in the populations studied, e) differences in selection and phenotypic characteristics of the study participants/cases and controls or, f) unmeasured population-specific environmental exposures that may confound the association [81, 91, 92].

Thus, methods used successfully to identify the genes underlying rare Mendelian diseases generally failed in the identification of the genetic basis of common disorders such as cancer, diabetes and heart disease. This suggested that most of the genetic contribution to complex diseases arises from multiple loci with individually small effects (Figure 5). The conceptual outline for association studies to identify common genetic variants underlying common complex diseases was first reported by Risch and colleagues in 1996 [93], and is now referred to as the common disease/common variant (CD/CV) hypothesis. The major assumption behind the

CD/CV hypothesis is that since the major diseases are common, so are the genetic variants that cause them. Moreover, common variants with low penetrance and modest risk are not subjected to the same negative selection as variants with strong phenotypic effect causing Mendelian diseases. Hence, the hypothesis states that common diseases are caused by multiple, high frequency genetic variants conferring cumulative incremental effects on disease risk [73, 94]. With these assumptions as a fundament, the next challenge became clear - to survey the common genetic variation in the genomes of a large number of individuals. This would be necessary in order to reveal the intricate genetic background of common complex diseases.

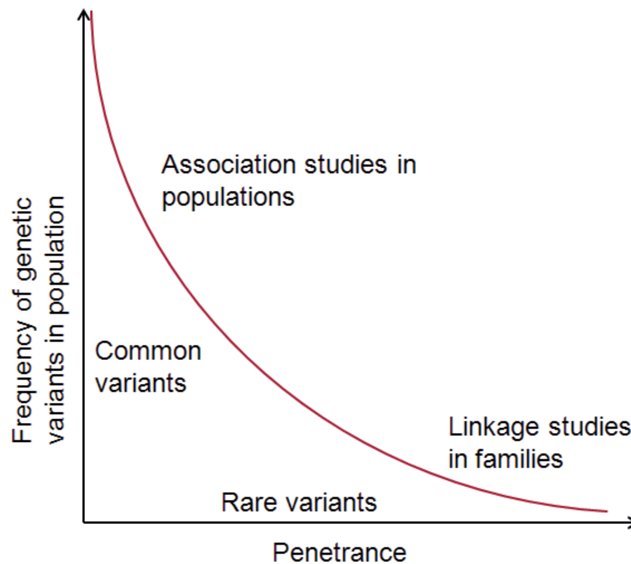

Figure 5 The allelic spectrum of disease – its all about number, frequency and penetrance. The allelic spectrum of disease relies on the number of genetic variants, their frequency in a population and on the penetrance (size of their phenotypic effect). Linkage studies have proved successful in identifying genetic variants causing rare Mendelian disorders, those with low-frequency and high penetrance. Complex diseases are believed to be caused by multiple genetic variants each conferring only low to modest risk for disease. Adapted from [73, 95].

The breakthrough came in 2006-2007 with the successful implementation of genome-wide association studies (GWAS). This new approach became possible as a result of the completion of the human genome sequence in 2001 [96, 97], the creation of SNP LD maps by the International HapMap Project [98] and great advancements in genotyping technology (efficient gene-chips) and tools for statistical handling [71]. Using SNP-based arrays and comparing the frequency of SNP alleles between cases and controls, the GWA approach allowed the investigators to detect genetic variants with modest phenotypic effects in a systematic and unbiased manner, provided that the variants had a high frequency in the population. These studies required large numbers of patients and cost several million dollars each. Due to the vast amount of genetic variants analyzed in a GWA study, a high number of statistical tests are performed, thus leading to a substantial risk of false positives owing to multiple testing. The important need for controlling this problem has resulted in the general use of a more stringent genome-wide significance level before an association is considered statistically significant. Current consensus has, based on a simulation study, defined a genome-wide significance level of $P < 5\times10^{-8}$ to account for 10^6 independent genome-wide hypotheses tested in a dense GWA (44), even though also $P < 10^{-7}$ has been suggested (186,187).

Approximately 951 GWA studies covering over two hundred distinct diseases and traits have been published by the second quarter of 2011, with nearly 1,450 SNP–trait associations reported as significant ($P < 5\times10^{-8}$) (Figure 6) [99, 100]. The upshot is that hundreds of common genetic variants have now been statistically linked with various diseases. Such associations are consistent with the common disease–common variant hypothesis, which posits that common diseases are attributable in part to allelic variants present in more than 1–5% of the population [94, 101, 102]. Hence, genome-wide association studies have, without doubt, provided valuable insights into the genetic architecture of common complex disorders. However, most variants identified so far confer relatively small increments in risk, and explain only a small proportion of familial clustering, thus leading to question of how the remaining ”missing heritability” can be explained. Possible sources of the missing heritability

and future research strategies, including and extending beyond current genome-wide association approaches, will be discussed in more detail in Chapter 5.

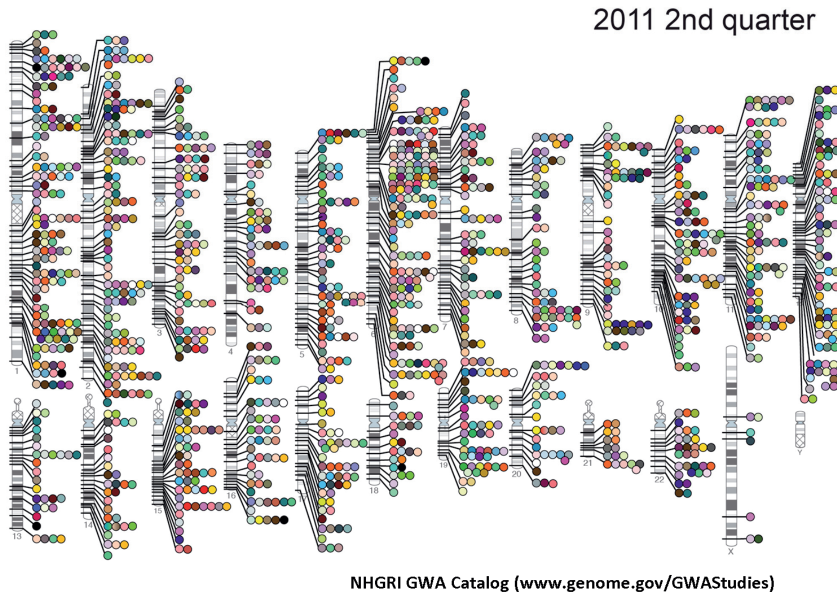

Figure 6 Published genome-wide associations reported as of June 2011. The circles indicate the chromosomal location of 1,449 published GWA at $p \leq 5x10^{-8}$ for 237 traits. Each disease type or trait is coded by color. From the National Human Genome Research Institute [99].

1.3 GENETIC PREDISPOSITION AND SUSCEPTIBILITY GENES FOR TYPE 2 DIABETES AND DIABETES RELATED TRAITS

1.3.1 Heritability of type 2 diabetes

Phenotypic variation among individuals may be attributable to genetics, environmental challenge and/or random events. Heritability is the proportion of phenotypic variation in a population that is due to genetic variation between individuals. Heritability of a trait or condition is often estimated on the basis of parent-offspring correlations for continuous traits or the ratio of the incidence in first-degree relatives of affected persons to the incidence in first-degree relatives of unaffected persons. Heritability is also frequently estimated by comparing resemblances between twins.

The clinical assessment of type 2 diabetes has often incorporated genetic information in the form of family history. Although very simple, family information has helped to raise clinical awareness for an individual patient's risk of type 2 diabetes due to the strong heritability of this disease. In contrast to a population risk of ~7%, family studies have estimated that the risk for type 2 diabetes among offspring is, respectively, 3.5-fold and 6-fold higher for those with a single diabetic parent and two diabetic parents compared with offspring without parental diabetes [103]. Furthermore, the higher concordance rate of type 2 diabetes in monozygotic versus dizygotic twins and the high prevalence of type 2 diabetes in specific ethnic groups such as Pima Indians and Mexican Americans, all lend support to the existence of genetic determinants for type 2 diabetes [78]. Overall, estimates have shown that 30%–70% of type 2 diabetes risk can be ascribed to genetics [104]. It is also evident, for example from a recent study in Finnish families, that type 2 diabetes-related intermediate and quantitative traits show substantial heritability [105]. The patterns of inheritance therefore suggest that type 2 diabetes and its related traits are both polygenic and heterogeneous; hence multiple genes are involved and different combinations of genes play a role in different subsets of individuals. How many risk

genes that exist and what their relative contributions are, remains somewhat uncertain. However, recent advances in genetic mapping of complex diseases have provided some information or at least great optimism in the dissection of the complex architecture of polygenic diseases such as type 2 diabetes.

1.3.2 Genetics of type 2 diabetes and intermediary phenotypes

In the past 10-15 years, huge resources have been devoted to finding type 2 diabetes genes. These efforts have included many candidate-gene studies and extensive efforts to fine-map linkage signals. Linkage analysis and subsequent positional fine-mapping of candidates have been mostly inconclusive, despite the detection of multiple genomic regions putatively linked to diabetes [106]. There is one notable exception, namely transcription factor 7-like 2 gene (*TCF7L2*). In 2006, the Icelandic company deCODE Genetics identified a common type 2 diabetes susceptibility variant in the *TCF7L2* gene region [107]. This result was interesting for two reasons. First, the variants that were found to alter risk did not explain the linkage signal, even though the investigators analyzed more than 200 markers across the region. This suggested that a non-candidate-gene or region-based association approaches, such as a GWAS, could have a great potential. Second, *TCF7L2* was a completely unexpected gene, thus demonstrating that a genome-wide approach could uncover previously unknown disease pathways [108].

Variants in many candidate genes were extensively studied by association studies in the pre-GWA era. In most instances, however, the initial association was not replicated in subsequent analyses. The candidate gene studies produced more unequivocal evidence for common variants involved in type 2 diabetes than did the linkage approach. The most robust candidate variants were the E23K variant in the *KCNJ11* gene [109-111], the P12A variant in the peroxisome proliferator-activated receptor-γ (*PPARG*) gene [112], and common variation in the *HNF1B* and the Wolfram syndrome 1 (*WFS1)* genes [113-115]. Rare mutations in all of these four genes are causing monogenic forms of diabetes [116-119], and two are targets of anti-

diabetic therapies. *KCNJ11* encodes a component of a potassium channel with a key role in β-cell physiology that is a target for the sulphonylurea class of drugs, and *PPARG* encodes a transcription factor involved in adipocyte differentiation that is a target for the thiazolodinedione class of drugs [78, 108].

In the spring of 2007, the results from the first wave of GWA studies investigating type 2 diabetes genes were published, namely the French, deCODE, DGI, WTCCC and FUSION studies [120-124]. These five independent GWA studies were all conducted using a two-stage strategy consisting of a GWA screen in an initial cohort of unrelated cases and controls followed by replication of the most significant findings in additional patients series. The initial GWAS were subsequently followed by five smaller GWA studies [125-129]. The screening and replication sets consisted primarily of European Whites, with the exception of the deCODE study which contained groups of Chinese and West Africans. Each of these early GWA studies of type 2 diabetes identified numerous potential susceptibility variants, but no less than nine loci emerged as being consistently associated with risk of type 2 diabetes across multiple studies. The nine loci were *TCF7L2*, solute carrier family 30, member 8 (*SLC30A8*), hematopoietically expressed homeobox (*HHEX*), CDK5 regulatory subunit-associated protein 1-like 1 (*CDKAL1*), cyclin-dependent kinase inhibitor 2A/2B (*CDKN2A/B*), insulin-like growth factor 2 mRNA-binding protein 2 (*IGF2BP2*), fat mass- and obesity-associated gene (*FTO*), *KCNJ11* and *PPARG*, among which three (*TCF7L2*, *KCNJ11* and *PPARG*) had previously been implicated in type 2 diabetes.

The *TCF7L2* gene is the most important type 2 diabetes susceptibility gene found to date [107]. Since its discovery, the association has been replicated in a variety of studies in subjects of different ethnicities [130-141]. In the U.K. population, the allelic odds ratio (OR) for the lead SNP (rs7903146, risk-allele frequency = 30%) is 1.36 and individuals carrying two risk (T) alleles are at nearly twice the risk of type 2 diabetes as are those with none [134]. The population attributable risk (PAR) is on the other hand somewhat lower, and varies with the variants' frequency in the

population. *TCF7L2* encodes a transcription factor in the Wnt-signaling pathway, which induces transcription of a number of genes, including proglucagon, in the intestine. Recent studies have shown that there is an increased expression of *TCF7L2* in the islets of pancreas in type 2 diabetes, which in turn results in impaired glucose-stimulated insulin secretion [142].

One of the most interesting regions to emerge from the first wave of GWAS for type 2 diabetes and CVD lies in a gene desert ~130 kb upstream of the *CDKN2B* gene on chromosome 9p21. Several SNPs in the 9p21 interval have demonstrated strong associations with coronary artery disease/myocardial infarction (MI) [143-146] and other vascular diseases such as stroke and intracranial and abdominal aneurisms [147-149]. All these SNPs are highly correlated (r^2 >0.8) and found in a ~60 kb LD-block. The 9p21 region also contains two adjacent, but distinct type 2 diabetes signals separated by a recombination hotspot; a strong signal mapped to an 11 kb LD-block (represented by rs10811661) and a second signal (rs564398) located ~100 kb in a telomeric direction from the type 2 diabetes-associated interval [120, 121, 124]. After the initial GWASs, several studies have confirmed the association with the implicated candidate SNPs in type 2 diabetes [150-153] and CVD [154-160] and extended the number of CVD phenotypes associated with the region [161-165]. This raised the possibility of a shared genetic or mechanistic link causing both CVD and diabetes within this region. In support, a significant interaction was found between poor glycemic control and a variant within the 9p21 region on the risk of coronary heart disease in patients with type 2 diabetes [166]. However, the effects of the disease susceptibility variants for the two major disease loci have shown to be independent, since type 2 diabetes risk variants do not seem to confer increased risk of cardiovascular disease or the other way around [147, 167].

The risk variants identified in the 9p21 interval by GWAS are in general located in non-coding regions, since most reported risk variants do not appear in mature transcripts, and there are no known micro-RNAs mapping to this region [168]. This suggests that their effects probably are mediated by influences on gene expression of

nearby genes in cis. Besides the coding sequences for the two cyclin-dependent kinase inhibitors, *CDKN2A* (*p16*^{INK4a}) including its alternative reading frame (*ARF*) transcript variant (*p14*^{ARF}), and *CDKN2B* (*p15*^{INK4b}), the region contains a large antisense non-coding RNA gene, designated *CDKN2BAS* (formerly termed *ANRIL*). Recent studies have shown that expression of these genes is co-regulated and that most of the confirmed risk variants are all correlated with *CDKN2BAS* expression, indicating that *CDKN2BAS* could play a role in *CDKN2B* regulation [168]. Hence, modulation of *CDKN2BAS* expression may mediate susceptibility to several important diseases.

The individual SNP rs10757278 has been highlighted as a potential causal variant for the association with coronary artery disease based on effects on expression of the *INK4/ARF* locus (*p15*^{INK4b}, *p16*^{INK4a}, *ARF* and *CDKN2BAS*) [169]. Moreover, the rs10757278 SNP have also been mapped to one of 33 newly identified enhancers in the 9p21 interval, in which the risk variant disrupts a transcription factor binding site, thus having functional relevance for an atherosclerosis-associated pathway in human endothelial cells [170].

The French GWA study, one of the first five GWA studies investigating type 2 diabetes genes, involved non-obese diabetics and revealed that a version of a gene encoding a protein that transports zinc in the pancreas, *SLC30A8*, increased the risk of type 2 diabetes [122]. Of all the new type 2 diabetes genes discovered by the GWA approach, *SLC30A8* are one of the few involving a non-synonymous polymorphism – an arginine to tryptophan substitution at amino acid 325. *SLC30A8* has also recently been identified as an auto-antigen in human type 1 diabetes [171]. In contrast to *SLC30A8*, most of the genes identified in the GWA screens would not be considered typical candidate genes for type 2 diabetes and in most cases the variants are located in non-coding regions in or near the gene.

In the first wave of GWAS, all studies had relatively small sample sizes and were therefore to some extent statistically underpowered to detect variants with modest effect sizes. In recognition of this, data from three GWA studies were combined by

the DIAGRAM consortium. Through meta-analysis comprising 10,128 individuals of European descent and ~2 million SNPs directly genotyped or imputed, followed by large-scale replication in up to 53,975 individuals, six additional type 2 diabetes susceptibility genes (*JAZF1*, *CDC123*, *TSPAN8*, *THADA*, *ADAMTS9*, *and NOTCH2*) were detected [172]. The initial GWA scans were mainly performed in cases and controls from European populations. For this reason GWA scans in other populations were warranted. The first GWA studies performed in Asian subjects with type 2 diabetes discovered a new gene, *KCNQ1*, which has later also been confirmed in European subjects [173, 174]. Furthermore, a single GWA study in Taiwanese demonstrated genome-wide associations with type 2 diabetes for two other loci, *SRR* and *PTPRD* [175]. Notably, most type 2 diabetes variants have been shown to have an impact on pancreatic β-cell function with a primary effect on insulin secretion rather than on insulin action [176]. A GWA study performed in French and Danish subjects revealed, however, a variant in the *IRS1* gene, which together with *PPARG* being one of a limited number of type 2 diabetes loci so far displaying a diabetogenic potential through affecting peripheral insulin sensitivity [177].

The GWA approach has further demonstrated that genetic studies of glycemic traits can identify type 2 diabetes risk loci. Follow-up signals for type diabetes from GWA scans for fasting glucose or insulin secretion revealed from 2008 to 2010 a whole new set of type 2 diabetes susceptibility loci. The melatonin-receptor gene (*MTNR1B*), which highlights the link between circadian and metabolic regulation [178], was found to be associated with levels of fasting glucose and risk of type 2 diabetes [179-181]. Follow-up signals of a fourth GWA scan for fasting glucose identified, in addition to *MTNR1B*, five other loci (*ADCY5*, *PROX1*, *GCK*, *GCKR* and *DGKB*) associated with type 2 diabetes [182]. Very recently, several studies have reported even larger meta-analyses of GWA data from both European and Asian ethnic groups, leading to the identification of several new loci for type 2 diabetes, including *RBMS1*, *DUSP9*, *KLF14*, *ARAP1*, *HMGA2*, *HNF1A*, *GRB14*, *ST6GAL1*, *VPS26A*, *HMG20A*, *AP3S2* and *HNF4A* [183-185]. Of these new loci, genetic and gene expression studies had previously suggested an important role for *KLF14* in metabolic disease. A recent

study demonstrated a network of genes whose expression was associated with *KLF14* variation in *trans*, providing a framework for understanding how *KLF14* influences disease risk [186]. Moreover, confirmation of a common variant associations at *HNF1A* and *HNF4A* [183, 185] added new loci to those known to harbor both rare mutations causing monogenic forms of diabetes and common variants predisposing to multifactorial diabetes. The number is now 7, the others being *PPARG*, *KCNJ11*, *WFS1*, *HNF1B* and *GCK*.

Overall, the power of genome-wide association studies, in combination with larger data sets, meta-analyses of the initial GWA studies, establishments of larger consortia (e.g DIAGRAM, GIANT, MAGIC), GWA scans conducted on intermediary diabetes phenotypes (e.g. fasting glucose) and the use of study samples of different ethnicities, have delivered a whole set of new susceptibility loci for type 2 diabetes over the last five years, now counting around 50 loci [23, 183]. The validated susceptibility loci along with their discovery method, cellular function and putative intermediary mechanism in diabetes are summarized in Table 2. The last reported type 2 diabetes susceptibility regions are individually only associated with a marginally increased risk for diabetes (OR<1.1), and can together explain only ~10% of the heritability seen for type 2 diabetes [187]. Clinical factors seem to predict the risk of diabetes development better than a sample of 16 genotyped type 2 diabetes associated SNPs, either alone or in combination [188]. The clinical utility of the genome wide association studies is therefore controversial and have been highly debated [189-191].

Table 2 Genetic regions (variants) associated with type 2 diabetes at genome-wide levels of statistical significance ($p < 10^{-8}$), listed by chromosome.

Chr	Gene region (lead SNP)	Discovery method (major ethnicity)	Cellular function and putative intermediary mechanism in diabetes	References
1	***PROX1*** (rs340874)	Follow-up of signals for T2D from GWA scan for FG (European)	Encodes the prospero-related homeobox 1. Implicated in cell proliferation and development. Associated with elevated FG.	[182]
1	***NOTCH2*** (rs10923931)	GWA meta-analysis (European)	Transmembrane receptor implicated in pancreatic organogenesis; regulates cell differentiation.	[172]
2	***GRB14*** (rs3923113)	GWA meta-analysis (South Asians)	Adaptor protein binding to insulin receptor and insulin-like growth factor receptors to inhibit kinase signaling. Associated with reduced insulin sensitivity.	[183]
2	***BCL11A*** (rs243021)	GWA meta-analysis (European)	Involved in both B- and T-lymphocyte development and β-cell function. Affects insulin response to glucose.	[185, 192]
2	***RBMS1*** (rs7593730)	GWA meta-analysis (European)	Encodes RNA-binding motif, single-stranded interacting protein 1. Implicated in DNA replication, gene transcription, cell cycle progression and apoptosis. Unknown diabetogenic mechanism.	[184]
2	***GCKR*** (rs780094)	Follow-up of signals for T2D from GWA scan for FG (European)	Glucokinase regulatory protein. Involved in signal transduction, glucose transport and sensing. Associated with FG, fasting insulin and HOMA-IR.	[182]
2	***IRS1*** (rs2943641)	Single GWA study (French, European)	Encodes insulin receptor substrate-1. Associates with reduced adiposity and impaired metabolic profile (e.g. visceral to subcutaneous fat ratio, IR, dyslipidemia, CVD, adiponectin levels).	[177, 185, 193]
2	***THADA*** (rs7578597)	GWA meta-analysis (European)	Thyroid adenoma-associated gene. Associates with PPAR; Involved in apoptosis. Associated with β-cell dysfunction, lower β-cell response to GLP-1 and reduced β-cell mass.	[172, 192]
3	***ST6GAL1*** (rs16861329)	GWA meta-analysis (South Asians)	Enzyme located in Golgi apparatus, involved in post-translational modification of cell-surface components by glycosylation.	[183]
3	***ADCY5*** (rs11708067)	Follow-up of signals for T2D from GWA scan for FG (European)	Encodes adenylate cyclase 5. Involved in signal transduction. Associated with elevated FG.	[182]

Chr	Gene region (lead SNP)	Discovery method (major ethnicity)	Cellular function and putative intermediary mechanism in diabetes	References
3	***ADAMTS9*** (rs4607103)	GWA meta-analysis (European)	Proteolytic enzyme. Affects insulin response to glucose. Primary effect on insulin action not driven by obesity.	[172, 192, 194]
3	***IGF2BP2*** (rs4402960)	Single GWA study (European)	Growth factor (IGF2-mRNA) binding protein. Involved in pancreatic development and stimulation of insulin action.	[120, 121, 124, 172, 173, 185, 195]
3	***PPARG**** (rs1801282)	Candidate study; Later confirmation by GWA studies	TRF involved i adipocyte development. TRF receptor for TZDs and prostaglandins. Effect on IR.	[112, 124, 196]
4	***WFS1**** (rs1801214)	Candidate study; later validated by GWA meta-analysis	Endoplasmic reticulum transmembrane protein involved in stress and β-cell apoptosis. Insulin response.	[113, 185, 197, 198]
5	***ZBED3*** (rs4457053)	GWA meta-analysis (European)	Encodes an axin-interacting protein activating wnt/beta-catenin signaling. Unknown diabetogenic mechanism.	[185]
6	***CDKAL1*** (rs7754840)	Single GWA study (Icelandic, European)	Cyklin kinase (CDK5) inhibitor. Involved in cell cycle regulation in the β-cell. Insulin response.	[120, 121, 123, 124, 172, 173, 185, 199, 200]
7	***KLF14*** (rs972283)	GWA meta-analysis (European)	Basic transcription element-binding protein. "Master switch" controlling other genes associated with BMI, insulin, glucose and cholesterol.	[185, 186]
7	***DGKB*** (rs972283)	Follow-up of signals for T2D from GWA scan for FG (European)	Encodes diacylglycerol kinase beta. Implicated in signal transduction. Associated with elevated FG.	[182]
7	***GCK**** (rs4607517)	Follow-up of signals for T2D from GWA scan for FG (European)	Encodes the enzyme glucokinase. Involved in signal transduction, glucose transport and sensing. Associated with elevated FG and HbA1c.	[115, 182]
7	***JAZF1*** (rs864745)	GWA meta-analysis (European)	Zinc-finger protein. Function as a transcriptional repressor. Associated with prostate cancer. Insulin response.	[172, 185, 201]
8	***TP53INP1*** (rs896854)	GWA meta-analysis (European)	Encodes the p53-dependent damage-inducible nuclear protein. May regulate p53-dependent apoptosis. Unknown diabetogenic mechanism.	[185]
8	***SLC30A8*** (rs13266634)	Single GWA study (French, European)	β-cell zinc transporter ZnT8. Involved in insulin storage and secretion. Associated with fasting proinsulin levels.	[120-122, 124, 185, 199, 202]
9	***TLE4***[a] (rs13292136)	GWA meta-analysis (European)	Encodes the transducin-like enhancer of split 4. Unknown diabetogenic mechanism.	[185]

Chr	Gene region (lead SNP)	Discovery method (major ethnicity)	Cellular function and putative intermediary mechanism in diabetes	References
9	***PTPRD*** (rs17584499)	Single GWA study (Taiwanese)	Encodes the tyrosine phosphatase receptor type D protein. Associated with increased HOMA-IR and may affect insulin signaling on its target cells.	[175, 203, 204]
9	***CDKN2A/B*** (rs10811661)	Single GWA study (European)	Cyclin-dependent kinase inhibitor and p15/16 tumor suppressor. Involved in islet development. Also associated with CVD and several cancers. Insulin response.	[120, 121, 157, 185, 195, 199, 205]
10	***VPS26A*** (rs1802295)	GWA meta-analysis (South Asians)	Multimeric protein involved in transport of proteins from endosomes to the trans-Golgi network. Expressed in pancreatic and adipose tissues.	[183]
10	***CDC123*** (rs12779790)	GWA meta-analysis (European)	Cell cycle kinase, required for S-phase entry. Affects different aspects of insulin response to glucose.	[172, 192, 206]
10	***HHEX*** (rs1111875)	Single GWA study (French, European)	TRF involved i pancreatic development. Might influence both insulin release and insulin sensitivity.	[120-122, 172, 185, 195, 199, 200, 207, 208]
10	***TCF7L2*** (rs7903146)	Linkage study; later confirmation by several GWAs (European)	TRF involved in wnt-signaling. Influencing insulin and glucagon secretion. Most important polygene identified for T2D.	[107, 120, 121, 123, 129-139, 142, 143, 172, 185, 199, 209-215]
11	***ARAP1***[b] (rs1552224)	GWA meta-analysis (European)	Associated with lower proinsulin levels, as well as lower β-cell function (HOMA-B and insulinogenic index).	[185, 213]
11	***HMGA2*** (rs1531343)	GWA meta-analysis (European)	Oncogene implicated in body size (height). Primary effect on insulin action not driven by obesity.	[185, 216, 217]
11	***MTNR1B*** (rs10830963)	Follow-up of signals for T2D from GWA scan for FG or IS	Receptor for melatonin. Involved in glucose homeostasis. Associated with increased FG and reduced β-cell function.	[179-181]
11	***KCNQ1*** (rs2237892, rs231362)[c]	Single GWA study (Japanese, Asian, European)	Encodes the pore-forming alfa subunit of $I_{Ka}K^{+}$ channel. Insulin response.	[173-175, 185, 199, 218]
11	***KCNJ11**** (rs5219)	Candidate study; later confirmation by GWAS	Inwardly rectifying potassium channel. Risk allele impairs insulin secretion.	[110, 111, 121, 124]
12	***HNF1A**** (rs7957197)	Candidate study; Later confirmation by GWA meta-analysis (European)	TRF essential for pancreatic β-cell development and function.	[115, 172, 185, 205, 219-222]
12	***TSPAN8*** (rs7961581)	GWA meta-analysis (European)	Cell surface glycoprotein implicated in GI cancers. Insulin response.	[172, 201]

Chr	Gene region (lead SNP)	Discovery method (major ethnicity)	Cellular function and putative intermediary mechanism in diabetes	References
13	***SPRY2*** (rs1359730)	Single GWA study (Chinese)	Inhibitor of tyrosine kinase signaling. Associated with body fat percentage. Homologs inhibit insulin receptor-transduced MAPK signaling. Regulates development of pancreas.	[193, 206]
15	***AP3S2*** (rs2028299)	GWA meta-analysis (South Asians)	Clathrin-associated adaptor complex expressed in adipocytes and pancreatic islets. Involved in vesicle transport and sorting. Unknown diabetogenic mechanism	[183]
15	***HMG20A*** (rs7178572)	GWA meta-analysis (South Asians)	High mobility group non-histone chromosomal protein influencing histone methylation. Involved in neuronal development. Unknown diabetogenic mechanism.	[183]
15	***C2CD4A*** (rs11071657)	Single GWA study (Japanese)	Nuclear calcium-dependent domain-containing protein. Impairs glucose-stimulated insulin response. Associated with levels of fasting glucose and proinsulin.	[213, 223, 224]
15	***ZFAND6*** (rs11634397)	GWA meta-analysis (European)	Encodes a zinc finger AN1 Domain-containing protein. Unknown diabetogenic mechanism.	[185]
15	***PRC1*** (rs8042680)	GWA meta-analysis (European)	Protein regulating cytokinesis. Unknown diabetogenic mechanism.	[185]
16	***FTO*** (rs8050136, rs9939609)	Single GWA study (British, European)	2-oxoglutarate-dependent demethylase. Alters BMI i general population.	[121, 124, 172, 185, 215, 225-228]
17	***SRR*** (rs391300)	Single GWA study (Taiwanese)	Encodes a serine racemase protein. May play a role in regulation of insulin and glucagon secretion.	[175]
17	***HNF1B**** (rs757210)	Candidate study	TRF involved in development of the kidney, pancreas, liver, and Mullerian duct. Implicated in MODY and renal cyst. Associated with prostate cancer.	[114, 115, 185, 222, 229]
20	***HNF4A**** (rs4812829)	Candidate study; Later replicated by GWA meta-analysis (South Asians)	Nuclear TRF expressed in liver. Regulates transcription of several genes, e.g. HNF1A. Elevated hepatic glucose production. Defective pancreatic β-cell function and impaired insulin secretion.	[183, 219, 222, 230-233]
X	***DUSP9*** (rs5945326)	GWA meta-analysis (European)	MAP kinase phosphatase. Decreased insulin release for male risk allele carriers. Up-regulated during adipocyte differentiation. Involved in insulin signaling and stress induced IR.	[185, 234, 235]

All loci have shown genome-wide statistical significance ($p \leq 10^{-8}$). For many of the loci several SNPs associate with type 2 diabetes, but only one (in some cases two) SNPs are listed. Abbreviations: Chr:

chromosome; CVD: cardiovascular disease; FG: fasting glucose; GI: gastrointestinal; IR: insulin resistance; IS: insulin secretion; T2D: type 2 diabetes; TRF: transcription factor; TZD: thiazolidinediones. [a]GWA signal re-annotated from CHCHD9 to TLE4. [b]Formerly known as CENTD2. [c]These SNPs are in low LD in Europeans (r^2 = 0.01), thus they most likely represent two independent association signals. *Genes also implicated in MODY, other monogenic forms of diabetes or rare genetic syndromes.

Genetics of glycemic traits and glucose homeostasis. Levels of circulating glucose are tightly regulated. Indicators of glycemic status such as fasting glucose levels, HbA1c and glucose levels 2-hour post-oral glucose tolerance test (OGTT) have all heritability estimates (47-59%, 34-36% and 33%, respectively) that make them amenable to genetic analysis [236, 237]. The GWA scans on fasting glucose in both diabetic and non-diabetic subjects have shown that genetic studies of glycemic traits can identify type 2 diabetes risk loci, as well as loci containing gene variants that are associated with a modest elevation in glucose levels [179-182, 238-240].

Several loci have also reached evidence for association with HbA1c in type 1 and type 2 diabetes, as well as in non-diabetic subjects, including loci near *FN3K*, *HFE*, *TMPRSS6*, *ANK1*, *SPTA1*, *ATP11A/TUBGCP3*, *HK1*, *MTNR1B*, *GCK*, *G6PC2/ABCB11*, *TCF7L2*, *SLC30A8*, *SORCS1*, *WDR72*, *GCS* and *BNC2* [241-245]. The associations with HbA1c may well in part be a function of hyperglycemia associated with five of the loci (*TCF7L2*, *SLC30A8*, *GCK*, *G6PC2* and *MTNR1B*) [179, 181, 202, 238, 239, 246, 247]. Most of the others have been classified as novel, but some variants map to loci where more rare variants cause various forms of hereditary anemia and iron storage disorders. Common variants at these loci likely influence HbA1c levels via erythrocyte biology. Seven non-glycemic loci have shown to account for a 0.19 (% HbA1c) difference between the extreme 10% tails of the risk score, and would reclassify approximately 2% of a general white population screened for diabetes with HbA1c [243].

The most recent diabetes intermediary trait assessed by a GWA scan was proinsulin level. Proinsulin is a precursor of mature insulin and C-peptide. Higher circulating proinsulin levels are associated with impaired β-cell function, raised glucose levels,

insulin resistance, and type 2 diabetes [242]. Nine SNPs at eight loci have shown association with proinsulin levels [213]. Two loci (*LARP6* and *SGSM2*) have not been previously related to metabolic traits, one (*MADD*) has been associated with fasting glucose, one (*PCSK1*) has been implicated in obesity, and four (*TCF7L2*, *SLC30A8*, *VPS13C/C2CD4A/B*, and *ARAP1*, formerly *CENTD2*) increase type 2 diabetes risk. The proinsulin-raising allele of *ARAP1* was associated with a lower fasting glucose, improved β-cell function and lower risk of type 2 diabetes. There is no doubt that genetic studies on glycemic traits have and will continue to illuminate the biology underlying glucose homeostasis and type 2 diabetes development.

1.3.3 Genetics of common obesity

Overweight and obesity are reaching epidemic dimensions on all continents and is closely linked to the rising prevalence of type 2 diabetes. The interactions between environmental and individual factors, including genetic makeup, explain the variability in body size between individuals in a given population [248]. Genetic variation may explain as much as 40-90% of the observed population variation in body mass index (BMI) [249, 250]. Thus, huge efforts have been made to dissect the underlying heritability of obesity in order to increase our knowledge of the biological processes involved and to highlight new pathways relevant for therapeutic interventions.

In the pre-GWA era, despite the huge efforts using genome-wide linkage studies and candidate gene association studies, only mutations in genes responsible for rare Mendelian forms of severe childhood obesity were identified (e.g. genes encoding leptin, leptin receptor and proopiomelanocortin) [86]. Few genetic variations had been unequivocally associated with BMI and risk of multifactorial obesity in population studies [251, 252], with the notable exception of low-frequency variants identified in the melanocortin-4 receptor (*MC4R*) gene explaining some 2-3 % of cases of severe obesity [253]. Since 2007, the GWA approach has been similarly productive as for type 2 diabetes in the identification of common variants influencing

obesity and quantitative measures of adiposity. Currently, over 30 loci influencing BMI and the risk of obesity have been identified [23].

In the spring of 2007, several independent GWA studies reported a strong correlation between body mass index and SNPs in the human *FTO* (the fat-mass and obesity-related) gene [121, 225, 226, 228]. The association was initially seen in a study of type 2 diabetes. The diabetogenic effect, however, proved to be mediated through adiposity. Common variants in the first intron define a risk allele predisposing to both childhood and adult obesity, with homozygotes for the risk allele weighing approximately 3 kilograms more and having a 1.67-fold increased risk of obesity compared to those homozygotes for the protective allele [226]. These findings have been extensively reproduced and the effect is fairly consistent throughout different study populations (reviewed in [254]). Nevertheless, a recent study demonstrated little evidence of association in African Americans, suggesting that the effect of *FTO* variants on adiposity phenotypes may show some genetic heterogeneity dependent on ethnicity [255].

It is not fully understood how the observed effects of the *FTO* gene variants on obesity is exerted. The gene is shown to have sequence homology with Fe(II)- and 2-oxoglutarate (2OG) oxygenases [256, 257], enzymes that are implicated in various processes such as DNA repair, fatty acid metabolism, and posttranslational modifications [258]. In vitro studies have shown that recombinant FTO protein has strong preference for the abundant N6-methyladenosine (m^6A) residues in RNA in vitro [259] and 3-methylthymidine (3-meT) in single-stranded DNA or 3-methyluracil (3-meU) in single-stranded RNA [256]. Whether an altered demethylase activity can explain the link between BMI and the *FTO* gene variants remains to been seen.

Furthermore, the *FTO* protein is expressed in a wide range of human tissues, both peripheral and central, with a particularly high expression in the brain [226, 256, 260, 261]. A study on mice demonstrated that feeding and fasting up-regulated and down-regulated the *FTO* mRNA levels, respectively [261]. The *FTO* variants do not appear to be involved in the regulation of energy expenditure, but may have a role in the

control of food intake and food choice [262]. Furthermore, dietary factors and physical activity may accentuate the susceptibility to obesity by the *FTO* variants [263-265]. Based on the relative rates of the *FTO* protein expression in adipose tissue sub-fractions, other studies point out a role of *FTO* on the level of the adipose tissue metabolism [266, 267]. The functional role of FTO in energy homeostasis remains elusive.

The *FTO* risk variants have also shown association with obesity-related conditions and diseases such as the metabolic syndrome, polycystic ovary syndrome (PCOS), and type 2 diabetes [227, 268, 269]. However, in most of the studies reported, these effects appear to be secondary to weight increase since the associations are attenuated or completely abolished after adjusting for BMI [227].

After the identification of the strong signal between *FTO* variants and obesity, several other interesting associations with BMI have been reported. Signals near genes implicated in features of neuronal function (e.g. *BDNF*, *SH2B1* and *NEGR1*) highlights the neuronal influence on body weight regulation [270-273]. Other studies, performed in cases and controls selected from the extremes of the BMI distribution have revealed an additional, partly overlapping, list of susceptibility loci [274-277]. Moreover, GWASs and meta-analysis on patterns of fat distribution have also identified a wide range of variants different from those influencing overall adiposity [278-281]. Interestingly, many of these variants reveal sexual dimorphism in the genetic basis of fat distribution [280].

2. AIMS

Vision

In this project, the overall aim was to establish a research design to investigate genotype-phenotype relations between common genetic variants, type 2 diabetes and type 2 diabetes-related traits using subjects from a large population-based Norwegian material - the second Nord-Trøndelag Health Study (HUNT2). By evaluating past and current findings based on GWA studies, we sought to validate and expand the results, and thus to contribute to the continuing exploration and dissection of the complex genetic background of type 2 diabetes and related phenotypes.

Specific aims

1) To examine and validate several newly identified type 2 diabetes susceptibility loci for association with type 2 diabetes, obesity and lipid measures in a completely unselected population of type 2 diabetes patients (paper I).

2) To investigate whether a genetic variant in the gene *FTO* affects type 2 diabetes risk entirely through its effect on BMI, and how this variant influences BMI across adult life in Scandinavian subjects using both cross-sectional and longitudinal perspectives (paper II).

3) To investigate the very interesting chromosome 9p21 region emerging from the GWAS for type 2 diabetes and cardiovascular diseases in an attempt to fine-map the functional variant(s) involved in type 2 diabetes and cardiovascular disease, and to investigate the relation between the association signals (paper III).

4) To evaluate whether four novel genetic variants affecting hemoglobin A1c levels in subjects with type 1 diabetes have an effect on glycemia in a population-based type 2 diabetes cohort (paper IV).

3. RESEARCH DESIGN AND METHODS

3.1 DEFINITION OF COHORTS

3.1.1 The Nord-Trøndelag Health Study

The Nord-Trøndelag Health Study (HUNT) is a longitudinal population-based health study and is considered the largest and most unique database of personal and family medical histories in Norway, and possibly in the world. The database is a product of tight collaboration and joint efforts between local, regional, national and international partners throughout the last 20 years. The data was collected during three intensive studies, HUNT1, HUNT2 and HUNT3. Compared to population studies in other countries, the HUNT study is somewhat unique, given that a total population within a geographical area was covered, the age range is wide, the study covers an extensive range of topics, and the participation rate was high. Nord-Trøndelag is a county with 127,000 inhabitants in the middle of Norway. The county is representative of Norway as a whole with regard to the economy, industry and sources of income, age distribution, morbidity and mortality rates. The population is stable both ethnically and geographically, with less than 3% of people of non-white origin and a net emigration level of 0.3% per year (1996–2000) [282].

HUNT1

HUNT1 was carried out in 1984-1986 to establish the health history of the people aged ≥ 20 years in Nord-Trøndelag. Of those invited, the attendance rate was 88.1% (74,599 participants). The main objectives in this health survey aimed at hypertension, diabetes, lung disease and quality of life. Each participant had his/her blood pressure, height and weight measured and was asked to answer two different questionnaires. For those over the age of 40 years, non-fasting serum glucose was measured. No blood samples were collected in HUNT1.

HUNT2

The second HUNT Study which took place from 1995-97 was partly a follow-up study of HUNT1. However, HUNT2 comprised a larger scientific program including data and blood samples from more than 66,000 subjects, of whom ~45,000 had participated in HUNT1. Of the 92,434 individuals considered eligible for the study, the participation rate was 71.3% [282]. The main objectives in this health survey focused on large public health issues like cardiovascular disease, diabetes, obstructive lung disease, osteoporosis and mental health. A personal invitation was distributed to all inhabitants 20 years or older. The invitation included a questionnaire 1 (Q1) and suggestions for time and place for a clinical examination. During the medical examination data were collected from two additional questionnaires (Q2 and Q3) and combined with blood and urine samples and various clinical measurements, some of them in sub-samples of the population. Four different versions of Q2 were handed out according to age and gender. Q3 was delivered to those individuals reporting hypertension, lung disease, diabetes or asthma in Q1. The participants were asked to complete Q2 and Q3 and returned them by mail. Thus, data are missing more often for Q2 and Q3 than for Q1. The clinical examination of each participant was carried out by specially trained nurses and technicians at established screening stations in the largest towns or by visiting smaller communities with transportable office and laboratory facilities. Blood sampling was done whenever subjects attended, thus considered to be in a random or a non-fasting state. Serum analyses were performed in fresh blood samples at Levanger Hospital. Glucose, total cholesterol, HDL cholesterol and triglycerides were measured by standard enzymatic methods. In those who confirmed to have diabetes in Q1 an extra tube of whole blood was drawn for analyses of HbA_{1c} [282].

3.1.2 Malmö Diet and Cancer Cohort (MDC)

The Malmö Diet and Cancer cohort [283] was a 10-year prospective case-control study with baseline examinations from March 1991 to October 1996 and consisted of 28,449 individuals. All men born 1923-1945 and all women born 1923-1950 and

living in Malmö were invited to participate. The main objective was to investigate whether a western diet could be associated with certain forms of cancer accounting for other life-style factors. Moreover, the study also aimed to act as a resource available for testing innovative hypotheses arising from other studies. Initially food intake, heredity, socio-economic factors, life-style pattern, occupational situation, previous and current diseases, symptoms and medications, were determined. Plasma/serum was collected. Nurses conducted anthropometric measurements and collected blood. Diabetes at baseline was identified through self-reported diabetes diagnosis, self-reported use of anti-diabetic therapy or fasting whole blood glucose levels.

3.1.3 Malmö Preventive Project (MPP)

The Malmö Preventive Project (MPP) was a preventive case-finding program for cardiovascular risk factors and alcohol abuse initiated in Malmö in 1974 [284, 285]. The major aim was to screen large stratum of the adult population in order to find high-risk individuals for preventive intervention. Subjects were invited to participate in a broad health-screening program, including a physical examination and a panel of laboratory tests. Additionally, every participant filled in a self-administered questionnaire on lifestyle, and medical history. 33,346 Swedish subjects (22,444 men and 10,902 women; mean age 49 years, 24.5% with impaired fasting glucose and/or impaired glucose tolerance from the city of Malmö in southern Sweden participated in a health screening during 1974-1992, with a 71% attendance rate. Eligible participants were invited to a re-screening visit during 2002-2006 at the Clinical Research Unit Medicine, University Hospital, Malmö. This screen included a physical examination and fasting blood samples for measurements of plasma glucose and lipids, as well as whole blood samples for DNA extraction.

3.2 Study samples

Our research group (The Bergen Diabetes Research Group) is collaborating with HUNT, and currently we have access to extensive clinical data and DNA samples from ~6,000 participants aged ≥ 20 years from HUNT2. Initially (in 2007), we had access to extensive clinical data and DNA-samples from 1,850 participants with diabetes and 2,956 random controls (dataset 1 in Figure 7). We had genomic DNA available for 1,850 diabetic participants, for 1,391 of whom more extensive clinical diabetes data were available. Subjects who were GAD antibody–positive and had diabetes onset before age 40 years, or before age 30 years with insulin treatment initiated during the first year of diagnosis, or had continuously been on insulin treatment since the year of diagnosis were considered not having type 2 diabetes and thus excluded from the type 2 diabetes group. For 459 diabetic subjects with no GAD antibody measurements, subjects with diabetes onset before age 40 years were also excluded from the type 2 diabetes group [231]. Based on these criteria we had three groups; one group consisting of 1,644 subjects with type 2 diabetes, one with 206 subjects with suspected type 1 diabetes or other forms of diabetes, and a group of 2,956 non-diabetic (self-reported) control participants.

Two years later we increased our sample size with DNA and clinical data from 300 new subjects with a history of myocardial infarction, 300 subjects with history of stroke and 500 new random controls. Altogether, dataset 2 consisted of 1,100 individuals and none of these were overlapping with dataset 1 (see Figure 7). In total, our current HUNT2 material consists of two different datasets, counting 5,906 individuals all together. In addition, we had access to data on weight, height and diabetes status from HUNT1 (1985) for 4,625 of the 5,906 subjects in HUNT2, giving a follow-up time of 10 years.

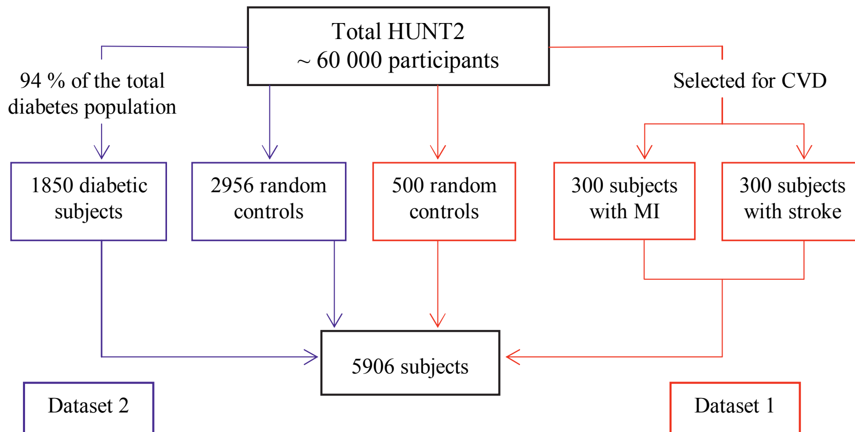

Figure 7 Schematic overview of our HUNT2 study sample. Our HUNT2 material consists of two different datasets, counting 5,906 individuals all together. Dataset 1: 1,850 diabetic subjects and 2,956 non-diabetic controls. Dataset 2: 300 subjects selected for myocardial infarction, 300 subjects selected for stroke and 500 random controls with neither cardiovascular disease nor diabetes.

Participants in HUNT2 did not always fill inn all questions in the various questionnaires, resulting in different numbers of valid responses in different parts of the HUNT database. In the dataset obtained from the HUNT biobank we explored that there were also varying levels of missing data for some of the variables collected from the clinical examinations and laboratory analyses. Thus, the study samples derived and the number of subjects that were enrolled in paper I-IV differed according to the year of initiation and the different research questions raised in each paper.

In paper II, we obtained summary statistics from the MPP and MDC studies and the study groups enrolled in paper II were derived as follows. In the MPP cohort, 33,346 Swedish participated in the health screening. Of those persons participating in the initial screening, 4,931 have died, and 551 were lost to follow-up. Of those twenty-five thousand invited to a re-screening visit, 17,284 persons participated in the re-screening; of these 1,223 were excluded because of incomplete medical information

or lack of a DNA sample (or type 2 diabetes at baseline). Thereby, 16,061 non-diabetic subjects, 2,063 of whom developed type 2 diabetes during follow-up, were included in the analyses. In the MDC cohort (n=30447), 27948 were successfully genotyped for *FTO* rs9939609. Out of those, 27,901 had information on BMI, age and diabetes status. After excluding individuals that were part of MPP and were genotyped for *FTO* (n=7,923), 19,978 remained and constituted the study sample. Differential missingness was thus not a problem here as all the genotyped MDC individuals with needed clinical information were included in the analyses, either as part of MDC study, or as part of MPP study.

3.3 GENETIC ANALYSES

3.3.1 SNP selection

In parallel with this work, putative new loci showing association with type 2 diabetes and related traits were identified at an increasing speed. Hence, at the initiation of our studies, inclusion criteria for SNPs were based on the available information at that time. The panel of SNPs selected could in many cases have been more wide-ranging and thus ideally extended. However, with reference to a multiplex level of ~25 SNPs to ensure a high rate of success (i.e. high genotyping efficiency) and the need to keep the costs of the SNP scoring to a reasonable level, we decided to type the best candidates related to our research questions of interest and from a cost/benefit standpoint.

In paper I, the selection of SNPs prioritized for genotyping was based on publicly available type 2 diabetes GWA data as of June 2007, i.e. Sladek et al., the DGI–FUSION–WTCCC collaboration and deCODE Genetics [120-124, 226]. We included only SNPs that were robustly replicated in at least two of the GWASs, i.e. *TCF7L2* (rs7903146 and rs12255372), *KCNJ11* (rs5219), *PPARG* (rs1801282), *IGF2BP2* (rs4402960), *CDKAL1* (rs7756992), *SLC30A8* (rs13266634), *CDKN2B* (rs10811661), *HHEX* (rs1111875) and *FTO* (rs9939609). Because of the close relation between the Norwegian and Swedish populations, we also included two SNPs in *PKN2*

(rs6698181) and *FLJ39370* (rs17044137), also known as *C4ORF32*, which had shown rather strong evidence of association in the DGI Scandinavian data set [120], although not in the UK sample sets [24]. Unfortunately, we had to exclude the SNPs in *TCF7L2*, *KCNJ11* and *PPARG* from further analysis and publication due to fact that we did not attain analysis rights from the HUNT Research Center publication committee for these three genes in the cohort.

In the second paper, wherein we aimed to validate and expand the results regarding the relations between *FTO*, type 2 diabetes and weight gain in subjects from three large Scandinavian cohorts, we decided to include only one SNP, rs9939609, which was reported by the first GWAS [226]. The reported obesity-associated *FTO* SNPs are in strong to perfect LD (pairwise r^2 >0.8; HapMap; CEU, release 21), hence rs9939609 functioned as a proxy for the cluster of SNPs in the first intron of *FTO* that have demonstrated association with both type 2 diabetes and BMI.

In paper III, we applied a fine-mapping approach and defined a target region for tagSNP selection across the 9p21 interval (Chr9:21,995,330-22,133,570, NCBI Build 36). 18 SNPs (rs3217986, rs523096, rs10965215, rs1759417, rs1333034, rs8181047, rs10811647, rs1333039, rs16905599, rs10811658, rs7045889, rs10811659, rs10757282, rs1333051, rs4977761, rs2065501, rs6475610, rs10757287) tagging a 138-kb region were selected using the Haploview implementation of the Tagger algorithm [286] using the following criteria: minor allele frequency (MAF) of >5% and pairwise r^2 >0.80. In addition, two previously GWAS-identified type 2 diabetes susceptibility variants (rs564398 and rs10811661) and three confirmed CVD-susceptibility variants (rs1333040, rs10757278 and rs1333049) were added to the SNP set, making a total of 23 SNPs. Four variants (rs1759417, rs1333049, rs7045889 and rs6475610) did not pass our quality control criteria and were excluded from further analyses. Thus, a total of 19 SNPs were included in the association analysis with type 2 diabetes, stroke, myocardial infarction, angina pectoris and coronary heart disease.

Although a wide panel of genetic variants have shown association with glycemic traits in both non-diabetic and diabetic subjects, we decided in paper IV to include only the four novel variants reported by Paterson and colleagues [241], which had shown the strongest GWA signals with glycemic control in subjects with type 1 diabetes collected from the Diabetes Control and Complications Trial (DCCT). These were the non-coding SNPs rs10810632, rs1358030, rs11624318 and rs566369 located in or close to the *BNC2*, *SORCS1*, *GSC* and *WDR72* genes, respectively.

3.3.2 Genotyping

A plethora of different SNP genotyping methods has been developed over the past years. One principal approach in SNP genotyping is to score multiple SNPs from each sample in a multiplexed fashion (more than one primer/probe pair per reaction). In paper I-IV, all samples from the HUNT2 cohort were genotyped using the multiplex MassARRAY® iPLEX™ System (SEQUENOM Inc., San Diego, CA, USA) at the Norwegian genotyping platform CIGENE, Ås, Norway. The Sequenom MassARRAY system utilizes a MALDI-TOF (Matrix-assisted laser desorption ionization - time of flight) mass spectrometry for genotyping. In brief, this technology is based on an allele-specific primer extension reaction where short primers are extended, according to the base composition in the template sequence, and separated by mass [287]. The differences in mass captured by the MALDI-TOF are automatically translated by the software into specific genotype calls.

The service provided by CIGENE included primer and assay design all the way through to production of genotypes. We were typically multiplexing at a level of 20 to 25 SNPs depending on the project and the ability to make proper designs.

In the MPP cohort, subjects were genotyped using the TaqMan allelic discrimination assay-by-design method on ABI 7900 (Applied Biosystems, Foster City, CA). An allelic discrimination assay is a multiplexed end-point (data collection at the end of the PCR process) assay that detects variants of a single nucleic acid sequence. The presence of two primer/probe pairs in each reaction allows genotyping of the two

possible variants at the SNP site in a target template sequence. The actual quantity of target sequence is not determined. For each sample, a unique pair of fluorescent dye detectors is used, for example, two TaqMan probes that target an SNP site. One fluorescent dye detector is a perfect match to the major allele 1 and the other fluorescent dye detector is a perfect match to the minor allele 2. The allelic discrimination assay classifies unknown samples as homozygotes or heterozygotes. Hence, the allelic discrimination assay measures the change in fluorescence of the dyes associated with the probes. In the MDC cohort, the genotyping was performed by a SEQUENOM MassARRAY platform, and/or TaqMan genotyping assay.

3.3.3 Quality Control (QC)

In paper I-IV we applied standard QC procedures to only allow for further analysis and interpretation of data with acceptable quality. Quality control was achieved by sample tracking methods, genotyping protocols, genotype design for samples including numbers and plating locations, internal control samples (duplicate samples from the same DNA collection), assay call rates, assay reproducibility and concordance with previously generated genotypes and frequencies. Finally, we persistently tested for deviations from Hardy-Weinberg proportions to detect failed assays or large-scale stratification separately in cases and controls.

3.4 ETHICS

3.4.1 Ethical issues and approvals

Both the core study and each sub-study of HUNT were approved by the Data Inspectorate of Norway and recommended by the Regional Committee for Medical Research Ethics, and all information from HUNT is treated according to the guidelines of the Data Inspectorate [282]. Participation in the HUNT study was voluntary, and written informed consent was obtained regarding the screening, subsequent control and follow-up, and to the use of data and blood samples for

research purposes. All data files had only a laboratory number and no personal identification number. Only the HUNT Research Center had access to the key, which could link the two numbers in question. We had no contact with the participants. The MDC and MPP studies and their follow-up analyses had been approved by the Regional Ethics Committee in Lund. Written informed consent was obtained from all participants.

The study protocols for paper I-IV were approved by the Regional Committee for Research Ethics and the Norwegian Data Inspectorate, and all studies were performed according to the latest version of the Helsinki Declaration.

3.4.2 Funding

The HUNT Study was funded by joint efforts of a large number of partners. Main contributions came from The Ministry of Health, through The National Institute of Public Health and The National Health Screening Service. The Nord-Trøndelag County Council, The Norwegian University of Science and Technology and The Norwegian Research Council were also providing essential funding. Sub-studies were supported by The Norwegian Research Council or a number of private organizations, like The Diabetes Association. Some sub-studies were supported by pharmaceutical industry such as the diabetes study by GlaxoSmithKline.

The MPP Study was supported by grants from the Swedish Research Council (including Linné grant 31475113580), the Heart and Lung Foundation, the Diabetes Research Society, a Nordic Center of Excellence Grant in Disease Genetics, the Diabetes Program at the Lund University, the European Foundation for the study of diabetes, the Påhlsson Foundation, the Craaford Foundation, the Novo Nordisk Foundation, the European Network of Genomic and Genetic Epidemiology and the Wallenberg Foundation. The MDC Study was supported by project grants from the Swedish Research Council, the European Foundation for the Study of Diabetes, the Novo Nordisk and Albert Påhlsson Foundations, a Linnaeus grant to the Lund University Diabetes Centre, and the Knut and Alice Wallenberg Foundation.

The genotyping was in part provided by the CIGENE technology platform (Ås, Norway), which is supported by the Functional Genomics Programme (FUGE) of the Research Council of Norway. This work has been supported by funds from the University of Bergen, Haukeland University Hospital, Helse Vest, Innovest and the Research Council of Norway. One of the co-authors of paper II (LG) has been a consultant for and served on advisory boards for Sanofi-Aventis, GlaxoSmithKline, Novartis, Merck, Tethys Bioscience and Xoma, and received lecture fees from Lilly and Novartis. All the others authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported. The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscripts.

4. MAIN RESULTS

4.1 PAPER I

The first wave of genome-wide association studies procured substantial support for several new type 2 diabetes susceptibility variants; however, these studies were mainly performed in carefully selected cases (sometimes the extremes of the phenotype) and controls which may not be representative of the entire type 2 diabetes populations. In this paper, using a completely unselected population of type 2 diabetes subjects and controls (the HUNT2 Study), we confirmed the association with type 2 diabetes for three SNPs in or near the genes *CDKN2B* (rs10811661, OR =1.20), *FTO* (rs9939609, OR=1.14), and *SLC30A8* (rs13266634, OR=1.20) and observed a borderline significant association for the variant near *IGF2BP2* (rs4402960, OR=1.10). Notably, the *FTO* variant, which was previously shown to be associated with diabetes probably via a primary effect on obesity, was still significant after adjustment for BMI. Moreover, the association results for the *HHEX* SNP (rs1111875) and the *CDKAL1* SNP (rs7756992) were non-significant and showed slightly lower ORs than in previous studies. We found no support for an association with the less consistently replicated *FLJ393370* (*C4ORF32*) or *PKN2* SNPs. We hypothesized that some SNPs could have a more pronounced effect on type 2 diabetes in obese participants than in non-obese participants; however, results were not significantly different between the obese and the non-obese groups.

Quantitative metabolic traits analysis demonstrated in agreement with previous studies that *FTO* was strongly associated with BMI. The association with obesity shown for *FTO* using BMI as phenotype could not be demonstrated using waist-to-hip ratio as a quantitative trait for obesity after adjustment for age, gender and diabetes status. We also demonstrated an association with triglyceride levels for the *FTO* SNP. Furthermore, the SNP in the vicinity of *CDKN2B* indicated association with waist-to-hip ratio and also a nominal association with cholesterol. We found no

evidence for association with quantitative metabolic traits for the SNPs in or near *PKN2*, *IGF2BP2*, *FLJ39370*, *CDKAL1*, *SLC30A8* and *HHEX* in this study.

4.2 PAPER II

This paper describes the relationship between a genetic variant (rs9939609) in *FTO*, type 2 diabetes and weight gain throughout adult life in 41,504 subjects from the Scandinavian HUNT, MDC and MPP studies. The initial analyses in the HUNT cohort followed by the meta-analyses of data from all three cohorts, revealed a highly significant association for rs9939609 with both type 2 diabetes (OR=1.13, P=4.5×10^{-8}) and the risk to develop incident type 2 diabetes during follow-up (OR=1.12, P=3.2×10^{-8}) after correction for age and gender. In contrast to previous findings, the associations with both type 2 diabetes and incident type 2 diabetes remained significant also after correction for BMI, waist circumference, waist-to-hip ratio or change in BMI over time (ΔBMI). We have previously suggested a link between SNPs in *FTO* and altered lipid profiles (paper I), but we were not able to confirm this finding in a meta-analysis in this paper.

The meta-analysis of the *FTO*-associated allele-wise effect on BMI using cross-sectional data from the HUNT, MPP and MDC studies confirmed the strong effect of the *FTO* SNP on BMI (0.28 kg/m^2 per risk allele, P= 2.0×10^{-26}), with no significant heterogeneity in the effect sizes for the risk allele between different adult age groups. Interestingly, using longitudinal data from the HUNT and MPP studies, we found no differences in change of BMI over time according to rs9939609 risk alleles overall (ΔBMI=0.0 (-0.05, 0.05)), or in any individual age stratum. Hence, our results indicate that the additional weight gain as a result of the *FTO* risk variant seems to occur relatively early in life, most likely before adulthood, and the relative BMI difference remains stable thereafter.

4.3 PAPER III

In this paper we applied a genetic fin-mapping approach by assessing 19 SNPs tagging a 138-kb region on chromosome 9p21 for association with type 2 diabetes, stroke, myocardial infarction, angina pectoris, and coronary heart disease. Through single point and haplotype analysis, we found evidence for only one common type 2 diabetes risk haplotype (OR=1.20) in the region. The risk haplotype was completely tagged by two markers, rs10757282 and rs10811661, and we mapped the break-up of this haplotype and confined a maximal 11-kb candidate region located 117-128 kb upstream of the *CDKN2B* gene. There was no evidence of additional type 2 diabetes association signals elsewhere in the region.

Furthermore, we confirmed the strong association between SNPs in the 60-kb cardiovascular disease block with angina pectoris and myocardial infarction (p>0.01). We observed two apparently independent and suggestive single SNP association signals for myocardial infarction in the 138-kb interval; one SNP located in a small 2-kb region just upstream the previously implicated type 2 diabetes risk block (rs2065501, OR 1.18), and one other located in the 3'UTR of the *CDKN2B* gene ~59-kb downstream of the well confirmed cardiovascular disease block (rs3217986, OR=0.70). The latter SNP was also significantly associated with angina. Notably, both of these two single SNP associations remained significant after conditioning upon the lead cardiovascular disease SNPs in the region

4.4 PAPER IV

Paper IV describes an attempt to evaluate whether four novel SNPs reported to affect HbA1c levels in subjects with type 1 diabetes had an effect on glycemia in a type 2 diabetes cohort. We observed allele frequencies similar to the frequencies reported in individuals with type 1 diabetes, however, in the individual SNP analysis, No significant associations with HbA1c or glucose levels were found for the *SORCS1*, *BNC2*, *GSC* or *WDR72* variants (all P-values >0.05). Although the observed effects

were non-significant and of much smaller magnitude than previously reported in type 1 diabetes, the *SORCS1* risk variant showed a direction consistent with increased HbA1c and glucose levels, with an observed effect of 0.11% and 0.13 mmol/l increase per risk allele for HbA1c and glucose, respectively. In contrast, the *WDR72* risk variant showed a borderline association with reduced HbA1c levels, and direction consistent with decreased glucose levels.

When we included all four variants in a combined genetic score model we observed no evidence for a relationship between increasing number of risk alleles and increasing HbA1c or glucose levels. Each additional risk allele demonstrated an increase in HbA1c of approximately 0.04%.

5 DISCUSSION

5.1 REPLICATION OF GENOTYPE-PHENOTYPE ASSOCIATIONS (PAPER I)

By performing the studies described in this thesis, reviewing the literature, and learning from the design of successful genetic association studies, four key features become clear:

1) Because of the moderate risk conferred by many common genetic variants, it is crucial to design an adequately powered study with large sample sizes that are carefully controlled to minimize bias.

2) Thorough phenotype characterization of study subjects is of importance to ensure appropriate categorical classifications and also to allow for statistical adjustment for confounding factors, especially related to environmental exposures.

3) SNP selection and detection is critical. There is a continuing attempt to catalog more SNPs across the genome and to explore new methods to assay SNP genotypes more densely.

4) Even statistically convincing associations require validation by replication in independent cohorts, and in cohorts of other ethnicities.

The latter will be more comprehensively discussed here since it reflects one of the fundaments for this thesis.

The importance of replication. During the last 10-15 years, genetic association studies have increasingly gained in importance and become an essential approach for unraveling the genetic architecture of complex diseases. Although genetic association studies offer a potentially powerful approach to detect genetic variants that influence susceptibility to common disease, the failure to replicate or validate association

results are rather common and a serious concern [288, 289]. Why does replication often fail?

Several reasons have been proposed to explain the inconsistent findings from association studies. Bias may be caused by factors that lead to systematic deviations from the true effect of a genetic association. Biases may operate at the level of a single study, a collection of studies (e.g. meta-analysis), or a research field at large. They may arise in: a) the study design (selection or recruitment of participants, retrospective or prospective collection of DNA samples, and method of gathering information on phenotypes, exposures and covariates); b) DNA extraction method; c) production of genotype data; d) raw data management; e) data processing and data analysis; f) reporting of analyses; g) integration of studies through meta-analyses or f) integration of meta-analyses into field synopses [81, 290, 291].

Two potential sources of bias are particularly recognized in genetic association analyses: population stratification and genotyping error. The magnitude of population stratification effects, reflecting different ancestral history that includes responses to natural selection, migration patterns, and founder events, remains a debated concern. They are expected to be small in well-designed studies, but the effect of population stratification on the results of association analyses are potentially more severe when small effects are studied in very large sample sets. Thus, a true association in one population is not always true in another population because of heterogeneity in genetic or environmental background. Biases such as population stratification can also be the reason for a false positive finding. On the contrary, non-replication could arise due to chance and to a false negative finding in the replication study. Hence, besides investigating possibly insufficient power and flaws or biases of the replication study, it is also recommended and important to investigate possible heterogeneity between the different samples studied.

Why is replication and validation important? The first reason is more general for all association studies and is based in the study design itself. A direct establishment of a casual relationship from an observed association is not possible because experimental

designs like randomization cannot be applied to genetic associations [289]. To infer causality of a genetic factor to a phenotype from an observed association, one also needs to consider other classical criteria or conditions such as strength, consistency, specificity, temporality, biological gradient, plausibility, analogy and coherence [289, 292]. To exemplify: The association should have a biological explanation (= plausibility), the result should agree with relationships between similar expositions/genotypes and diseases (= analogy), and the result should not conflict with existing knowledge (= coherence). However, for genetic associations, only in some fortunate cases a biological explanation or analogy situation might be available. Thus, inferring causality from a genetic association really depends upon the current biological knowledge [289].

Also from the statistical viewpoint the need for replication is a major criterion. Considering the apparently high proportion of false positive reports in the literature, very stringent criteria for interpreting association studies are needed [89]. The threshold of replication has been a matter of considerable debate in recent years [293-295]. With genome-wide testing of hundreds of thousands of polymorphisms, many would argue that P-values at least $< 10^{-7}$ are needed for establishing a strong candidacy for further replication and validation. More conventional thresholds of statistical significance like $P <0.05$ may still be used appropriately for the replication and validation phase of a proposed association, if the discovery data are excluded, the replication is limited to a specific polymorphism and a specific model of association is analyzed [290]. Nevertheless, a single, nominally significant association should be viewed as tentative until it has been independently replicated in other studies.

A third reason for why replication and validation is so important is more specific for genetic association studies. The history of genetic association studies has frequently shown that in the first study of an association, the effect is overestimated, and that there is just a modest correlation between effects in the first and the following studies on the same association [90, 91, 289, 293]. This disagreement between effects can be explained by either true difference between the populations studied or be a result of a

phenomenon known as the" winner's curse" [296], an event based on the fact that the associations with the strongest effects are overestimated. Linked to regression to the mean [297], this phenomenon takes place primarily because within a small sample, a weak effect becomes significant only if the effect is overestimated. Furthermore, the "winner's curse" phenomenon is further strengthened by selective reporting of analyses, possibly biased interpretation of results and publication, and other forms of bias. As a consequence, the first report of an association is more likely to be a false positive and the effect overestimated, thus emphasizing the need for replication or validation of the association results. Replication and validation efforts of genetic association results are therefore highly emphasized by the scientific community [289].

In this project, the overall aim was to establish a research design to investigate genotype-phenotype relations between common genetic variants, type 2 diabetes and type 2 diabetes-related traits. The whole-genome association scans, in combination with large data sets has shown its promise and delivered a whole set of new susceptibility loci for type 2 diabetes [23]. To get a better estimate of the true effects conferred by the susceptibility loci identified in the first wave of GWAS, we sought replication and validation of the genetic variants using a large population-based cohort – HUNT2. Paper I describes one of the first replication studies of the new type 2 diabetes susceptibility variants. The paper provided some important insights, as it confirmed most, but not all, previously identified loci. Thus, our results show that GWAS findings can be generalized to a completely unselected population such as the Norwegian HUNT study (Figure 8). We confirmed the diabetes association at the *SLC30A8* locus and the rs10811661 SNP located on chromosome 9p21, 125 kb from the nearest gene (*CDKN2B*). Our data were less compelling with regard to the SNPs tested near *IGF2BP2*, *HHEX* and *CDKAL1*. However, there was a trend in the same direction and of the same magnitude as in previous reports [120-124]. These genetic variants have been confirmed to be associated with type 2 diabetes, although in some cases with modest evidence in the initial stages and strengthened evidence only in combined analyses [121]. Our findings probably reflect the possibility that the associations are stronger in certain subgroups and that very large sample sizes are

needed to formally replicate the *IGF2BP2*, *HHEX* and *CDKAL1* data. It should also be emphasized that the risk variants had not been fine-mapped and that even subtle differences between different populations might affect linkage disequilibrium between test and disease variants. The DGI Study defined two loci, *FLJ39370* and *PKN2*, as interesting for follow-up studies [120], whereas the WTCCC/UKT2D and FUSION studies showed conflicting results [121, 124]. We were not able to detect any association between type 2 diabetes and *FLJ39370* or *PKN2* candidate SNPs in the Norwegian sample, which is supposed to be genetically closely related to other Scandinavian populations. Hence, these two SNPs probably represent false positive results in the DGI whole-genome scan.

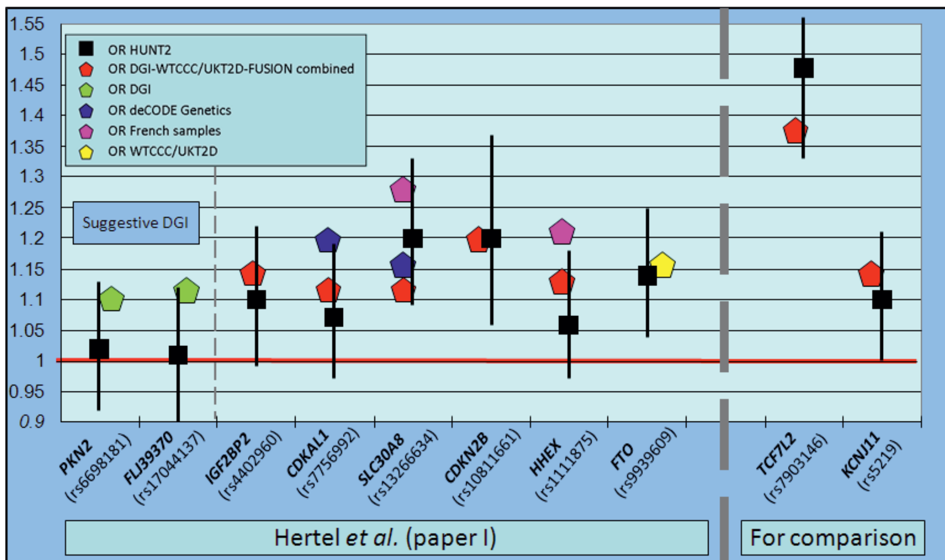

Figure 8 Association results for confirmed type 2 diabetes susceptibility loci based on GWA results from the DGI [120] , WTCCC/UKT2D [124], deCODE Genetics [123], French samples [122] and the DGI-WTCCC-FUSION combined analysis [120] compared to results from our study in HUNT2 (Paper I) . The size effects that we observed are close to the estimates from the previous studies, indicating that the design of the first round of the whole-genome scans seems to pinpoint risk-alleles that show generality, at least in other Northern European populations.

We also validated the HUNT samples by genotyping known type 2 diabetes risk variants in *TCF7L2*, *KCNJ11* and *PPARG*. However, as previously mentioned, we had to exclude these three SNPs from further analysis and publication due to fact that we did not attain analysis rights from the HUNT Research Center publication committee for these three genes in the cohort. We found a significant association with type 2 diabetes for the SNPs rs7903146 and rs12255372 in *TCF7L2* and rs5219 in *KCNJ11*, similar to recent data reported by Thorsby et al. [214], and with an OR similar to other studies [107, 109], indicating that the HUNT2 population contains a representative diabetes cohort and that our genotyping strategy was robust.

The HUNT2 study includes a well-characterized population from a clearly defined region of Norway where participants were recruited without regard to disease status and where the controls were drawn from the same population. Thus, there was no clear selection bias that could arise when conducting genetic studies [209]. A recent study by Almgren and colleagues [105] demonstrated that the strongest heritability for type 2 diabetes was seen in patients with age at onset 35-60 years (h^2 = 0.69). Inclusion of patients with onset up to 75 years made the heritability estimate drop to 0.31. Thus, to detect stronger genetic effects in type 2 diabetes, it seems reasonable to restrict inclusion of patients to those with age at onset 35-60 years. Of the total diabetes population in HUNT2, we had data from 94%, comprising 1644 type 2 diabetes cases. The mean age at diagnosis for these patients were 58.4 (± 12.1) years. Thus, we believe that our study contained a proper type 2 diabetes cohort for detection of the most important genetic effects.

Although we did not have the opportunity to formally test for possible population sub-structures with the limited numbers of markers genotyped, we believe that the type 2 diabetes cohort of HUNT2 is more, rather than less homogeneous than other type 2 diabetes sample collections. Furthermore, the allele frequencies and the size effects are similar to previous publications, arguing against problems with population stratification. However, although appropriate and detailed sampling of cases and controls may diminish problems of population stratification, even in well-designed

studies modest amounts of population stratification can be detected [298]. The effect of population stratification on the results of association analyses are, however, potentially more severe when small effects are studied in very large sample sets [299].

As a replication study, the sample size of ~3,500 patients and control participants was only powered to detect relatively small ORs (ranging from 1.1 – 1.4) at a nominal significance. Our study had 60-90% power for detection of the different OR estimates for the susceptibility variants reported in the first GWA wave on type 2 diabetes. We argue that it is not necessary to correct for multiple comparisons when using diabetes as a trait, since this could be considered a pure replication study. The NCI-NHGRI Working Group on Replication in Association Studies has set up precise criteria for establishing a positive replication [293]: a) the sample size (N) should be sufficient; b) the phenotype should be similar; c) the replication population should be similar – if the confirmation sample stems from a population which is different than that from which the original sample was drawn (e.g. ethnicity, phenotype, time etc), validation of the genetic association is attempted; d) the phenotype should be more or less similar, e) the effect should be similar in magnitude and in same direction; f) the replication marker(s) should be the same or in very high LD, selected with a strong rationale; g) the genetic model should be the same; h) P-values of combined analysis should be smaller than the initial P-value; and finally, i) the report of replication should be at the same level of detail as for the initial report.

In view of all of the above-mentioned criteria, Paper I represents a positive replication and validation of the type 2 diabetes susceptibility variants reported in the first wave of GWAS on type 2 diabetes. However, for additional phenotypic traits, the results were more explorative and further studies are needed to address whether our observed associations with waist-to-hip ratio and triacylglycerol levels represent true effects or spurious associations.

As part of the WTCCC study, Frayling et al. [226] reported that SNPs in the first intron of the *FTO* gene were highly associated with type 2 diabetes and BMI,

suggesting that the *FTO* locus exerts its primary effect on adiposity and that it subsequently has an impact on type 2 diabetes [226]. In Paper I, we replicated the association for the same *FTO* variant (rs9939609), both with regard to BMI and type 2 diabetes. Interestingly, the association between rs9939609 and type 2 diabetes and BMI remained significant after adjustment for BMI and diabetes status, respectively. It is also noteworthy that rs9939609 demonstrated a strong association with triacylglycerol levels, which was not abolished after correcting for diabetes status. Thus, our data suggested that the relation between *FTO* and BMI/diabetes is more complex than initially thought. These findings became the fundament of our next research questions (Paper II).

5.2 LARGE-SCALE META-ANALYSES (PAPER II)

Genetic association studies generally focus on single phenotypes, but associated variants may actually influence multiple traits. In some cases, the phenotypes are correlated, such as the association of several obesity susceptibility loci with type 2 diabetes, obesity and other obesity-related phenotypes [300]. As mentioned in section 1.3.3, the obesity-associated risk variants of *FTO* have shown association with obesity-related conditions and diseases such as the metabolic syndrome, PCOS, and type 2 diabetes [227, 268, 269]. In most of the studies reported, these effects appear to be secondary to weight increase since the associations are attenuated or completely abolished after adjusting for BMI [227].

However, based on our findings in Paper I, we raised the following questions: Could variations in *FTO* alter metabolic traits (in particular type 2 diabetes, HDL and triglycerides levels) independent of its effect on BMI, at least in some other populations? We also noted that the *FTO*-associated allele-wise increase in BMI remained surprisingly constant across all adult age-groups in our cross-sectional sample from HUNT2. Our longitudinal data on weight-gain between 1985 (HUNT1) and 1995 (HUNT2) showed that there were no detectable differences in weight-gain between carriers with high versus low risk genotypes. Hence, we questioned whether the *FTO*-associated BMI difference is established relatively early in life and thereafter stable across adulthood. To seek replication and validation of our initial results from HUNT (Paper I), the findings needed to be tested in independent and large patient materials, preferentially of Scandinavian origin and by using both cross-sectional and longitudinal perspectives.

We therefore initiated a Nordic collaborative project with two Swedish research groups having access to both extensive clinical and genetic data from the MPP and MDC cohorts (for further details see sections 3.1.2 and 3.1.3). Initially, we were interested in performing a pooled analysis of all three datasets using individual level data to properly account for covariates et cetera. Unfortunately, we could not get

access to individual data from the MPP and MDC cohorts. Certainly, access to genetic data across studies is an important aspect of identifying new loci or to validate already reported genetic associations. However, even sharing summary-level data, such as allele frequencies, inherently carries some degree of privacy risk to study participants [301]. As an alternative, we therefore decided to use a meta-analysis approach by performing identical analyses in all three cohorts using exactly the same covariates and methods, followed by meta-analysis of the results.

Meta-analysis across multiple genetic association studies with combined cohort sizes of tens of thousands of individuals often uncovers many more associated loci than the original individual studies or delivers more confident estimates for already reported associations. Hence, in order to circumvent problems with lack of statistical power, meta-analysis has been increasingly applied in genetic epidemiology [90]. Even though larger sample sizes theoretically deliver more confident estimates of association there are in general several pitfalls related to meta-analysis such as publication biases [302], heterogeneity between studies because of overestimation of effect sizes in small studies of low quality [303], and heterogeneity between studies introduced due to confounding by ethnic origin, age, gender, or other measured or unmeasured variables. Therefore, meta-analysis can be a useful tool in dissecting the genetics of complex diseases and traits, provided its methods are properly applied and interpreted [304].

Paper II is one of the largest studies to date investigating the effect of *FTO* sequence variants on type 2 diabetes and BMI across the whole range of adult ages and in a longitudinal perspective. Through a meta-analysis comprising over 41,000 Scandinavian individuals, we demonstrate that a common variant of *FTO* does not mediate type 2 diabetes risk entirely through its influence on BMI. In an attempt to capture the complex relationship between *FTO*, BMI, and type 2 diabetes during the life course, we also performed an analysis on incident type 2 diabetes in over 20,000 non-diabetic individuals followed up for over 10 years. Over 3,000 developed type 2 diabetes during follow-up, and the *FTO* variant was strongly associated with the

incident risk of type 2 diabetes. The results remained similar when we controlled for BMI at baseline (before diabetes was diagnosed), ΔBMI, or waist circumference and/or waist-to-hip ratio as covariates in the regression analyses. None of the covariates alone or in combination with BMI changed our results notably. *FTO* still conferred an increased risk for type 2 diabetes.

Our data therefore contrast previous results reported in most of the populations studied to date, including Europeans [121, 225-227]. Consistent with our finding, Sanghera et al., in a material of North Indian ethnicity, have reported a significant association between type 2 diabetes and *FTO* gene variants that was unaffected by controlling for BMI [305]. Similarly, a study in South Asian Indians also found a strong association of *FTO* variants with type 2 diabetes, which did not seem to be mediated entirely through BMI [306]. The finding that rs9939609 is associated with type 2 diabetes in Asian populations after accounting for BMI and waist circumference has recently been confirmed by two different meta-analysis; one meta-analysis of four studies comprising 8,091 South Asian individuals [307] and in a meta-analysis of 32 populations including 96,551 East and South Asians [308].

Several of the Asian studies suggested that the varying results between Europeans and Asians could be explained by ethnic differences in the way that the *FTO* gene influences susceptibility to diabetes. Scandinavians are considered to display a low level of historical and genetically heterogeneity [309], thus, one could speculate that variation in other environmental factors that contribute to the development of type 2 diabetes could be the reason for the apparently disparate results between Scandinavian and other European populations. Arguing against that this is a spurious finding in whites, however, similar observations have been recorded in white American type 2 diabetes cases in the ARIC study [310]; in 283 subjects with and 2,601 subjects without type 2 diabetes mellitus in the French Multinational MONItoring of Trends and Determinants in CArdiovascular Disease (MONICA) Study [311] and in a partially overlapping sample of LADA patients in the HUNT2 population [312].

A single BMI measurement may not fully capture the effect of the *FTO* gene on total fat mass. One can also probably expect heterogeneity between cohorts as BMI has been steadily increasing in most populations during the last decades and therefore will vary according to when the cohort was characterized. In the current study, the mean BMI in MPP, which started in the 70ies, was lower than in MDC. Considering that a single BMI estimate may not be an accurate measure of obesity in our populations, we applied longitudinal measures of BMI together with waist circumference and/or waist-to-hip ratio as covariates in the regression analyses. We also allowed for a more flexible effect of BMI using second and third order polynomial functions of BMI as covariates. None of the covariates alone or in combination with BMI changed our results notably. Other reasons for the diverging results could be differences in selection or recruitment of cases and control subjects between studies, differences in undetected key effects at early age, or population-specific environmental factors that may interact with the way *FTO* works to influence the risk of type 2 diabetes.

The HUNT2, MDC and MPP samples have previously been validated by genotyping of known type diabetes risk variants indicating that the cohorts contain a representative diabetes cohort and that the genotyping strategies are robust (see Paper I and [188]). In the MDC cohort, only data on *FTO* has been published in the whole material. However, we would like to inform that there is a manuscript in preparation where the rs7903146 in *TCF7L2* was genotyped and analyzed for risk of incident type 2 diabetes. Each additional risk allele of rs7903146 was associated with 36 % (95 % CI: 25-47 %, $p=4 \times 10^{-19}$) increased risk of diabetes which is comparable to the risk estimates observed in other studies. One possible limitation of our study is that diabetes diagnosis was self-reported in the HUNT and MDC cohorts, while confirmed from patient records or based upon a fasting plasma glucose concentration in the MPP cohort. On the other hand, self-reported information has been shown to be a reliable source of information for epidemiological studies focusing on diabetes mellitus [313]. Moreover, the effect of the *FTO* variant on type 2 diabetes risk was stronger in the HUNT cohort than in the MPP and, in particular, the MDC cohorts. It is therefore of relevancy to point out that if data are combined – in a multi-stage

design, a sequential design, or in meta-analyses – a positive result can always be driven by the large effect from the first study only, even if it is unbiased.

How sequence variation in *FTO* could possibly affect type 2 diabetes risk in other forms than through increased adiposity, remains elusive. No studies so far have reported evidence for an association between the *FTO* SNPs and glucose tolerance or insulin sensitivity independent of BMI [152, 200, 227, 314]. Studies have suggested a link between SNPs in *FTO* and inflammation [315, 316] and altered lipid profiles [268], which both are factors that are linked to increased risk for type 2 diabetes. The link between the *FTO* variant and altered lipid levels found in Paper I could however not be confirmed in a meta-analysis reported in Paper II, indicating that these associations might be a spurious finding. Non-fasting measurements for the HUNT2 individuals and/or a population-based sample could be two other explanations why we observe associations with lipids in HUNT2 and not in the HUNT2-MPP-MDC combined analysis. Furthermore, given its expression pattern in relevant brain regions [260], it has been argued that the *FTO* gene could play a role in circadian rhythms [254]. Abnormal circadian rhythms have been shown to affect the risk of type 2 diabetes and other metabolic diseases [317].

A different interesting aspect is that the rs9939609 SNP may affect the primary allelic *FTO* transcript levels [318], and correlations have been observed in peripheral tissues between BMI of tissue donors and *FTO* mRNA expression levels [319]. It is also noteworthy that three recent *FTO* expression studies suggest a BMI-independent role in type 2 diabetes. One study found no association between *FTO* expression and BMI in islet cells [320]. Another study reported an inverse correlation between *Fto* mRNA and glucose in mice after correction for body weight [321]. Finally, a third study found an increase of *FTO* mRNA and protein levels in muscle from type 2 diabetic patients compared with healthy lean control subjects or BMI-matched obese non-diabetic individuals [322]. The latter also suggests that increased *FTO* expression in type 2 diabetic patients contributes to reduced mitochondrial oxidative capacities, lipid accumulation, and oxidative stress, all associated with type 2 diabetes.

Furthermore, it has been suggested that FTO may play a role in epigenetic regulation taking into consideration that FTO have been shown to function as a demethylation enzyme [256, 323] combined with the suggestion of FTO as a transcriptional co-activator that enhances the transactivation potential of the CCAAT/enhancer binding proteins from unmethylated as well as methylation-inhibited promoters [324, 325]. It is also possible that the rs9939609 SNP (or a SNP in strong LD) affects another gene in the region, which has the potential to alter type 2 diabetes risk independently of BMI [326].

Paper II supports the hypothesis that *FTO* is a type 2 diabetes susceptibility gene; however, more evidence and further studies are warranted to get a more complete picture of the effect of the *FTO* gene variation on adiposity and glucose homeostasis. Some directions have been suggested [324]: a) Large-scale type 2 diabetes case-control studies matched at the individual level for age, gender and adiposity measures such as BMI, body fat content and body adiposity index; b) Comparison of incident type 2 diabetes cases to nested controls in large-sized longitudinal cohorts in which careful collection of obesity and type 2 diabetes-related deep phenotypes have been performed. It is important to perform such studies in individuals of different ethnicities due to the fact that the relationship between adiposity and risk for type 2 diabetes varies according to an individual's ethnic background [324, 327]

In Paper II, we confirmed the strong and well-established association between the rs9939609 *FTO* SNP and BMI. We also found significant cross-sectional association between *FTO* and other anthropometric measures such as waist-to-hip ratio and waist and hip circumference. Moreover, both in our initial analysis of the HUNT population and in the meta-analysis, we observed effect sizes and allele frequencies of magnitudes consistent with previous studies conducted in Europeans [121, 225, 226, 228]. Altogether, this strongly confirms the role of *FTO* as a genetic factor influencing which individuals are at risk of becoming obese.

The plethora of studies conducted on *FTO* the last years have shown that the association between *FTO* sequence variants and BMI is not established at birth,

because birth weight and ponderal index (kg/m^3) at birth are not affected by *FTO* variants [226, 328]. The association seems to evolve gradually and becomes detectable before the age of seven [226, 329, 330]. It is not clear how *FTO* genotype affects BMI after adolescence and develops during the life course [331-338], although a recent longitudinal Finnish study suggested that the effect may continue into adulthood since they found an association between rs9939609 and BMI at age 31, which could not be explained by the BMI at age 14 [339]. In the initial analyses performed in the HUNT2 population we observed that the relative difference in mean BMI among individuals with different rs9939609 genotypes remained more or less stable across all adult ages. The non-significant SNP–age interactions observed for all the different obesity-related traits examined in the HUNT2 population further strengthened that change in those traits with age were not dependent on the individual's *FTO* rs9939609 genotype. In accordance with the observations from the HUNT2 cohort, our meta-analysis revealed no heterogeneity across the effect-sizes for the *FTO* risk allele between the different age groups.

On its own, the results from the cross-sectional analyses of Paper II were informative. However, the evidence of change across age could have been inferred from differences between the age groups. The cross-sectional design could also have been affected by the cohort effect where age differences show trends particular to a specific group and not true developmental changes. In this study, we had the unique opportunity to subsequently carry out investigations using longitudinal data analyzing change of BMI over time. We found no evidence that the BMI difference between genotypes increased over time as there was no evidence of heterogeneity across the different adult age groups. In accordance with the previous studies conducted by Hunt et al. [331] and Wangensteen et al [338], our results suggest that genetic susceptibility to weight gain in terms of higher BMI induced by the *FTO* SNP rs9939609 do not increase with age in adults. Hence, because our study primarily comprised individuals that were above 30 years of age (98.7%), current evidence suggest that the *FTO* variant increases BMI in the first 2 to 3 decades of life, and from then on the BMI difference between the genotypes becomes more or less constant throughout life.

In contrast, a recent report from the Genetic Analysis Workshop 16 managing data of 7,130 Caucasian individuals from the Framingham Heart Study showed that genetic susceptibility to weight gain induced by the *FTO* SNP rs9939609 increased with age [334]. The same pattern was observed in a small study comprising adult Filipino women, where the minor allele of rs9939609 was associated with longitudinal BMI, indicating a relatively constant genotype effect over a period of 22 years [335]. In contrast, Jacobsen et al. [332] and Qi et al. [336] observed an opposite pattern in which the association between the FTO SNP and the risk of obesity tended to decline in men at older age. The association was constant across different age groups in women [336]. Although our cross-sectional analyses in the HUNT population could indicate a smaller effect of *FTO* on BMI for individuals between 60 and 80 years, the meta-analysis detected no significant support for a decrease in the associations between rs9939609 and BMI at older age. Furthermore, we detected no significant differences between men and women when analyzed separately.

In a recent longitudinal study of Danish men, the *FTO* risk allele was associated with fatness during early childhood, maintained a constant effect on BMI during later childhood and facilitated a further increase in BMI during adulthood [333]. In our study, we observed no additional weight gain in adulthood attributable to the *FTO* genotype. In other words, in the Scandinavian population *FTO* seems to induce weight gain for a period in early life, and thereafter no further weight gain or loss later in life.

Our study included both males and females, and height and weight were measured in all three cohorts. Participants from both the HUNT and MDC cohorts were part of an all population-inclusive survey with high attendance, which limits possible selection biases. Even though the participation rate in HUNT generally was fairly high compared to most other studies in Norway and abroad [340, 341], there is always a potential selection problem. In HUNT2, data from young age groups, especially in men, should be analyzed with some caution [282]. However, a comprehensive non-participation study after HUNT1 could not find evidence of selection in health

measures in young age groups [342]. Old non-participants, however, had significantly more health problems than participants of the same age [282].

Recent models of *Fto* deficiency in mice support the idea that alleles associated with risk of obesity will cause up- or dysregulation of *FTO* and that inactivation of *FTO* protect against obesity. Furthermore, mice with loss of Fto function appear to have reduced fat mass, mainly due to increased energy expenditure and not decreased energy intake [316, 343]. Interestingly, studies in humans have demonstrated an opposite pattern, indicating a relationship between the *FTO* risk alleles and energy consumption, but not energy expenditure [262]. More recently, it has also been shown that high-fat diets and low physical activity levels may interact and modify the susceptibility to obesity by the *FTO* variants [263-265, 344]. This could be one of the reasons for the discordant results regarding the effect of these variants on BMI in a longitudinal perspective, since type of diets and level of physical activity may have varied between the populations studied and the time the data were collected.

5.3 GENETIC FINE-MAPPING FOR IDENTIFICATION OF FURTHER PHENOTYPIC AND GENETIC COMPLEXITY (PAPER III)

After the identification of a disease-associated region by GWAS, comprehensive studies of sequence variation in the region are essential to identify the full set of variants that might explain the association signal. This is due to the fact that a GWAS is neither a candidate-gene approach nor directly intended to detect causative variants. Rather, GWAS uses SNPs as naturally occurring genetic markers to map genomic risk alleles for the trait in question. Since GWAS arrays do not capture DNA variation in each region completely, it has been assumed that causal variants partially captured by LD (e.g. due to location near recombination hotspots or low MAF) might show stronger association with the phenotype than the tag SNPs used in GWAS. HapMap and GWAS arrays contain primarily variants with MAF >5%. First-generation GWAS studies have therefore generally not tested variants of quite low frequency that might have larger effects on disease risk. In contrast to exact replication, as performed in Paper I, local replication refers to the analysis of the original marker(s) detected in the initial report plus other markers in the same region that were not part of the original study. This fine-mapping approach plays an important role in the subsequent steps of GWA studies to assemble a more complete catalog of variation present in an associated region and to test it for association with the phenotype of interest when it is assumed that the associated marker is not the causal variant and possibly not even the variant that best tags it [345].

Non-coding variants at human chromosome 9p21 near *CDKN2A* and *CDKN2B* have repeatedly been associated with type 2 diabetes [120, 121, 124, 172], myocardial infarction [144, 145, 346], aneurysm [147], stroke [148] and at least five types of cancers [347-352]. Hence, this region appears as one of the most interesting regions to emerge from the first-generation of GWAS. In Paper I we demonstrated the significance of the SNP rs10811661 located close the *CDKN2B* gene with increased odds of having type 2 diabetes in the HUNT2 population. The purpose of Paper III

was to explore this finding by performing a fine-map approach. We attempted to dissect the genetic complexity of a 138-kb interval on Chr 9p21 with respect to type 2 diabetes and cardiovascular disease, and subsequently to investigate the relation between the association signals.

By analyzing 19 SNPs, Paper III to some extent highlights the genetic complexity of the chromosome 9p21 region. The significant association between marker rs10811661 and type 2 diabetes, was in agreement with our previous results obtained for this marker in more or less identical cases and controls from the HUNT2 population (Paper I). However, in Paper III, we revealed a haplotype association tagged by rs10811661 and rs10757282 that was more strongly associated with type 2 diabetes than either SNP alone. This finding suggests that these SNPs might tag a rarer risk haplotype harboring a causative allele or that the candidate region harbors several causal polymorphisms.

We found no evidence of additional type 2 diabetes association signals elsewhere in the 9p21 region as suggested by some [124], but not all first-generation GWASs on type 2 diabetes [120, 121]. The status of rs564398 as a type 2 diabetes susceptibility variant appears to be largely driven by the WTCCC scan and UK data since several other studies and replication efforts, including Paper III, have turned out negative for this variant [120, 150]. A meta-analysis conducted on type 2 diabetes and polymorphisms on chromosome 9p21 demonstrated an overall OR of only 1.08 for the T allele of the SNP rs564398 [353]. Thus, the failure to replicate rs564398 as a type 2 diabetes susceptibility variant in HUNT2 may possibly be due to lack in statistical power to detect association or to the fact that a different population was studied.

The risk variants identified in the 9p21 interval by GWASs are in general located in non-coding regions, suggesting that their effects probably are mediated by influences on gene expression of nearby genes in *cis*, since most reported risk variants do not appear in mature transcripts, and there are no known micro-RNAs mapping to this region [168]. On the other hand, if numerous cis-acting effects are present at a locus,

determination of a disease association by fine-mapping may not be possible. Investigation of gene expression rather than disease phenotype increases the power to map cis-acting effects [168], and should perhaps therefore be considered as an better fine-map approach for this region. Strong LD, however, might limit the ability to separate the potential individual effects that SNPs could have on expression. Thus, such a fine-mapping approach would be most appropriate to perform in African populations, since Africans have less LD in this region compared to Caucasians [167, 168, 354].

In Paper III, we further confirmed the associations for SNPs in the 9p21 region with MI, angina pectoris and any coronary heart disease, among which rs1333040 and rs10757278 demonstrated the strongest associations. The associations were driven by those subjects having both MI and angina, probably validating the underlying diagnosis of coronary heart disease, but this could also be a marker of the severity of the disease. Expression studies of the *INK4*/*ARF* locus ($p15^{INK4b}$, $p16^{INK4a}$, *ARF* and *CDKN2BAS*) have highlighted the rs10757278 SNP as a potential casual variant for the association with coronary artery disease [169]. Moreover, this marker is further mapped to one of 33 newly identified enhancers in the 9p21 interval, in which the risk variant disrupts a transcription factor binding site, thus having functional relevance for an atherosclerosis associated pathway in human endothelial cells [170].

We performed association analyses conditioning on the lead SNPs in the region and observed two independently and potential single SNP association signals for MI, which are located close to, but not in LD with the former and well-confirmed cardiovascular disease-associated region. One of these signals, the rs3217986 SNP, is located in the 3’ UTR of the *CDKN2B* gene as well as in intron 1 of the non-protein coding *CDKN2B* antisense RNA, *CDKN2BAS*. It is therefore tempting to hypothesize that the rs3217986 risk variant exerts its effect on MI susceptibility by influencing expression of one or both of these two genes. At present, there are no reports on whether the rs3217986 risk variant is correlated with expression of *CDKN2B* and/or *CDKN2BAS,* thus this hypothesis needs further examination. However, we have

realized that more in-depth haplotype analyses, as performed for type 2 diabetes, should also have been performed for the multiple cardiovascular outcomes of Paper III. It is important to investigate whether there are haplotypes that are more significantly associated with cardiovascular disease than the single top GWA SNPs rs1333040 and rs10757278, since the aim of this study was to explore and find better tags of the causal variant and not only to test the association of the region. Although rs10757278 is suggested to be casual for some diseases, it is still not certain for all the cardiovascular diseases.

We found no association between SNPs of the 9p21 region and stroke, as suggested by a previous GWASs [148]. One could anticipate that SNPs in this interval associate with ischemic stroke, but not hemorrhagic stroke, since studies have indicated that sequence variation on chromosome 9p21 influences atherosclerosis development and progression [164]. Participants of the HUNT2 survey were identified having stroke through a self-administered questionnaire, hence details regarding type of stroke (hemorrhagic versus ischemic), or subtypes (atherothrombotic versus cardioembolic) were not available. This could be one explanation why we failed to confirm the associations between variants in the region and stroke. Interestingly, in a sub-analysis some SNPs demonstrated tentative associations with stroke in subjects with type 2 diabetes, but not in those without type 2 diabetes. In support of this, a significant interaction has been found between poor glycemic control and a variant within the 9p21 region on the risk of coronary heart disease [166].

Despite their close proximity, there seems to be no apparent overlap between the CVD and type 2 diabetes risk regions. Theories with reference to the disease mechanism mediated by the causative risk variants of the 9p21 interval have increased in numbers during the last years. However, since most of them still remain exploratory, the exact nature of the causative variants and the target proteins or genes is still somewhat elusive. The disease mechanism and the causative risk variants may possibly differ between CVD and type 2 diabetes.

Paper III has several limitations. Data on MI, angina, stroke and diabetes were self-reported, not medically confirmed. There are, however, studies that support the agreement between medical records and self-reported MI, stroke and diabetes [313, 355, 356]. In the present manuscript, we estimated the odds ratios based on additive effects of allele dosage for all SNPs. It would have been relevant to test whether the results remained similar under a recessive or dominant model. The P-values presented are point-wise estimates of an individual SNPs significance and based on 1000 permutations implemented in the regression models. While certain types of permutation procedures can correct for multiple comparisons, they do not do so by definition. To properly account for multiple comparisons in the current study, the observed test statistic should have been compared to the permuted statistics of all SNPs in each replicate. Thus, our results are not corrected for multiple comparisons.

For our detailed analysis of the 9p21 region, we selected 18 SNPs tagging a 138 kb interval using the Haploview implementation of the Tagger algorithm [286] applying the following criteria: MAF of >5% and pairwise r^2 >0.80, according to HapMap CEU LD data. In addition, we added two previously GWAS-identified type 2 diabetes susceptibility variants (rs564398 and rs10811661) and three confirmed CVD susceptibility variants (rs1333040, rs10757278 and rs1333049) to the SNP set, making a total of 23 SNPs. Of these, 19 SNPs were included in the subsequent analyses since four SNPs did not pass quality control criteria. Thus, our tagging of the 138 kb region is not best possible given the abovementioned criteria, owing to the fact that we did not include other SNPs to replace the four excluded SNPs. Furthermore, for fine-mapping purposes, investigators have been more relaxed about MAF >5%, given our knowledge of the importance of rarer variants.

There are several approaches that would have been better suited for a more comprehensively assessment of the genetic variation in the region such as imputation of variants directly from targeted sequencing, from a genotyped reference panel derived from sequencing or from 1000 Genomes Project low-coverage data. A recent study conducted by Shea and colleagues compared different strategies to fine-map the

association of common SNPs at chromosome 9p21 with type 2 diabetes and myocardial infarction [345]. By applying the more sophisticated fine-mapping approaches listed above, they did not find evidence for stronger association at chromosome 9p21 to SNPs in moderate LD with the initial tag SNPs. Moreover, they did not observe lower-frequency variants with effect sizes that could individually explain the common variant associations, but they did identify additional common variants in LD with the SNPs with the strongest GWAS signals that could underlie each association. These additional common variants are important to identify, since a complete list of all variants that might explain the association signals in the region will be attractive for subsequent functional studies aiming to understand how non-coding variants at 9p21 can lead to such varied and clinically relevant phenotypic associations [345].

5.4 DISCOVERY OF DISEASE RISK GENES BY DEFINING THE GENETIC BACKGROUND OF INTERMEDIARY PHENOTYPES (PAPER IV)

Definition and accuracy of the phenotype is one of the key issues in designing any genetic study for which the goal is to detect genetic susceptibility variants. To increase the chance of finding relevant susceptibility genes for complex diseases, one may have to look more narrowly and start dissecting them according to their pathophysiology or clinical characteristics. Most multifactorial diseases can be considered as cases, a binary definition versus controls. Some of them can also be characterized by or decomposed into descriptive quantitative traits, frequently called partial phenotypes or intermediate phenotypes. Most physiological systems have a hierarchical component to them, leading from the gene to its product, to intermediate phenotypes of greater complexity, to the ultimate phenotypes used to diagnose disease (Figure 9).

Especially obesity and type 2 diabetes can by far be approached through the quantification of descriptive traits, which could further be used as quantitative traits for genetic analysis. In contrast to genome-wide searches, genetic analysis of an intermediate or low-level phenotype is eased by greater proximity to the genetic variant, thus reducing the effects of other factors surrounding and possibly influencing the effect of individual genes [357].

Studying intermediate or low-level phenotypes can also bridge the gap between the two strategies of modern complex disease research: the “top-down” approach, linking complex phenotypes with genotypes, and the “bottom-up” approach, which starts with the gene of interest and then works up to the complex phenotype [358]. GWAS are basically a top-down approach, whereas most candidate gene studies are typical of the bottom-up approach. Trying to relate common genetic variants to intermediate phenotypes is an attempt to make sense of what is going on between the top and the bottom. Studying intermediate phenotypes may also help to get by both epistatic

interaction between genes and pleiotropy [357]. This approach has now been widely used in genome scans for new susceptibility loci for type 2 diabetes and obesity. It is important to bear in mind that the intermediate phenotype can be implicated in both the causal mechanisms and be a result of the disease status. Nevertheless, it is also important that the preferred intermediate phenotype relates to the ultimate phenotype of interest.

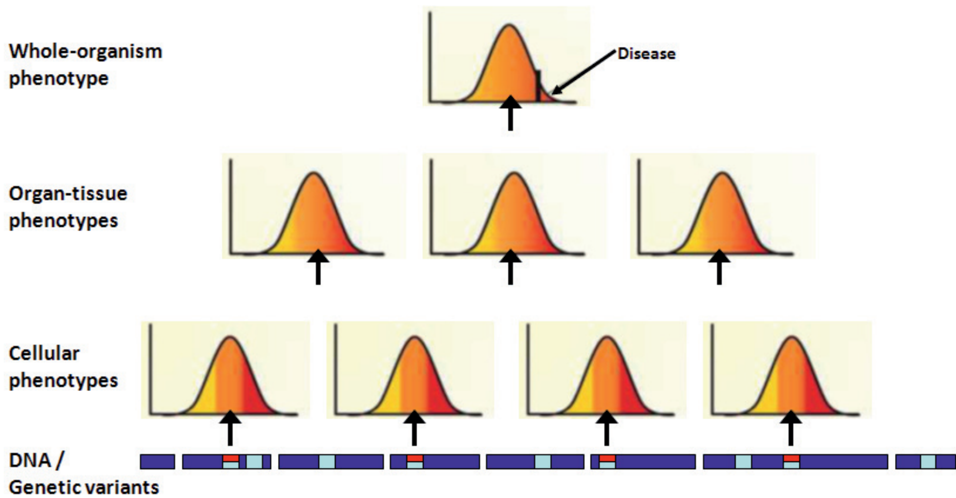

Figure 9 The effect of common genetic variants on different hierarchical components of a physiological system. Yellow to red gradients in the eight normal distributions represent the effect of each of the different red and blue common unlinked genetic variants at the different ends of the spectrum of the phenotypes. The power to investigate the genetic variants associated with the phenotypes decreases as one goes from cellular phenotypes to the ultimate phenotypes used to diagnose disease. Redrawn and modified from [359].

Poor glucose control (expressed as high HbA1c) is a well-known risk factor for long-term diabetic complications and has for a long time been recognized as an interesting intermediate phenotype in relation to diabetes. There is variability in glycemic control both between and within individuals, and HbA1c have a heritability estimate that makes it as a phenotype suited for genetic analysis [236, 237]. Initiatives dissecting

the genetic influences of HbA1c levels will be important not only to better understand the genetic and biologic determinants of HbA1c variation in the general population, but also to inform the healthcare system to focus diabetes diagnosis and care more centrally on HbA1c [242].

Linkage studies of measures of glycemia (e.g. fasting glucose or HbA1c), predominantly in non-diabetic individuals, did not lead to the identification of novel genes. More recently, several GWA scans on fasting glucose in both diabetic and non-diabetic subjects have shown that genetic studies of glycemic traits can identify type 2 diabetes risk loci, as well as loci containing gene variants that are associated with a modest elevation in glucose levels [179-182, 238-240]. Several loci have also reached evidence for association with HbA1c in diabetic individuals, as well as in non-diabetic subjects [242-245].

To identify loci for glycemic control in persons with type 1 diabetes, Paterson and colleagues [241] analyzed longitudinal repeated measures of HbA1c from the Diabetes Control and Complications Trial (DCCT). They performed a GWAS using the mean of quarterly HbA1c values measured over 6.5 years, separately in the conventional (n=667) and intensive (n=637) treatment groups of the DCCT. They identified a major locus for HbA1c levels in the conventional treatment group near the *SORCS1* gene, which was also associated with mean glucose. The same locus was confirmed using HbA1c in the intensive treatment group. In addition, three other loci, namely 14q32.13 (*GSC*), 9p22 (*BNC2*), 15q21.3 (*WDR72*) achieved evidence close to genome-wide significance in the intensive group. All four loci were carried forward for replication in two independent replication samples. These included the GoKinD study, a case–control collection of patients with type 1 diabetes with and without diabetic nephropathy, and healthy subjects from the MAGIC meta-analysis. The *SORCS1* association was replicated in GoKinD controls and the *BNC2* association with HbA1c was replicated in non-diabetic individuals from MAGIC. Both SNPs were also associated with diabetes complications in the expected direction; *SORCS1*

with hypoglycemia (and less robustly with both retinopathy and nephropathy) and *BNC2* with microvascular end points [241, 242]

Although a major locus for HbA1c levels in type 1 diabetes was identified, a number of questions remained. These included evaluation in other groups such as non-white individuals, type 1 diabetes not meeting DCCT eligibility criteria, and type 2 diabetes. In either case, it appears that at least the top signals reported from the DCCT study deserve further study [241, 360].

Paper IV is the first attempt to evaluate the effect of the SNPs found by Paterson [241] with regard to glycemic control in type 2 diabetes. We typed 1486 subjects with type 2 diabetes from the HUNT2 cohort, and subsequently assessed the effect of the four top signals reported from the DCCT study on HbA1c and non-fasting glucose levels individually and in a combined genetic score by summing the number of risk alleles carried by each patient. We detected no significant associations with HbA1c or glucose individually or by using the genetic score model and partially inconsistent direction of associations – the *WDR72* SNP showed a borderline association with reduced and not increased HbA1c levels in our study. The different pathophysiology between type 1 and type 2 diabetes could be one explanation why our results did not support that the four investigated loci are genetic susceptibility factors for glycemic control in type 2 diabetes.

There are some prior data suggesting a role of the *SORCS1* gene in glycemic traits. *SORCS1* encodes a sortilin-related vacuolar protein sorting 10 domain-containing receptor, which binds to platelet-derived growth factor. A quantitative trait locus for fasting insulin in the syntenic region in mice has been described [361], with further independent evidence obtained in rats for post-intra-peritoneal glucose tolerance [362]. Two studies have also demonstrated modest evidence for association between SNPs in *SORCS1* and fasting insulin, insulin sensitivity and insulin resistance in humans [125, 363]. However, no association has been found with type 2 diabetes. Considering our results in light of the previously reported results and features for

SORCS1, we can not refute a possible link between common variants in *SORCS1* and glycemic control in type 2 diabetes.

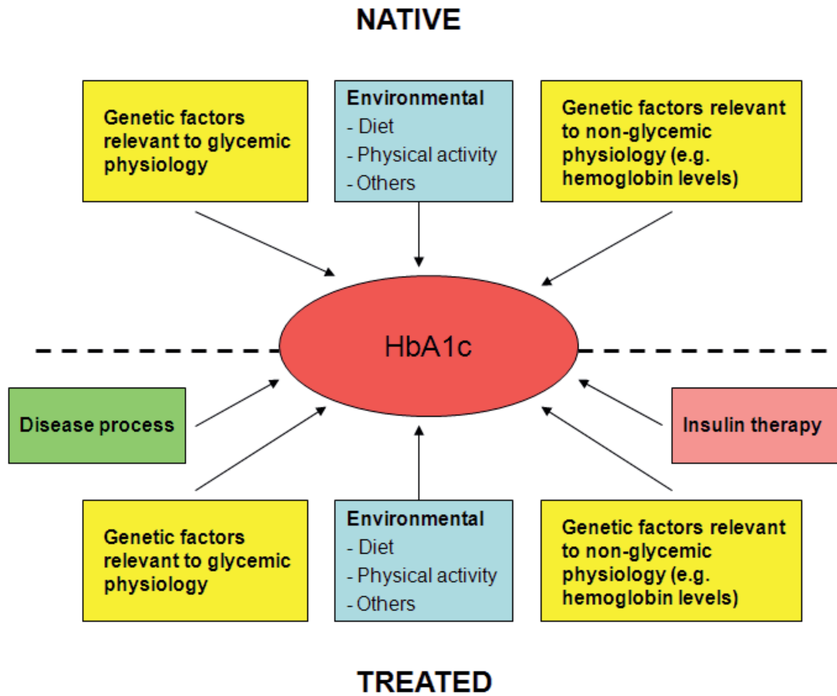

Figure 10 A schematic illustration exemplifying the potential contributors to HbA1c levels in the native (top) or treated (bottom) states. Modified and redrawn from [360].

In this regard, it should be noted that there are several limitations of our study. We had access to only one HbA1c and non-fasting blood glucose value for each case, in contrast to the repeated measurements used by the DCCT investigators during the course of a carefully controlled clinical trial. Furthermore, the use of HbA1c as a quantitative trait modulated by genetic factors must be taken with caution in the context of pharmacological treatment, since treatment as an environmental variable may overwhelm the genetic signal [360]. This is, above all, true for insulin as a therapeutic modality. Except for hypoglycemia and some other practical limitations,

insulin dosage can be gradually increased until the target HbA1c has been reached. Thus, higher insulin doses should be able to overcome most genetic barriers against lowering HbA1c [360]. Thus, for treated HbA1c, which was the case for most of our diabetic patients, the non-genetic contributions to the variability in the trait plays a much larger role than that in the native state (Figure 10). Whereas the DCCT investigators attempted to control for this, we, on the other hand, had no access to information on medical treatment in the current study. Thus, our data may be confounded by environmental factors and cannot be considered a straight-forward replication study.

In the future, it will be important to expand the study of genetic influences on HbA1c to pre-diabetic and diabetic populations, even though the confounding effects of treatment might obscure any role of these polymorphisms in the diabetic population. Additionally, genetic associations may be revealed from studies of low-to-intermediate frequency variants through imputation from the 1000 Genomes Project, direct association using whole-genome sequencing data, and in-depth replication and locus fine-mapping. These genetic efforts will hopefully contribute to the detection of new loci involved in hemoglobin glycation, glucose metabolism, and diabetes [242].

"One can prefer to be an optimist or a pessimist, but the best approach is to be an empiricist."

- Eric S. Lander,
Leader of the international Human Genome Project

5.5 CONCLUDING REMARKS AND FUTURE PERSPECTIVES

Throughout Paper I - IV, we have shown that SNPs near *IGFBP2*, *CDKAL1*, *SLC30A8*, *CDKN2B*, *HHEX* and *FTO* are associated with diabetes in non-selected patients with type 2 diabetes from a population-based, Norwegian sample (the HUNT2 survey). Furthermore, we have identified that a variant in *FTO* alters type 2 diabetes risk partly independent of its observed effect on BMI, and that the additional weight gain as a result of the *FTO* risk variant seems to occur before adulthood, and the BMI difference remains stable thereafter. Through a fine-mapping approach of the 9p21 region we confirmed the association of variants with type 2 diabetes and coronary heart disease. We also identified one haplotype strongly associated with type 2 diabetes and two potentially novel and independent signals associated with CVD in the region. Finally, we have shown that four recently reported loci affecting glycemic control in type 1 diabetes patients had no apparent effect on HbA1c levels in type 2 diabetes, neither individually nor by using a combined genetic score model.

Despite the stimulating success of the recent SNP-based genome scans, the results of studies investigating specific complex diseases such as type 2 diabetes, indicate that the approach frequently identifies common variants that account for only a small fraction (< 10%) of the heritable component of the disease. Although these studies have provided some new biological insights, only a limited amount of the heritable factor of any complex trait has been identified and it remains a challenge to clarify the functional link between associated variants and phenotypic traits. It is, nevertheless, crucial to continue to perform in-depth follow-up studies for these and future susceptibility loci in unselected samples of patients and control participants,

since inclusion criteria based on age of onset, family history of the trait and BMI may affect the type of loci detected.

Several resources and procedures that are considered necessary to progress from current findings of GWA studies have been highlighted [364]: a) even larger samples from various populations of different ethnicities for multiple diseases or traits; b) a more complete synopsis and better knowledge of common variation across the genome in different populations; c) improved technical and analytical methods for investigation of structural variation (e.g CNVs) in large samples; d) improved next-generation sequencing strategies for detection and further investigation of low frequency variation; e) development of novel and more sophisticated statistical and computational methods for analysis of high-throughput sequence data from large samples; f) collection of spatio-temporal gene expression data from densely genotyped human samples to enable search for associations between genetic variants and gene expression; g) improved genome annotation, particularly of non-coding regions, since SNPs usually occur in non-coding regions more frequently than in coding regions; h) applicable functional assays for associated genes; i) advanced animal models or suitable *in vitro* models, wherein possible causal variants can be evaluated; j) large cohort studies with available DNA samples and coordinated measurements of environmental exposures and disease outcomes; k) improved analytical methods for comprehensive assessment of gene×gene and gene×environment effects; and l) methods for evaluation of the role of epigenetics in the inherited risk of disease.

A recent interesting study adapted the techniques used in GWA studies, and instead conducted a pilot environment-wide association (EWA) study to consider 266 separate environmental factors with diabetes [365]. The factors most associated with diabetes had size effects on type 2 diabetes that are comparable to the highest risk gene loci found in GWA studies. Future studies might therefore benefit from combining GWA and EWA data and methodologies, to consider the combined effects of genes and environment [365].

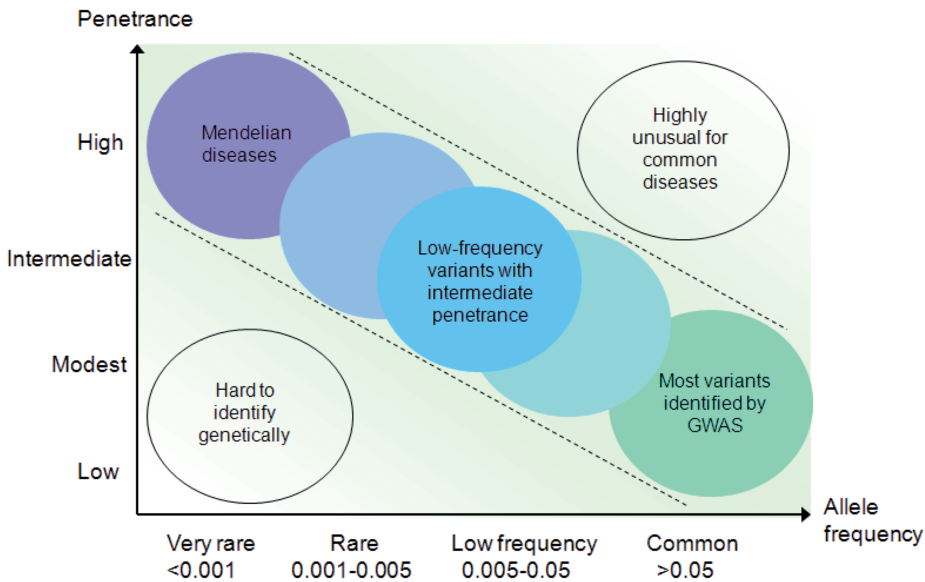

Figure 11 Possibility of identifying genetic variants by risk allele frequency and penetrance (strength of genetic effect). Frequency and the effect are essential factors when it comes to the likelihood of detecting genetic variants. Most emphasis lies in identifying genetic associations with characteristics shown within the diagonal lines. Future research adapting next-generation sequencing strategies will in particular focus on identifying low-frequency variants with intermediate penetrance. Adapted from [366, 367].

For diabetes, obesity and lipid disorders there are several examples of rare mutations causing monogenic forms of disease and more common at-risk variants in the same gene [219, 368]. Several genes implicated in MODY has been validated as type 2 diabetes genes through GWAS (see Table 2 in section 1.3.2). Mutations in the *HNF1A* gene are the most common cause of MODY, and there is a substantial variation in the age at diabetes diagnosis, even within families where diabetes is caused by the same mutation. In a recent study we therefore investigated the hypothesis that common polygenic variants that predispose to type 2 diabetes might account for the difference in age at diagnosis in *HNF1A*- MODY [369]. In a sample of 410 individuals with a known mutation in the *HNF1A* gene, we assessed the effect

of 15 robustly associated type 2 diabetes variants on the age at diagnosis both individually and in a combined genetic score. We observed that each risk variant for type 2 diabetes was found to lower the age of MODY onset by 0.35 years, independent of other genetic and environmental modifiers. Thus, this is one of the first studies to demonstrate that clinical characteristics of a monogenic disease (i.e. MODY) can be influenced by common variants that predispose to the polygenic form of that disease (i.e., common type 2 diabetes). There will definitely be more examples to come of polygenic risk variants of complex diseases influencing the phenotype of the corresponding monogenic diseases [369, 370].

The current generation of genome-wide association studies tags common SNPs to some extent adequately, but is less efficient in detecting SNPs of lower allele frequency or in technologically difficult genomic regions [366]. These limitations have been widely recognized and it now seems more likely that each common disease is mostly caused by large numbers of rare variants, ones too rare to have been catalogued by the HapMap. Recent advances in next-generation sequencing technologies could rapidly facilitate substantial progress. This conclusion is based on the assumption that much of the missing genetic control is due to gene variants that are too rare to be picked up by GWA studies and have relatively large effects on risk [371] (Figure 11). Both large-scale exome-sequencing and whole-genome sequencing are currently being performed in numerous individuals of different ethnicities and will without doubt contribute to identification of rare variants and serve as a reference for other studies.

As an example, genome-wide association studies have revealed that common non-coding variants in *MTNR1B* gene increase type 2 diabetes risk [179, 180]. Although the strongest association signal was highly significant, its contribution to type 2 diabetes risk was only modest (OR of ~1.10–1.15). By performing large-scale exon re-sequencing in 7,632 Europeans, including 2,186 individuals with type 2 diabetes, Bonnefond and colleagues [372] identified 40 non-synonymous variants, including 36 very rare variants (MAF <0.1%), associated with type 2 diabetes. Four of these were

very rare with complete loss of melatonin binding and signaling capabilities. This study establishes a solid link, both genetically and functionally, between the *MTNR1B* gene and type 2 diabetes risk, and further highlights the importance of rare variants in health and disease.

Another way to identify rare genetic variants with expectedly large effects is to study families with particular accumulation of disease. Genetic variations that cause disease will occur with much higher frequency in the affected relatives than in a reference population. Exome-sequencing will then be an appropriate approach for tracking down the gene variations, as recently demonstrated for the *MITF* gene and malignant melanoma [373]. Sequencing of families with extreme quantitative traits could also be an important next step in the dissection of the genetics of type 2 diabetes.

Finally, most genetic research has an overall aim of translating the findings into progression in clinical care. The mechanistic insights generated by gene discovery might identify new therapeutic targets and lead to novel pharmaceutical and preventative approaches. In addition, there is a growing belief that individual patterns of genetic predisposition will be valuable in health-care delivery and contribute to the development of a more personalized medicine. However, the development of personalized medicine beyond monogenic diseases expects a more complete picture of the genetic predisposition and the disease mechanisms. Hopefully, in a few years large-scale genome-wide re-sequencing efforts have provided a systematic and a more complete description of the associations between genome sequence variation and major clinical phenotypes, thus facilitating the use of personalized medicine [23].

References

1 Permutt MA, Wasson J, Cox N. Genetic epidemiology of diabetes. *J Clin Invest* 2005; **115:** 1431-9.

2 Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2011; **34 Suppl 1:** S62-9.

3 Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med* 1998; **15:** 539-53.

4 Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Res Clin Pract* 2009; **87:** 4-14.

5 Midthjell K, Kruger O, Holmen J, Tverdal A, Claudi T, Bjorndal A, Magnus P. Rapid changes in the prevalence of obesity and known diabetes in an adult Norwegian population. The Nord-Trondelag Health Surveys: 1984-1986 and 1995-1997. *Diabetes Care* 1999; **22:** 1813-20.

6 Stene LC, Midthjell K, Jenum AK*, et al.* [Prevalence of diabetes mellitus in Norway]. *Tidsskr Nor Laegeforen* 2004; **124:** 1511-4.

7 Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 1997; **20:** 1183-97.

8 Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 2003; **26 Suppl 1:** S5-20.

9 WHO/International Diabetes Federation. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia: report of a WHO/IDF consultation. 2006.

10 Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus. Abbreviated report of a WHO consultation. 2011**:** 25.

11 Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. *Lancet* 1998; **352:** 837-53.

12 Khaw KT, Wareham N. Glycated hemoglobin as a marker of cardiovascular risk. *Curr Opin Lipidol* 2006; **17:** 637-43.

13 Pradhan AD, Rifai N, Buring JE, Ridker PM. Hemoglobin A1c predicts diabetes but not cardiovascular disease in nondiabetic women. *Am J Med* 2007; **120:** 720-7.

14 Farmer A, Fox R. Diagnosis, classification, and treatment of diabetes. *BMJ* 2011; **342:** d3319.

15 Pozzilli P, Buzzetti R. A new expression of diabetes: double diabetes. *Trends Endocrinol Metab* 2007; **18:** 52-7.

16 Tuomi T. Type 1 and type 2 diabetes: what do they have in common? *Diabetes* 2005; **54 Suppl 2:** S40-5.

17 van Deutekom AW, Heine RJ, Simsek S. The islet autoantibody titres: their clinical relevance in latent autoimmune diabetes in adults (LADA) and the classification of diabetes mellitus. *Diabet Med* 2008; **25:** 117-25.

18 Lin J, Zhou ZG, Wang JP, Zhang C, Huang G. From Type 1, through LADA, to type 2 diabetes: a continuous spectrum? *Ann N Y Acad Sci* 2008; **1150:** 99-102.

19 Fourlanos S, Dotta F, Greenbaum CJ, Palmer JP, Rolandsson O, Colman PG, Harrison LC. Latent autoimmune diabetes in adults (LADA) should be less latent. *Diabetologia* 2005; **48:** 2206-12.

20 Pociot F, Akolkar B, Concannon P*, et al.* Genetics of type 1 diabetes: what's next? *Diabetes* 2010; **59:** 1561-71.

21 Barrett JC, Clayton DG, Concannon P*, et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 2009; **41:** 703-7.

22 Gonzalez EL, Johansson S, Wallander MA, Rodriguez LA. Trends in the prevalence and incidence of diabetes in the UK: 1996-2005. *J Epidemiol Community Health* 2009; **63:** 332-6.

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

24 Fajans SS, Bell GI, Polonsky KS. Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N Engl J Med* 2001; **345:** 971-80.

25 Murphy R, Ellard S, Hattersley AT. Clinical implications of a molecular genetic classification of monogenic beta-cell diabetes. *Nat Clin Pract Endocrinol Metab* 2008; **4:** 200-13.

26 Online Mendelian Inheritance in Man, OMIM. Baltimore, MD. MIM Number: 606391: Accessed 09.23.2011: Johns Hopkins University.

27 Eide SA, Raeder H, Johansson S, Midthjell K, Sovik O, Njolstad PR, Molven A. Prevalence of HNF1A (MODY3) mutations in a Norwegian population (the HUNT2 Study). *Diabet Med* 2008; **25:** 775-81.

28 Njolstad PR, Hertel JK, Sovik O, Raeder H, Johansson S, Molven A. [Progress in diabetes genetics]. *Tidsskr Nor Laegeforen* 2010; **130:** 1145-9.

29 De Leon DD, Stanley CA. Permanent Neonatal Diabetes Mellitus. In: Pagon RA, ed, *GeneReviews*. Seattle (WA): University of Washington, Seattle. 1993-2011.

30 Njolstad PR, Sagen JV, Bjorkhaug L*, et al.* Permanent neonatal diabetes caused by glucokinase deficiency: inborn error of the glucose-insulin signaling pathway. *Diabetes* 2003; **52:** 2854-60.

31 Pearson ER, Flechtner I, Njolstad PR*, et al.* Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. *N Engl J Med* 2006; **355:** 467-77.

32 Sagen JV, Raeder H, Hathout E*, et al.* Permanent neonatal diabetes due to mutations in KCNJ11 encoding Kir6.2: patient characteristics and initial response to sulfonylurea therapy. *Diabetes* 2004; **53:** 2713-8.

33 Jones CW. Gestational diabetes and its impact on the neonate. *Neonatal Netw* 2001; **20:** 17-23.
34 Metzger BE, Lowe LP, Dyer AR*, et al.* Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med* 2008; **358:** 1991-2002.
35 Lobner K, Knopff A, Baumgarten A*, et al.* Predictors of postpartum diabetes in women with gestational diabetes mellitus. *Diabetes* 2006; **55:** 792-7.
36 Boney CM, Verma A, Tucker R, Vohr BR. Metabolic syndrome in childhood: association with birth weight, maternal obesity, and gestational diabetes mellitus. *Pediatrics* 2005; **115:** e290-6.
37 Eckel RH, Kahn SE, Ferrannini E*, et al.* Obesity and type 2 diabetes: what can be unified and what needs to be individualized? *Diabetes Care* 2011; **34:** 1424-30.
38 Wang J, Luben R, Khaw KT, Bingham S, Wareham NJ, Forouhi NG. Dietary energy density predicts the risk of incident type 2 diabetes: the European Prospective Investigation of Cancer (EPIC)-Norfolk Study. *Diabetes Care* 2008; **31:** 2120-5.
39 Hectors TL, Vanparys C, van der Ven K*, et al.* Environmental pollutants and type 2 diabetes: a review of mechanisms that can disrupt beta cell function. *Diabetologia* 2011; **54:** 1273-90.
40 Anderson JW, Kendall CW, Jenkins DJ. Importance of weight management in type 2 diabetes: review with meta-analysis of clinical studies. *J Am Coll Nutr* 2003; **22:** 331-9.
41 Schulze MB, Heidemann C, Schienkiewitz A, Bergmann MM, Hoffmann K, Boeing H. Comparison of anthropometric characteristics in predicting the incidence of type 2 diabetes in the EPIC-Potsdam study. *Diabetes Care* 2006; **29:** 1921-3.
42 Pierce M, Keen H, Bradley C. Risk of diabetes in offspring of parents with non-insulin-dependent diabetes. *Diabet Med* 1995; **12:** 6-13.
43 Shai I, Jiang R, Manson JE, Stampfer MJ, Willett WC, Colditz GA, Hu FB. Ethnicity, obesity, and risk of type 2 diabetes in women: a 20-year follow-up study. *Diabetes Care* 2006; **29:** 1585-90.
44 Nichols GA, Hillier TA, Brown JB. Progression from newly acquired impaired fasting glusose to type 2 diabetes. *Diabetes Care* 2007; **30:** 228-33.
45 Gress TW, Nieto FJ, Shahar E, Wofford MR, Brancati FL. Hypertension and antihypertensive therapy as risk factors for type 2 diabetes mellitus. Atherosclerosis Risk in Communities Study. *N Engl J Med* 2000; **342:** 905-12.
46 Mooradian AD. Dyslipidemia in type 2 diabetes mellitus. *Nat Clin Pract Endocrinol Metab* 2009; **5:** 150-9.
47 Hu FB, Manson JE, Stampfer MJ, Colditz G, Liu S, Solomon CG, Willett WC. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N Engl J Med* 2001; **345:** 790-7.
48 Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. *Diabetes Care* 2002; **25:** 1862-8.
49 Whincup PH, Kaye SJ, Owen CG*, et al.* Birth weight and risk of type 2 diabetes: a systematic review. *JAMA* 2008; **300:** 2886-97.

50 Legro RS. Type 2 diabetes and polycystic ovary syndrome. *Fertil Steril* 2006; **86 Suppl 1:** S16-7.
51 DECODE Study Group: Age- and sex-specific prevalences of diabetes and impaired glucose regulation in 13 European cohorts. *Diabetes Care* 2003; **26:** 61-9.
52 Dechenes CJ, Verchere CB, Andrikopoulos S, Kahn SE. Human aging is associated with parallel reductions in insulin and amylin release. *Am J Physiol* 1998; **275:** E785-91.
53 Bloomgarden ZT. Type 2 diabetes in the young: the evolving epidemic. *Diabetes Care* 2004; **27:** 998-1010.
54 International Diabetes Federation. IDF Diabetes Atlas, 4th edn. Brussels, Belgium: International Diabetes Federation. 2009.
55 Boden G. Fatty acids and insulin resistance. *Diabetes Care* 1996; **19:** 394-5.
56 Wei Y, Chen K, Whaley-Connell AT, Stump CS, Ibdah JA, Sowers JR. Skeletal muscle insulin resistance: role of inflammatory cytokines and reactive oxygen species. *Am J Physiol Regul Integr Comp Physiol* 2008; **294:** R673-80.
57 Rabe K, Lehrke M, Parhofer KG, Broedl UC. Adipokines and insulin resistance. *Mol Med* 2008; **14:** 741-51.
58 Kim JA, Wei Y, Sowers JR. Role of mitochondrial dysfunction in insulin resistance. *Circ Res* 2008; **102:** 401-14.
59 Robertson RP, Harmon J, Tran PO, Tanaka Y, Takahashi H. Glucose toxicity in beta-cells: type 2 diabetes, good radicals gone bad, and the glutathione connection. *Diabetes* 2003; **52:** 581-7.
60 Robertson RP, Harmon J, Tran PO, Poitout V. Beta-cell glucose toxicity, lipotoxicity, and chronic oxidative stress in type 2 diabetes. *Diabetes* 2004; **53 Suppl 1:** S119-24.
61 Hull RL, Westermark GT, Westermark P, Kahn SE. Islet amyloid: a critical entity in the pathogenesis of type 2 diabetes. *J Clin Endocrinol Metab* 2004; **89:** 3629-43.
62 Stumvoll M, Goldstein BJ, van Haeften TW. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* 2005; **365:** 1333-46.
63 Haffner SM, D'Agostino R, Jr., Mykkanen L*, et al.* Insulin sensitivity in subjects with type 2 diabetes. Relationship to cardiovascular risk factors: the Insulin Resistance Atherosclerosis Study. *Diabetes Care* 1999; **22:** 562-8.
64 Stern MP. Do non-insulin-dependent diabetes mellitus and cardiovascular disease share common antecedents? *Ann Intern Med* 1996; **124:** 110-6.
65 Nathan DM, Buse JB, Davidson MB, Ferrannini E, Holman RR, Sherwin R, Zinman B. Medical management of hyperglycemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy: a consensus statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care* 2009; **32:** 193-203.
66 Nathan DM, Cleary PA, Backlund JY*, et al.* Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *N Engl J Med* 2005; **353:** 2643-53.

67 Taubes G. Diabetes. Paradoxical effects of tightly controlled blood sugar. *Science* 2008; **322:** 365-7.

68 Monti MC, Lonsdale JT, Montomoli C, Montross R, Schlag E, Greenberg DA. Familial risk factors for microvascular complications and differential male-female risk in a large cohort of American families with type 1 diabetes. *J Clin Endocrinol Metab* 2007; **92:** 4650-5.

69 Reaven G. The metabolic syndrome or the insulin resistance syndrome? Different names, different concepts, and different goals. *Endocrinol Metab Clin North Am* 2004; **33:** 283-303.

70 Zimmet P, Alberti KG, Shaw J. Global and societal implications of the diabetes epidemic. *Nature* 2001; **414:** 782-7.

71 Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science* 2008; **322:** 881-8.

72 Pennisi E. Breakthrough of the year. Human genetic variation. *Science* 2007; **318:** 1842-3.

73 Rahim NG, Harismendy O, Topol EJ, Frazer KA. Genetic determinants of phenotypic diversity in humans. *Genome Biol* 2008; **9:** 215.

74 Levy S, Sutton G, Ng PC*, et al.* The diploid genome sequence of an individual human. *PLoS Biol* 2007; **5:** e254.

75 Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet* 2006; **7:** 85-97.

76 Iafrate AJ, Feuk L, Rivera MN*, et al.* Detection of large-scale variation in the human genome. *Nat Genet* 2004; **36:** 949-51.

77 Estivill X, Armengol L. Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet* 2007; **3:** 1787-99.

78 Doria A, Patti ME, Kahn CR. The emerging genetic architecture of type 2 diabetes. *Cell Metab* 2008; **8:** 186-200.

79 Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C. Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* 2011.

80 Pasaniuc B, Avinery R, Gur T, Skibola CF, Bracci PM, Halperin E. A generic coalescent-based framework for the selection of a reference panel for imputation. *Genet Epidemiol* 2010; **34:** 773-82.

81 Hattersley AT, McCarthy MI. What makes a good genetic association study? *Lancet* 2005; **366:** 1315-23.

82 Hopper JL, Bishop DT, Easton DF. Population-based family studies in genetic epidemiology. *Lancet* 2005; **366:** 1397-406.

83 McCarthy MI. Growing evidence for diabetes susceptibility genes from genome scan data. *Curr Diab Rep* 2003; **3:** 159-67.

84 Vaxillaire M, Froguel P. Genetic basis of maturity-onset diabetes of the young. *Endocrinol Metab Clin North Am* 2006; **35:** 371-84, x.

85 Waterfield T, Gloyn AL. Monogenic β-cell dysfunction in children: clinical phenotypes, genetic etiology and mutational pathways. *Pediatric Health* 2008; **2:** 517-32.

86 O'Rahilly S. Human genetics illuminates the paths to metabolic disease. *Nature* 2009; **462:** 307-14.
87 Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994; **265:** 2037-48.
88 Risch NJ. Searching for genetic determinants in the new millennium. *Nature* 2000; **405:** 847-56.
89 Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003; **33:** 177-82.
90 Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet* 2001; **29:** 306-9.
91 Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005; **2:** e124.
92 Zondervan KT, Cardon LR. The complex interplay among factors that influence allelic association. *Nat Rev Genet* 2004; **5:** 89-100.
93 Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; **273:** 1516-7.
94 Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet* 2001; **17:** 502-10.
95 Petretto E, Liu ET, Aitman TJ. A gene harvest revealing the archeology and complexity of human disease. *Nat Genet* 2007; **39:** 1299-301.
96 Lander ES, Linton LM, Birren B*, et al.* Initial sequencing and analysis of the human genome. *Nature* 2001; **409:** 860-921.
97 Venter JC, Adams MD, Myers EW*, et al.* The sequence of the human genome. *Science* 2001; **291:** 1304-51.
98 Frazer KA, Ballinger DG, Cox DR*, et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; **449:** 851-61.
99 Hindorff LA, Junkins HA, Hall PN, Mehta JP, TA M. A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies. Accessed 09.25.2011.
100 Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 2009; **106:** 9362-7.
101 Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science* 1997; **278:** 1580-1.
102 Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001; **69:** 124-37.
103 Meigs JB, Cupples LA, Wilson PW. Parental transmission of type 2 diabetes: the Framingham Offspring Study. *Diabetes* 2000; **49:** 2201-7.
104 Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. *Diabetologia* 1999; **42:** 139-45.
105 Almgren P, Lehtovirta M, Isomaa B*, et al.* Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia* 2011.

106 Lillioja S, Wilton A. Agreement among type 2 diabetes linkage studies but a poor correlation with results from genome-wide association studies. *Diabetologia* 2009; **52:** 1061-74.

107 Grant SF, Thorleifsson G, Reynisdottir I*, et al.* Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 2006; **38:** 320-3.

108 Frayling TM. Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat Rev Genet* 2007; **8:** 657-62.

109 Florez JC, Burtt N, de Bakker PI*, et al.* Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region. *Diabetes* 2004; **53:** 1360-8.

110 Gloyn AL, Weedon MN, Owen KR*, et al.* Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* 2003; **52:** 568-72.

111 Nielsen EM, Hansen L, Carstensen B*, et al.* The E23K variant of Kir6.2 associates with impaired post-OGTT serum insulin response and increased risk of type 2 diabetes. *Diabetes* 2003; **52:** 573-7.

112 Altshuler D, Hirschhorn JN, Klannemark M*, et al.* The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 2000; **26:** 76-80.

113 Sandhu MS, Weedon MN, Fawcett KA*, et al.* Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet* 2007.

114 Gudmundsson J, Sulem P, Steinthorsdottir V*, et al.* Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 2007; **39:** 977-83.

115 Winckler W, Weedon MN, Graham RR*, et al.* Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes* 2007; **56:** 685-93.

116 Barroso I, Gurnell M, Crowley VE*, et al.* Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature* 1999; **402:** 880-3.

117 Gloyn AL, Pearson ER, Antcliff JF*, et al.* Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. *N Engl J Med* 2004; **350:** 1838-49.

118 Inoue H, Tanizawa Y, Wasson J*, et al.* A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). *Nat Genet* 1998; **20:** 143-8.

119 Nishigori H, Yamada S, Kohama T*, et al.* Frameshift mutation, A263fsinsGG, in the hepatocyte nuclear factor-1beta gene associated with diabetes and renal dysfunction. *Diabetes* 1998; **47:** 1354-5.

120 Saxena R, Voight BF, Lyssenko V*, et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007; **316:** 1331-6.

121 Scott LJ, Mohlke KL, Bonnycastle LL*, et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007; **316:** 1341-5.

122 Sladek R, Rocheleau G, Rung J*, et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007; **445:** 881-5.

123 Steinthorsdottir V, Thorleifsson G, Reynisdottir I*, et al.* A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 2007; **39:** 770-5.

124 Zeggini E, Weedon MN, Lindgren CM*, et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007; **316:** 1336-41.

125 Florez JC, Manning AK, Dupuis J*, et al.* A 100K genome-wide association scan for diabetes and related traits in the Framingham Heart Study: replication and integration with other genome-wide datasets. *Diabetes* 2007; **56:** 3063-74.

126 Hanson RL, Bogardus C, Duggan D*, et al.* A search for variants associated with young-onset type 2 diabetes in American Indians in a 100K genotyping array. *Diabetes* 2007; **56:** 3045-52.

127 Hayes MG, Pluzhnikov A, Miyake K*, et al.* Identification of type 2 diabetes genes in Mexican Americans through genome-wide association studies. *Diabetes* 2007; **56:** 3033-44.

128 Rampersaud E, Damcott CM, Fu M*, et al.* Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: evidence for replication from diabetes-related quantitative traits and from independent populations. *Diabetes* 2007; **56:** 3053-62.

129 Salonen JT, Uimari P, Aalto JM*, et al.* Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. *Am J Hum Genet* 2007; **81:** 338-45.

130 Cauchi S, Meyre D, Dina C*, et al.* Transcription factor TCF7L2 genetic study in the French population: expression in human beta-cells and adipose tissue and strong association with type 2 diabetes. *Diabetes* 2006; **55:** 2903-8.

131 Chandak GR, Janipalli CS, Bhaskar S*, et al.* Common variants in the TCF7L2 gene are strongly associated with type 2 diabetes mellitus in the Indian population. *Diabetologia* 2007; **50:** 63-7.

132 Damcott CM, Pollin TI, Reinhart LJ*, et al.* Polymorphisms in the transcription factor 7-like 2 (TCF7L2) gene are associated with type 2 diabetes in the Amish: replication and evidence for a role in both insulin secretion and insulin resistance. *Diabetes* 2006; **55:** 2654-9.

133 Florez JC, Jablonski KA, Bayley N*, et al.* TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N Engl J Med* 2006; **355:** 241-50.

134 Groves CJ, Zeggini E, Minton J*, et al.* Association analysis of 6,736 U.K. subjects provides replication and confirms TCF7L2 as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. *Diabetes* 2006; **55:** 2640-4.

135 Hayashi T, Iwamoto Y, Kaku K, Hirose H, Maeda S. Replication study for the association of TCF7L2 with susceptibility to type 2 diabetes in a Japanese population. *Diabetologia* 2007; **50:** 980-4.

136 Humphries SE, Gable D, Cooper JA*, et al.* Common variants in the TCF7L2 gene and predisposition to type 2 diabetes in UK European Whites, Indian Asians and Afro-Caribbean men and women. *J Mol Med* 2006; **84:** 1-10.

137 Lehman DM, Hunt KJ, Leach RJ*, et al.* Haplotypes of transcription factor 7-like 2 (TCF7L2) gene and its upstream region are associated with type 2 diabetes and age of onset in Mexican Americans. *Diabetes* 2007; **56:** 389-93.

138 Saxena R, Gianniny L, Burtt NP*, et al.* Common single nucleotide polymorphisms in TCF7L2 are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals. *Diabetes* 2006; **55:** 2890-5.

139 Scott LJ, Bonnycastle LL, Willer CJ*, et al.* Association of transcription factor 7-like 2 (TCF7L2) variants with type 2 diabetes in a Finnish sample. *Diabetes* 2006; **55:** 2649-53.

140 Zeggini E, McCarthy MI. TCF7L2: the biggest story in diabetes genetics since HLA? *Diabetologia* 2007; **50:** 1-4.

141 Zhang C, Qi L, Hunter DJ, Meigs JB, Manson JE, van Dam RM, Hu FB. Variant of transcription factor 7-like 2 (TCF7L2) gene and the risk of type 2 diabetes in large cohorts of U.S. women and men. *Diabetes* 2006; **55:** 2645-8.

142 Lyssenko V, Lupi R, Marchetti P*, et al.* Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. *J Clin Invest* 2007; **117:** 2155-63.

143 The Wellcome Trust Case Control ConsortiumGenome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447:** 661-78.

144 Helgadottir A, Thorleifsson G, Manolescu A*, et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 2007; **316:** 1491-3.

145 McPherson R, Pertsemlidis A, Kavaslar N*, et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* 2007; **316:** 1488-91.

146 Samani NJ, Erdmann J, Hall AS*, et al.* Genomewide association analysis of coronary artery disease. *N Engl J Med* 2007; **357:** 443-53.

147 Helgadottir A, Thorleifsson G, Magnusson KP*, et al.* The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat Genet* 2008; **40:** 217-24.

148 Matarin M, Brown WM, Singleton A, Hardy JA, Meschia JF. Whole genome analyses suggest ischemic stroke and heart disease share an association with polymorphisms on chromosome 9p21. *Stroke* 2008; **39:** 1586-9.

149 Newton-Cheh C, Cook NR, VanDenburgh M, Rimm EB, Ridker PM, Albert CM. A common variant at 9p21 is associated with sudden and arrhythmic cardiac death. *Circulation* 2009; **120:** 2062-8.

150 Duesing K, Fatemifar G, Charpentier G*, et al.* Strong association of common variants in the CDKN2A/CDKN2B region with type 2 diabetes in French Europids. *Diabetologia* 2008; **51:** 821-6.

151 Horikawa Y, Miyake K, Yasuda K*, et al.* Replication of genome-wide association studies of type 2 diabetes susceptibility in Japan. *J Clin Endocrinol Metab* 2008; **93:** 3136-41.

152 Ng MC, Park KS, Oh B*, et al.* Implication of genetic variants near TCF7L2, SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGF2BP2, and FTO in type 2 diabetes and obesity in 6,719 Asians. *Diabetes* 2008; **57:** 2226-33.

153 Wu Y, Li H, Loos RJ*, et al.* Common variants in CDKAL1, CDKN2A/B, IGF2BP2, SLC30A8, and HHEX/IDE genes are associated with type 2 diabetes and impaired fasting glucose in a Chinese Han population. *Diabetes* 2008; **57:** 2834-42.

154 Assimes TL, Knowles JW, Basu A*, et al.* Susceptibility locus for clinical and subclinical coronary artery disease at chromosome 9p21 in the multi-ethnic ADVANCE study. *Hum Mol Genet* 2008; **17:** 2320-8.

155 Hiura Y, Fukushima Y, Yuno M*, et al.* Validation of the association of genetic variants on chromosome 9p21 and 1q41 with myocardial infarction in a Japanese population. *Circ J* 2008; **72:** 1213-7.

156 Larson MG, Atwood LD, Benjamin EJ*, et al.* Framingham Heart Study 100K project: genome-wide associations for cardiovascular disease outcomes. *BMC Med Genet* 2007; **8 Suppl 1:** S5.

157 Schunkert H, Gotz A, Braund P*, et al.* Repeated replication and a prospective meta-analysis of the association between chromosome 9p21.3 and coronary artery disease. *Circulation* 2008; **117:** 1675-84.

158 Shen GQ, Li L, Rao S*, et al.* Four SNPs on chromosome 9p21 in a South Korean population implicate a genetic locus that confers high cross-race risk for development of coronary artery disease. *Arterioscler Thromb Vasc Biol* 2008; **28:** 360-5.

159 Shen GQ, Rao S, Martinelli N*, et al.* Association between four SNPs on chromosome 9p21 and myocardial infarction is replicated in an Italian population. *J Hum Genet* 2008; **53:** 144-50.

160 Zhou L, Zhang X, He M, Cheng L, Chen Y, Hu FB, Wu T. Associations between single nucleotide polymorphisms on chromosome 9p21 and risk of coronary heart disease in Chinese Han population. *Arterioscler Thromb Vasc Biol* 2008; **28:** 2085-9.

161 Abdullah KG, Li L, Shen GQ*, et al.* Four SNPS on chromosome 9p21 confer risk to premature, familial CAD and MI in an American Caucasian population (GeneQuest). *Ann Hum Genet* 2008; **72:** 654-7.

162 Chen Z, Qian Q, Ma G*, et al.* A common variant on chromosome 9p21 affects the risk of early-onset coronary artery disease. *Mol Biol Rep* 2009; **36:** 889-93.

163 Wahlstrand B, Orho-Melander M, Delling L*, et al.* The myocardial infarction associated CDKN2A/CDKN2B locus on chromosome 9p21 is associated with stroke independently of coronary events in patients with hypertension. *J Hypertens* 2009; **27:** 769-73.

164 Ye S, Willeit J, Kronenberg F, Xu Q, Kiechl S. Association of genetic variation on chromosome 9p21 with susceptibility and progression of atherosclerosis: a population-based, prospective study. *J Am Coll Cardiol* 2008; **52:** 378-84.

165 Zee RY, Ridker PM. Two common gene variants on chromosome 9 and risk of atherothrombosis. *Stroke* 2007; **38:** e111.

166 Doria A, Wojcik J, Xu R*, et al.* Interaction between poor glycemic control and 9p21 locus on risk of coronary artery disease in type 2 diabetes. *JAMA* 2008; **300:** 2389-97.

167 Broadbent HM, Peden JF, Lorkowski S*, et al.* Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. *Hum Mol Genet* 2008; **17:** 806-14.

168 Cunnington MS, Santibanez Koref M, Mayosi BM, Burn J, Keavney B. Chromosome 9p21 SNPs Associated with Multiple Disease Phenotypes Correlate with ANRIL Expression. *PLoS Genet* 2010; **6:** e1000899.

169 Liu Y, Sanoff HK, Cho H*, et al.* INK4/ARF transcript expression is associated with chromosome 9p21 variants linked to atherosclerosis. *PLoS One* 2009; **4:** e5027.

170 Harismendy O, Notani D, Song X*, et al.* 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature* 2011; **470:** 264-8.

171 Wenzlau JM, Juhl K, Yu L*, et al.* The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *Proc Natl Acad Sci U S A* 2007; **104:** 17040-5.

172 Zeggini E, Scott LJ, Saxena R*, et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008; **40:** 638-45.

173 Unoki H, Takahashi A, Kawaguchi T*, et al.* SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet* 2008; **40:** 1098-102.

174 Yasuda K, Miyake K, Horikawa Y*, et al.* Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* 2008; **40:** 1092-7.

175 Tsai FJ, Yang CF, Chen CC*, et al.* A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. *PLoS Genet* 2010; **6:** e1000847.

176 Florez JC. Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes? *Diabetologia* 2008; **51:** 1100-10.

177 Rung J, Cauchi S, Albrechtsen A*, et al.* Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet* 2009; **41:** 1110-5.

178 Mulder H, Nagorny CL, Lyssenko V, Groop L. Melatonin receptors in pancreatic islets: good morning to a novel type 2 diabetes gene. *Diabetologia* 2009; **52:** 1240-9.

179 Bouatia-Naji N, Bonnefond A, Cavalcanti-Proenca C*, et al.* A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. *Nat Genet* 2009; **41:** 89-94.

180 Lyssenko V, Nagorny CL, Erdos MR*, et al.* Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat Genet* 2009; **41:** 82-8.

181 Prokopenko I, Langenberg C, Florez JC*, et al.* Variants in MTNR1B influence fasting glucose levels. *Nat Genet* 2009; **41:** 77-81.

182 Dupuis J, Langenberg C, Prokopenko I*, et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 2010; **42:** 105-16.

183 Kooner JS, Saleheen D, Sim X*, et al.* Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet* 2011; **43:** 984-9.

184 Qi L, Cornelis MC, Kraft P*, et al.* Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum Mol Genet* 2010; **19:** 2706-15.

185 Voight BF, Scott LJ, Steinthorsdottir V*, et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 2010; **42:** 579-89.

186 Civelek M, Lusis AJ. Conducting the metabolic syndrome orchestra. *Nat Genet* 2011; **43:** 506-8.

187 Billings LK, Florez JC. The genetics of type 2 diabetes: what have we learned from GWAS? *Ann N Y Acad Sci* 2010; **1212:** 59-77.

188 Lyssenko V, Jonsson A, Almgren P*, et al.* Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 2008; **359:** 2220-32.

189 Hardy J, Singleton A. Genomewide association studies and human disease. *N Engl J Med* 2009; **360:** 1759-68.

190 Hirschhorn JN. Genomewide association studies--illuminating biologic pathways. *N Engl J Med* 2009; **360:** 1699-701.

191 Kraft P, Hunter DJ. Genetic risk prediction--are we there yet? *N Engl J Med* 2009; **360:** 1701-3.

192 Simonis-Bik AM, Nijpels G, van Haeften TW*, et al.* Gene variants in the novel type 2 diabetes loci CDC123/CAMK1D, THADA, ADAMTS9, BCL11A, and MTNR1B affect different aspects of pancreatic beta-cell function. *Diabetes* 2009; **59:** 293-301.

193 Kilpelainen TO, Zillikens MC, Stancakova A*, et al.* Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat Genet* 2011; **43:** 753-60.

194 Boesgaard TW, Gjesing AP, Grarup N*, et al.* Variant near ADAMTS9 known to associate with type 2 diabetes is related to insulin resistance in offspring of type 2 diabetes patients--EUGENE2 study. *PLoS One* 2009; **4:** e7236.

195 Grarup N, Rose CS, Andersson EA*, et al.* Studies of association of variants near the HHEX, CDKN2A/B, and IGF2BP2 genes with type 2 diabetes and impaired insulin release in 10,705 Danish subjects: validation and extension of genome-wide association studies. *Diabetes* 2007; **56:** 3105-11.

196 Deeb SS, Fajas L, Nemoto M*, et al.* A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nat Genet* 1998; **20:** 284-7.

197 Florez JC, Jablonski KA, McAteer J*, et al.* Testing of diabetes-associated WFS1 polymorphisms in the Diabetes Prevention Program. *Diabetologia* 2008; **51:** 451-7.

198 Franks PW, Rolandsson O, Debenham SL*, et al.* Replication of the association between variants in WFS1 and risk of type 2 diabetes in European populations. *Diabetologia* 2008; **51:** 458-63.

199 Takeuchi F, Serizawa M, Yamamoto K*, et al.* Confirmation of multiple risk Loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. *Diabetes* 2009; **58:** 1690-9.

200 Pascoe L, Tura A, Patel SK*, et al.* Common variants of the novel type 2 diabetes genes CDKAL1 and HHEX/IDE are associated with decreased pancreatic beta-cell function. *Diabetes* 2007; **56:** 3101-4.

201 Grarup N, Andersen G, Krarup NT*, et al.* Association testing of novel type 2 diabetes risk alleles in the JAZF1, CDC123/CAMK1D, TSPAN8, THADA, ADAMTS9, and NOTCH2 loci with insulin release, insulin sensitivity, and obesity in a population-based sample of 4,516 glucose-tolerant middle-aged Danes. *Diabetes* 2008; **57:** 2534-40.

202 Staiger H, Machicao F, Stefan N*, et al.* Polymorphisms within novel risk loci for type 2 diabetes determine beta-cell function. *PLoS One* 2007; **2:** e832.

203 Below JE, Gamazon ER, Morrison JV*, et al.* Genome-wide association and meta-analysis in populations from Starr County, Texas, and Mexico City identify type 2 diabetes susceptibility loci and enrichment for expression quantitative trait loci in top signals. *Diabetologia* 2011; **54:** 2047-55.

204 Chang YC, Chiu YF, Liu PH*, et al.* Replication of Genome-Wide Association Signals of Type 2 Diabetes in Han Chinese in a Prospective Cohort. *Clin Endocrinol (Oxf)* 2011.

205 Parra EJ, Below JE, Krithika S*, et al.* Genome-wide association study of type 2 diabetes in a sample from Mexico City and a meta-analysis of a Mexican-American sample from Starr County, Texas. *Diabetologia* 2011; **54:** 2038-46.

206 Shu XO, Long J, Cai Q*, et al.* Identification of new genetic risk variants for type 2 diabetes. *PLoS Genet* 2010; **6**.

207 Stancakova A, Kuulasmaa T, Paananen J*, et al.* Association of 18 confirmed susceptibility loci for type 2 diabetes with indices of insulin release, proinsulin conversion, and insulin sensitivity in 5,327 nondiabetic Finnish men. *Diabetes* 2009; **58:** 2129-36.

208 Staiger H, Stancakova A, Zilinskaite J*, et al.* A candidate type 2 diabetes polymorphism near the HHEX locus affects acute glucose-stimulated insulin release in European populations: results from the EUGENE2 study. *Diabetes* 2008; **57:** 514-7.

209 De Silva NM, Steele A, Shields B*, et al.* The transcription factor 7-like 2 (TCF7L2) gene is associated with Type 2 diabetes in UK community-based

cases, but the risk allele frequency is reduced compared with UK cases selected for genetic studies. *Diabet Med* 2007; **24:** 1067-72.

210 Freathy RM, Weedon MN, Bennett A*, et al.* Type 2 diabetes TCF7L2 risk genotypes alter birth weight: a study of 24,053 individuals. *Am J Hum Genet* 2007; **80:** 1150-61.

211 Kimber CH, Doney AS, Pearson ER*, et al.* TCF7L2 in the Go-DARTS study: evidence for a gene dose effect on both diabetes susceptibility and control of glucose levels. *Diabetologia* 2007; **50:** 1186-91.

212 Melzer D, Murray A, Hurst AJ*, et al.* Effects of the diabetes linked TCF7L2 polymorphism in a representative older population. *BMC Med* 2006; **4:** 34.

213 Strawbridge RJ, Dupuis J, Prokopenko I*, et al.* Genome-Wide Association Identifies Nine Common Variants Associated With Fasting Proinsulin Levels and Provides New Insights Into the Pathophysiology of Type 2 Diabetes. *Diabetes* 2011; **60:** 2624-34.

214 Thorsby PM, Midthjell K, Gjerlaugsen N, Holmen J, Hanssen KF, Birkeland KI, Berg JP. Comparison of genetic risk in three candidate genes (TCF7L2, PPARG, KCNJ11) with traditional risk factors for type 2 diabetes in a population-based study--the HUNT study. *Scand J Clin Lab Invest* 2009; **69:** 282-7.

215 Timpson NJ, Lindgren CM, Weedon MN*, et al.* Adiposity-related heterogeneity in patterns of type 2 diabetes susceptibility observed in genome-wide association data. *Diabetes* 2009; **58:** 505-10.

216 Anand A, Chada K. In vivo modulation of Hmgic reduces obesity. *Nat Genet* 2000; **24:** 377-80.

217 Weedon MN, Lettre G, Freathy RM*, et al.* A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nat Genet* 2007; **39:** 1245-50.

218 Cui B, Zhu X, Xu M*, et al.* A genome-wide association study confirms previously reported loci for type 2 diabetes in Han Chinese. *PLoS One* 2011; **6:** e22353.

219 Holmkvist J, Almgren P, Lyssenko V*, et al.* Common variants in maturity-onset diabetes of the young genes and future risk of type 2 diabetes. *Diabetes* 2008; **57:** 1738-44.

220 Winckler W, Burtt NP, Holmkvist J*, et al.* Association of common variation in the HNF1alpha gene region with risk of type 2 diabetes. *Diabetes* 2005; **54:** 2336-42.

221 Weedon MN, Owen KR, Shields B*, et al.* A large-scale association analysis of common variation of the HNF1alpha gene with type 2 diabetes in the U.K. Caucasian population. *Diabetes* 2005; **54:** 2487-91.

222 Bonnycastle LL, Willer CJ, Conneely KN*, et al.* Common variants in maturity-onset diabetes of the young genes contribute to risk of type 2 diabetes in Finns. *Diabetes* 2006; **55:** 2534-40.

223 Grarup N, Overvad M, Sparso T*, et al.* The diabetogenic VPS13C/C2CD4A/C2CD4B rs7172432 variant impairs glucose-stimulated

insulin response in 5,722 non-diabetic Danish individuals. *Diabetologia* 2011; **54:** 789-94.
224 Yamauchi T, Hara K, Maeda S*, et al.* A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nat Genet* 2010; **42:** 864-8.
225 Dina C, Meyre D, Gallina S*, et al.* Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet* 2007; **39:** 724-6.
226 Frayling TM, Timpson NJ, Weedon MN*, 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:** 889-94.
227 Freathy RM, Timpson NJ, Lawlor DA*, 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:** 1419-26.
228 Scuteri A, Sanna S, Chen WM*, et al.* Genome-Wide Association Scan Shows Genetic Variants in the FTO Gene Are Associated with Obesity-Related Traits. *PLoS Genet* 2007; **3:** e115.
229 Frayling TM, Colhoun H, Florez JC. A genetic link between type 2 diabetes and prostate cancer. *Diabetologia* 2008; **51:** 1757-60.
230 Weedon MN, Owen KR, Shields B*, et al.* Common variants of the hepatocyte nuclear factor-4alpha P2 promoter are associated with type 2 diabetes in the U.K. population. *Diabetes* 2004; **53:** 3002-6.
231 Johansson S, Raeder H, Eide SA*, et al.* Studies in 3,523 Norwegians and meta-analysis in 11,571 subjects indicate that variants in the hepatocyte nuclear factor 4 alpha (HNF4A) P2 region are associated with type 2 diabetes in Scandinavians. *Diabetes* 2007; **56:** 3112-7.
232 Yamagata K, Furuta H, Oda N*, et al.* Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). *Nature* 1996; **384:** 458-60.
233 Barroso I, Luan J, Wheeler E*, et al.* Population-specific risk of type 2 diabetes conferred by HNF4A P2 promoter variants: a lesson for replication studies. *Diabetes* 2008; **57:** 3161-5.
234 Emanuelli B, Eberle D, Suzuki R, Kahn CR. Overexpression of the dual-specificity phosphatase MKP-4/DUSP-9 protects against stress-induced insulin resistance. *Proc Natl Acad Sci U S A* 2008; **105:** 3545-50.
235 Nielsen T, Sparso T, Grarup N*, et al.* Type 2 diabetes risk allele near CENTD2 is associated with decreased glucose-stimulated insulin release. *Diabetologia* 2011; **54:** 1052-6.
236 Meigs JB, Panhuysen CI, Myers RH, Wilson PW, Cupples LA. A genome-wide scan for loci linked to plasma levels of glucose and HbA(1c) in a community-based sample of Caucasian pedigrees: The Framingham Offspring Study. *Diabetes* 2002; **51:** 833-40.
237 Pilia G, Chen WM, Scuteri A*, et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet* 2006; **2:** e132.

238 Bouatia-Naji N, Rocheleau G, Van Lommel L*, et al.* A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels. *Science* 2008; **320:** 1085-8.

239 Chen WM, Erdos MR, Jackson AU*, et al.* Variations in the G6PC2/ABCB11 genomic region are associated with fasting glucose levels. *J Clin Invest* 2008; **118:** 2620-8.

240 Meigs JB, Manning AK, Fox CS, Florez JC, Liu C, Cupples LA, Dupuis J. Genome-wide association with diabetes-related traits in the Framingham Heart Study. *BMC Med Genet* 2007; **8 Suppl 1:** S16.

241 Paterson AD, Waggott D, Boright AP*, et al.* A genome-wide association study identifies a novel major locus for glycemic control in type 1 diabetes, as measured by both A1C and glucose. *Diabetes* 2010; **59:** 539-49.

242 Soranzo N. Genetic Determinants of Variability in Glycated Hemoglobin (HbA(1c)) in Humans: Review of Recent Progress and Prospects for Use in Diabetes Care. *Curr Diab Rep* 2011.

243 Soranzo N, Sanna S, Wheeler E*, et al.* Common variants at ten genomic loci influence hemoglobin A1C levels via glycemic and non-glycemic pathways. *Diabetes* 2010; **59:** 3229-39.

244 Franklin CS, Aulchenko YS, Huffman JE*, et al.* The TCF7L2 diabetes risk variant is associated with HbA(C) levels: a genome-wide association meta-analysis. *Ann Hum Genet* 2010; **74:** 471-8.

245 Pare G, Chasman DI, Parker AN, Nathan DM, Miletich JP, Zee RY, Ridker PM. Novel association of HK1 with glycated hemoglobin in a non-diabetic population: a genome-wide evaluation of 14,618 participants in the Women's Genome Health Study. *PLoS Genet* 2008; **4:** e1000312.

246 Cauchi S, Meyre D, Choquet H*, et al.* TCF7L2 variation predicts hyperglycemia incidence in a French general population: the data from an epidemiological study on the Insulin Resistance Syndrome (DESIR) study. *Diabetes* 2006; **55:** 3189-92.

247 Weedon MN, Clark VJ, Qian Y*, et al.* A common haplotype of the glucokinase gene alters fasting glucose and birth weight: association in six studies and population-genetics analyses. *Am J Hum Genet* 2006; **79:** 991-1001.

248 Swinburn BA, Sacks G, Hall KD, McPherson K, Finegood DT, Moodie ML, Gortmaker SL. The global obesity pandemic: shaped by global drivers and local environments. *Lancet* 2011; **378:** 804-14.

249 Hjelmborg JB, Fagnani C, Silventoinen K*, et al.* Genetic influences on growth traits of BMI: a longitudinal study of adult twins. *Obesity (Silver Spring)* 2008; **16:** 847-52.

250 Wardle J, Carnell S, Haworth CM, Plomin R. Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment. *Am J Clin Nutr* 2008; **87:** 398-404.

251 Saunders CL, Chiodini BD, Sham P*, et al.* Meta-analysis of genome-wide linkage studies in BMI and obesity. *Obesity (Silver Spring)* 2007; **15:** 2263-75.

252 Rankinen T, Zuberi A, Chagnon YC*, et al.* The human obesity gene map: the 2005 update. *Obesity (Silver Spring)* 2006; **14:** 529-644.

253 Larsen LH, Echwald SM, Sorensen TI, Andersen T, Wulff BS, Pedersen O. Prevalence of mutations and functional analyses of melanocortin 4 receptor variants identified among 750 men with juvenile-onset obesity. *J Clin Endocrinol Metab* 2005; **90:** 219-24.

254 Fawcett KA, Barroso I. The genetics of obesity: FTO leads the way. *Trends Genet* 2010; **26:** 266-74.

255 Wing MR, Ziegler JM, Langefeld CD*, et al.* Analysis of FTO gene variants with obesity and glucose homeostasis measures in the multiethnic Insulin Resistance Atherosclerosis Study cohort. *Int J Obes (Lond)* 2011; **35:** 1173-82.

256 Gerken T, Girard CA, Tung YC*, et al.* The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* 2007; **318:** 1469-72.

257 Sanchez-Pulido L, Andrade-Navarro MA. The FTO (fat mass and obesity associated) gene codes for a novel member of the non-heme dioxygenase superfamily. *BMC Biochem* 2007; **8:** 23.

258 Rose NR, McDonough MA, King ON, Kawamura A, Schofield CJ. Inhibition of 2-oxoglutarate dependent oxygenases. *Chem Soc Rev* 2011; **40:** 4364-97.

259 Jia G, Fu Y, Zhao X*, et al.* N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* 2011.

260 Fredriksson R, Hagglund M, Olszewski PK*, 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:** 2062-71.

261 Stratigopoulos G, Padilla SL, LeDuc CA*, et al.* Regulation of Fto/Ftm gene expression in mice and humans. *Am J Physiol Regul Integr Comp Physiol* 2008; **294:** R1185-96.

262 Cecil JE, Tavendale R, Watt P, Hetherington MM, Palmer CN. An obesity-associated FTO gene variant and increased energy intake in children. *N Engl J Med* 2008; **359:** 2558-66.

263 Andreasen CH, Stender-Petersen KL, Mogensen MS*, et al.* Low physical activity accentuates the effect of the FTO rs9939609 polymorphism on body fat accumulation. *Diabetes* 2008; **57:** 95-101.

264 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:** 1418-25.

265 Kilpeläinen TO, Qi L, Brage S*, 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:** e1001116.

266 Kloting N, Schleinitz D, Ruschke K*, et al.* Inverse relationship between obesity and FTO gene expression in visceral adipose tissue in humans. *Diabetologia* 2008; **51:** 641-7.

267 Wahlen K, Sjolin E, Hoffstedt J. The common rs9939609 gene variant of the fat mass- and obesity-associated gene FTO is related to fat cell lipolysis. *J Lipid Res* 2008; **49:** 607-11.

268 Al-Attar SA, Pollex RL, Ban MR*, et al.* Association between the FTO rs9939609 polymorphism and the metabolic syndrome in a non-Caucasian multi-ethnic sample. *Cardiovasc Diabetol* 2008; **7:** 5.

269 Barber TM, Bennett AJ, Groves CJ*, et al.* Association of variants in the fat mass and obesity associated (FTO) gene with polycystic ovary syndrome. *Diabetologia* 2008; **51:** 1153-8.

270 Loos RJ, Lindgren CM, Li S*, et al.* Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* 2008; **40:** 768-75.

271 Speliotes EK, Willer CJ, Berndt SI*, et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42:** 937-48.

272 Thorleifsson G, Walters GB, Gudbjartsson DF*, et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 2009; **41:** 18-24.

273 Willer CJ, Speliotes EK, Loos RJ*, et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 2009; **41:** 25-34.

274 Benzinou M, Creemers JW, Choquet H*, et al.* Common nonsynonymous variants in PCSK1 confer risk of obesity. *Nat Genet* 2008; **40:** 943-5.

275 Meyre D, Delplanque J, Chevre JC*, et al.* Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat Genet* 2009; **41:** 157-9.

276 Scherag A, Dina C, Hinney A*, et al.* Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and german study groups. *PLoS Genet* 2010; **6:** e1000916.

277 Walters RG, Jacquemont S, Valsesia A*, et al.* A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* 2010; **463:** 671-5.

278 Chambers JC, Elliott P, Zabaneh D*, et al.* Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat Genet* 2008; **40:** 716-8.

279 Heard-Costa NL, Zillikens MC, Monda KL*, et al.* NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. *PLoS Genet* 2009; **5:** e1000539.

280 Heid IM, Jackson AU, Randall JC*, et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat Genet* 2010; **42:** 949-60.

281 Lindgren CM, Heid IM, Randall JC*, et al.* Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet* 2009; **5:** e1000508.

282 Holmen J, Midthjell K, Krüger Ø*, et al.* The Nord-Trøndelag Health Study 1995-97 (HUNT2): Objectives, contents, methods and participation. *Norw J Epidemiol* 2003; **13:** 19-32.

283 Berglund G, Elmstahl S, Janzon L, Larsson SA. The Malmo Diet and Cancer Study. Design and feasibility. *J Intern Med* 1993; **233:** 45-51.

284 Berglund G, Nilsson P, Eriksson KF, Nilsson JA, Hedblad B, Kristenson H, Lindgarde F. Long-term outcome of the Malmo preventive project: mortality and cardiovascular morbidity. *J Intern Med* 2000; **247:** 19-29.

285 Nilsson P, Berglund G. Prevention of cardiovascular disease and diabetes: lessons from the Malmo Preventive Project. *J Intern Med* 2000; **248:** 455-62.

286 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21:** 263-5.

287 Storm N, Darnhofer-Patel B, van den Boom D, Rodi CP. MALDI-TOF mass spectrometry-based SNP genotyping. *Methods Mol Biol* 2003; **212:** 241-62.

288 Yang W, Kelly T, He J. Genetic epidemiology of obesity. *Epidemiol Rev* 2007; **29:** 49-61.

289 Konig IR. Validation in genetic association studies. *Brief Bioinform* 2011; **12:** 253-8.

290 Ioannidis JP, Boffetta P, Little J*, et al.* Assessment of cumulative evidence on genetic associations: interim guidelines. *Int J Epidemiol* 2008; **37:** 120-32.

291 Cordell HJ, Clayton DG. Genetic association studies. *Lancet* 2005; **366:** 1121-31.

292 Hill AB. The Environment and Disease: Association or Causation? *Proc R Soc Med* 1965; **58:** 295-300.

293 Chanock SJ, Manolio T, Boehnke M*, et al.* Replicating genotype-phenotype associations. *Nature* 2007; **447:** 655-60.

294 Ioannidis JP. Non-replication and inconsistency in the genome-wide association setting. *Hum Hered* 2007; **64:** 203-13.

295 Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004; **96:** 434-42.

296 Kraft P. Curses--winner's and otherwise--in genetic epidemiology. *Epidemiology* 2008; **19:** 649-51; discussion 57-8.

297 Bland JM, Altman DG. Regression towards the mean. *BMJ* 1994; **308:** 1499.

298 Freedman ML, Reich D, Penney KL*, et al.* Assessing the impact of population stratification on genetic association studies. *Nat Genet* 2004; **36:** 388-93.

299 Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. *Nat Genet* 2004; **36:** 512-7.

300 Li S, Zhao JH, Luan J*, et al.* Genetic predisposition to obesity leads to increased risk of type 2 diabetes. *Diabetologia* 2011; **54:** 776-82.

301 Craig DW, Goor RM, Wang Z*, et al.* Assessing and managing risk when sharing aggregate genetic variant data. *Nat Rev Genet* 2011; **12:** 730-6.

302 Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. *J Clin Epidemiol* 2000; **53:** 207-16.

303 Ioannidis JP, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG. Genetic associations in large versus small studies: an empirical assessment. *Lancet* 2003; **361:** 567-71.

304 Kavvoura FK, Ioannidis JP. Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. *Hum Genet* 2008; **123:** 1-14.

305 Sanghera DK, Ortega L, Han S*, et al.* Impact of nine common type 2 diabetes risk polymorphisms in Asian Indian Sikhs: PPARG2 (Pro12Ala), IGF2BP2, TCF7L2 and FTO variants confer a significant risk. *BMC Med Genet* 2008; **9:** 59.

306 Yajnik CS, Janipalli CS, Bhaskar S*, et al.* FTO gene variants are strongly associated with type 2 diabetes in South Asian Indians. *Diabetologia* 2009; **52:** 247-52.

307 Rees SD, Islam M, Hydrie MZ*, et al.* An FTO variant is associated with Type 2 diabetes in South Asian populations after accounting for body mass index and waist circumference. *Diabet Med* 2011; **28:** 673-80.

308 Li H, Kilpelainen TO, Liu C*, et al.* Association of genetic variation in FTO with risk of obesity and type 2 diabetes with data from 96,551 East and South Asians. *Diabetologia* 2011.

309 Als TD, Jorgensen TH, Borglum AD, Petersen PA, Mors O, Wang AG. Highly discrepant proportions of female and male Scandinavian and British Isles ancestry within the isolated population of the Faroe Islands. *Eur J Hum Genet* 2006; **14:** 497-504.

310 Bressler J, Kao WH, Pankow JS, Boerwinkle E. Risk of type 2 diabetes and obesity is differentially associated with variation in FTO in whites and African-Americans in the ARIC study. *PLoS One* 2010; **5:** e10521.

311 Legry V, Cottel D, Ferrieres J*, et al.* Effect of an FTO polymorphism on fat mass, obesity, and type 2 diabetes mellitus in the French MONICA Study. *Metabolism* 2009; **58:** 971-5.

312 Pettersen E, Skorpen F, Kvaloy K, Midthjell K, Grill V. Genetic heterogeneity in latent autoimmune diabetes is linked to various degrees of autoimmune activity: results from the Nord-Trondelag Health Study. *Diabetes* 2010; **59:** 302-10.

313 Midthjell K, Holmen J, Bjorndal A, Lund-Larsen G. Is questionnaire information valid in the study of a chronic disease such as diabetes? The Nord-Trondelag diabetes study. *J Epidemiol Community Health* 1992; **46:** 537-42.

314 Do R, Bailey SD, Desbiens K*, et al.* Genetic variants of FTO influence adiposity, insulin sensitivity, leptin levels, and resting metabolic rate in the Quebec Family Study. *Diabetes* 2008; **57:** 1147-50.

315 Samaras K, Botelho NK, Chisholm DJ, Lord RV. Subcutaneous and visceral adipose tissue FTO gene expression and adiposity, insulin action, glucose metabolism, and inflammatory adipokines in type 2 diabetes mellitus and in health. *Obes Surg* 2010; **20:** 108-13.

316 Church C, Lee S, Bagg EA*, et al.* A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. *PLoS Genet* 2009; **5:** e1000599.

317 Gimble JM, Bray MS, Young A. Circadian biology and sleep: missing links in obesity and metabolism? *Obes Rev* 2009; **10 Suppl 2:** 1-5.

318 Berulava T, Horsthemke B. The obesity-associated SNPs in intron 1 of the FTO gene affect primary transcript levels. *Eur J Hum Genet* 2010; **18:** 1054-6.
319 Zabena C, Gonzalez-Sanchez JL, Martinez-Larrad MT*, et al.* The FTO obesity gene. Genotyping and gene expression analysis in morbidly obese patients. *Obes Surg* 2009; **19:** 87-95.
320 Kirkpatrick CL, Marchetti P, Purrello F*, et al.* Type 2 diabetes susceptibility gene expression in normal or diabetic sorted human alpha and beta cells: correlations with age or BMI of islet donors. *PLoS One* 2010; **5:** e11053.
321 Poritsanos NJ, Lew PS, Mizuno TM. Relationship between blood glucose levels and hepatic Fto mRNA expression in mice. *Biochem Biophys Res Commun* 2010; **400:** 713-7.
322 Bravard A, Lefai E, Meugnier E*, et al.* FTO is increased in muscle during type 2 diabetes, and its overexpression in myotubes alters insulin signaling, enhances lipogenesis and ROS production, and induces mitochondrial dysfunction. *Diabetes* 2011; **60:** 258-68.
323 Han Z, Niu T, Chang J*, et al.* Crystal structure of the FTO protein reveals basis for its substrate specificity. *Nature* 2010; **464:** 1205-9.
324 Meyre D. Is FTO a type 2 diabetes susceptibility gene? *Diabetologia* 2012.
325 Wu Q, Saunders RA, Szkudlarek-Mikho M, Serna Ide L, Chin KV. The obesity-associated Fto gene is a transcriptional coactivator. *Biochem Biophys Res Commun* 2010; **401:** 390-5.
326 Ragvin A, Moro E, Fredman D*, et al.* Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to HHEX, SOX4, and IRX3. *Proc Natl Acad Sci U S A* 2010; **107:** 775-80.
327 de Koning L, Gerstein HC, Bosch J*, et al.* Anthropometric measures and glucose levels in a large multi-ethnic cohort of individuals at risk of developing type 2 diabetes. *Diabetologia* 2010; **53:** 1322-30.
328 Lopez-Bermejo A, Petry CJ, Diaz M, Sebastiani G, de Zegher F, Dunger DB, Ibanez L. The association between the FTO gene and fat mass in humans develops by the postnatal age of two weeks. *J Clin Endocrinol Metab* 2008; **93:** 1501-5.
329 Hakanen M, Raitakari OT, Lehtimaki T*, et al.* FTO genotype is associated with body mass index after the age of seven years but not with energy intake or leisure-time physical activity. *J Clin Endocrinol Metab* 2009; **94:** 1281-7.
330 Rzehak P, Scherag A, Grallert H*, et al.* Associations between BMI and the FTO Gene Are Age Dependent: Results from the GINI and LISA Birth Cohort Studies up to Age 6 Years. *Obes Facts* 2010; **3:** 173-80.
331 Hunt SC, Stone S, Xin Y*, et al.* Association of the FTO gene with BMI. *Obesity (Silver Spring)* 2008; **16:** 902-4.
332 Jacobsson JA, Riserus U, Axelsson T, Lannfelt L, Schioth HB, Fredriksson R. The common FTO variant rs9939609 is not associated with BMI in a longitudinal study on a cohort of Swedish men born 1920-1924. *BMC Med Genet* 2009; **10:** 131.

333 Jess T, Zimmermann E, Kring SI*, et al.* Impact on weight dynamics and general growth of the common FTO rs9939609: a longitudinal Danish cohort study. *Int J Obes (Lond)* 2008; **32:** 1388-94.

334 Luan J, Kerner B, Zhao JH, Loos RJ, Sharp SJ, Muthen BO, Wareham NJ. A multilevel linear mixed model of the association between candidate genes and weight and body mass index using the Framingham longitudinal family data. *BMC Proc* 2009; **3 Suppl 7:** S115.

335 Marvelle AF, Lange LA, Qin L, Adair LS, Mohlke KL. Association of FTO with obesity-related traits in the Cebu Longitudinal Health and Nutrition Survey (CLHNS) Cohort. *Diabetes* 2008; **57:** 1987-91.

336 Qi L, Kang K, Zhang C*, et al.* Fat mass-and obesity-associated (FTO) gene variant is associated with obesity: longitudinal analyses in two cohort studies and functional test. *Diabetes* 2008; **57:** 3145-51.

337 Tabara Y, Osawa H, Guo H*, et al.* Prognostic significance of FTO genotype in the development of obesity in Japanese: the J-SHIPP study. *Int J Obes (Lond)* 2009; **33:** 1243-8.

338 Wangensteen T, Egeland T, Akselsen H, Holmen J, Undlien D, Retterstol L. FTO Genotype and Weight Gain in Obese and Normal Weight Adults From a Norwegian Population Based Cohort (the HUNT Study). *Exp Clin Endocrinol Diabetes* 2010.

339 Kaakinen M, Laara E, Pouta A*, et al.* Life-course analysis of a fat mass and obesity-associated (FTO) gene variant and body mass index in the Northern Finland Birth Cohort 1966 using structural equation modeling. *Am J Epidemiol* 2010; **172:** 653-65.

340 Truelsen T, Nielsen N, Boysen G, Gronbaek M. Self-reported stress and risk of stroke: the Copenhagen City Heart Study. *Stroke* 2003; **34:** 856-62.

341 Wilsgaard T, Jacobsen BK, Schirmer H, Thune I, Lochen ML, Njolstad I, Arnesen E. Tracking of cardiovascular risk factors: the Tromso study, 1979-1995. *Am J Epidemiol* 2001; **154:** 418-26.

342 Holmen J, Midthjell K, Forsen L, Skjerve K, Gorseth M, Oseland A. [A health survey in Nord-Trondelag 1984-86. Participation and comparison of attendants and non-attendants]. *Tidsskr Nor Laegeforen* 1990; **110:** 1973-7.

343 Fischer J, Koch L, Emmerling C, Vierkotten J, Peters T, Bruning JC, Ruther U. Inactivation of the Fto gene protects from obesity. *Nature* 2009; **458:** 894-8.

344 Vimaleswaran KS, Li S, Zhao JH*, et al.* Physical activity attenuates the body mass index-increasing influence of genetic variation in the FTO gene. *Am J Clin Nutr* 2009; **90:** 425-8.

345 Shea J, Agarwala V, Philippakis AA*, et al.* Comparing strategies to fine-map the association of common SNPs at chromosome 9p21 with type 2 diabetes and myocardial infarction. *Nat Genet* 2011; **43:** 801-5.

346 Kathiresan S, Voight BF, Purcell S*, et al.* Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 2009; **41:** 334-41.

347 Falchi M, Bataille V, Hayward NK*, et al.* Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. *Nat Genet* 2009; **41:** 915-9.

348 Sherborne AL, Hosking FJ, Prasad RB*, et al.* Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. *Nat Genet* 2010; **42:** 492-4.

349 Shete S, Hosking FJ, Robertson LB*, et al.* Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet* 2009; **41:** 899-904.

350 Stacey SN, Sulem P, Masson G*, et al.* New common variants affecting susceptibility to basal cell carcinoma. *Nat Genet* 2009; **41:** 909-14.

351 Turnbull C, Ahmed S, Morrison J*, et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* 2010; **42:** 504-7.

352 Wrensch M, Jenkins RB, Chang JS*, et al.* Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. *Nat Genet* 2009; **41:** 905-8.

353 Cugino D, Gianfagna F, Santimone I, de Gaetano G, Donati MB, Iacoviello L, Di Castelnuovo A. Type 2 diabetes and polymorphisms on chromosome 9p21: A meta-analysis. *Nutr Metab Cardiovasc Dis* 2011.

354 Silander K, Tang H, Myles S, Jakkula E, Timpson NJ, Cavalli-Sforza L, Peltonen L. Worldwide patterns of haplotype diversity at 9p21.3, a locus associated with type 2 diabetes and coronary heart disease. *Genome Med* 2009; **1:** 51.

355 Bergmann MM, Jacobs EJ, Hoffmann K, Boeing H. Agreement of self-reported medical history: comparison of an in-person interview with a self-administered questionnaire. *Eur J Epidemiol* 2004; **19:** 411-6.

356 Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol* 2004; **57:** 1096-103.

357 Bougneres P. Genetics of obesity and type 2 diabetes: tracking pathogenic traits during the predisease period. *Diabetes* 2002; **51 Suppl 3:** S295-303.

358 Sillanpaa MJ, Corander J. Model choice in gene mapping: what and why. *Trends Genet* 2002; **18:** 301-7.

359 Dermitzakis ET. From gene expression to disease risk. *Nat Genet* 2008; **40:** 492-3.

360 Florez JC. A Genome-Wide Association Study of Treated A1C - A Genetic Needle in an Environmental Haystack? *Diabetes* 2010; **59:** 332-4.

361 Clee SM, Yandell BS, Schueler KM*, et al.* Positional cloning of Sorcs1, a type 2 diabetes quantitative trait locus. *Nat Genet* 2006; **38:** 688-93.

362 Granhall C, Park HB, Fakhrai-Rad H, Luthman H. High-resolution quantitative trait locus analysis reveals multiple diabetes susceptibility loci mapped to intervals<800 kb in the species-conserved Niddm1i of the GK rat. *Genetics* 2006; **174:** 1565-72.

363 Goodarzi MO, Lehman DM, Taylor KD*, et al.* SORCS1: a novel human type 2 diabetes susceptibility gene suggested by the mouse. *Diabetes* 2007; **56:** 1922-9.
364 McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet* 2008; **17:** R156-65.
365 Patel CJ, Bhattacharya J, Butte AJ. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. *PLoS One* 2010; **5:** e10746.
366 Manolio TA, Collins FS, Cox NJ*, et al.* Finding the missing heritability of complex diseases. *Nature* 2009; **461:** 747-53.
367 McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; **9:** 356-69.
368 Edmondson AC, Rader DJ. Genome-wide approaches to finding novel genes for lipid traits: the start of a long road. *Circ Cardiovasc Genet* 2008; **1:** 3-6.
369 Lango Allen H, Johansson S, Ellard S*,* Shields B, Hertel JK *et al.* Polygenic risk variants for type 2 diabetes susceptibility modify age at diagnosis in monogenic HNF1A diabetes. *Diabetes* 2010; **59:** 266-71.
370 Molven A, Njølstad PR. Role of molecular genetics in transforming diagnosis of diabetes mellitus. *Expert Rev Mol Diagn* 2011; **11:** 313-20.
371 Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 2010; **11:** 415-25.
372 Bonnefond A, Clement N, Fawcett K*, et al.* Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nat Genet* 2012; **44:** 297-301.
373 Yokoyama S, Woods SL, Boyle GM*, et al.* A novel recurrent mutation in MITF predisposes to familial and sporadic melanoma. *Nature* 2011; **480:** 99-103.