UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH AND MEDICAL SCIENCES

A decade of genome discoveries in type 2 diabetes and metabolism

Doctoral dissertation

Niels Grarup, MD, PhD

A decade of genome discoveries in type 2 diabetes and metabolism

Niels Grarup, MD, PhD
Novo Nordisk Foundation Center for Basic Metabolic Research
University of Copenhagen

Doctoral dissertation
Faculty of Health and Medical Sciences
University of Copenhagen
2020

The Faculty of Health and Medical Sciences at the University of Copenhagen has accepted this dissertation for public defence for the doctoral degree in Medical Science.

Copenhagen, 26 June 2020

Ulla Wewer, Head of Faculty

The defence will take place on 27 November 2020 at 2 pm in Victor Haderup Auditorium at the Panum Institute, Blegdamsvej 3B, 2200 Copenhagen N, Denmark.

Professor Sten Madsbad, University of Copenhagen, will chair the defence.

Assessment Committee:

Professor Hindrik Mulder, Lund University.

Professor Torben Kruse, University of Southern Denmark.

Professor Allan Flyvbjerg, University of Copenhagen (Chairman).

ISBN 978-87-93510-59-3

CONTENTS

LIST OF PAPERS ON WHICH THE DISSERTATION IS BASED

[1]

Grarup N*, Rose CS*, Andersson EA, Andersen G, Nielsen AL, Albrechtsen A, Clausen JO, Rasmussen SS, Jørgensen T, Sandbæk A, Lauritzen T, Schmitz O, Hansen T, Pedersen O: 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, 56:3105-3111

[2]

Grarup N, Andersen G, Krarup NT, Albrechtsen A, Schmitz O, Jørgensen T, Borch-Johnsen K, Hansen T, Pedersen O: 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, 57:2534-2540

[3]

Boesgaard TW*, **Grarup N***, Jørgensen T, Borch-Johnsen K, Meta-Analysis of Glucose and Insulin-Related Trait Consortium (MAGIC), Hansen T, Pedersen O: Variants at *DGKB*/*TMEM195*, *ADRA2A*, *GLIS3* and *C2CD4B* loci are associated with reduced glucose-stimulated beta cell function in middle-aged Danish people. *Diabetologia* 2010, 53:1647-1655

[4]

Grarup N, Overvad M, Sparsø T, Witte DR, Pisinger C, Jorgensen T, Yamauchi T, Hara K, Maeda S, Kadowaki T, Hansen T, Pedersen O: The diabetogenic *VPS13C*/*C2CD4A*/*C2CD4B* rs7172432 variant impairs glucose-stimulated insulin response in 5,722 non-diabetic Danish individuals. *Diabetologia* 2011, 54:789-794

[5]

Albrechtsen A*, **Grarup N***, Li Y*, Sparsø T*, Tian G*, Cao H, Jiang T, Kim SY, Korneliussen T, Li Q, Nie C, Wu R, Skotte L, Morris AP, Ladenvall C, Cauchi S, Stancakova A, Andersen G, Astrup A, Banasik K, Bennett AJ, Bolund L, Charpentier G, Chen Y, Dekker JM, Doney AS, Dorkhan M, Forsen T, Frayling TM, Groves CJ, Gui Y, Hallmans G, Hattersley AT, He K, Hitman GA, Holmkvist J, Huang S, Jiang H, Jin X, Justesen JM, Kristiansen K, Kuusisto J, Lajer M, Lantieri O, Li W, Liang H, Liao Q, Liu X, Ma T, Ma X, Manijak MP, Marre M, Mokrosinski J, Morris AD, Mu B, Nielsen AA, Nijpels G, Nilsson P, Palmer CN, Rayner NW, Renstrom F, Ribel-Madsen R, Robertson N, Rolandsson O, Rossing P, Schwartz TW, DESIR Study Group, Slagboom PE, Sterner M, the Diagram Consortium, Tang M, Tarnow L, Tuomi T, van't Riet E, van Leeuwen N, Varga TV, Vestmar MA, Walker M, Wang B, Wang Y, Wu H, Xi F, Yengo L, Yu C, Zhang X, Zhang J, Zhang Q, Zhang W, Zheng H, Zhou Y, Altshuler D, 't Hart LM, Franks PW, Balkau B, Froguel P, McCarthy MI, Laakso M, Groop L, Christensen C, Brandslund I, Lauritzen T, Witte DR, Linneberg A, Jørgensen T, Hansen T, Wang J, Nielsen R, Pedersen O: Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. *Diabetologia* 2013;56:298-310

[6]

Grarup N*, Sulem P*, Sandholt CH*, Thorleifsson G, Ahluwalia TS, Steinthorsdottir V, Bjarnason H, Gudbjartsson DF, Magnusson OT, Sparsø T, Albrechtsen A, Kong A, Masson G, Tian G, Cao H, Nie C, Kristiansen K, Husemoen LL, Thuesen B, Li Y, Nielsen R, Linneberg A, Olafsson I, Eyjolfsson GI, Jørgensen T, Wang J, Hansen T, Thorsteinsdottir U, Stefansson K, Pedersen O: Genetic architecture of vitamin B12 and folate levels uncovered applying deeply sequenced large datasets. *PLoS Genet* 2013, 9:e1003530, 12 pages

[7]

Moltke I*, **Grarup N***, Jørgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, Korneliussen TS, Andersen MA, Nielsen TS, Krarup NT, Gjesing AP, Zierath JR, Linneberg A, Wu X, Sun G, Jin X, Al-Aama J, Wang J, Borch-Johnsen K, Pedersen O, Nielsen R, Albrechtsen A, Hansen T: A common Greenlandic *TBC1D4* variant confers muscle insulin resistance and type 2 diabetes. *Nature* 2014, 512:190-193

[8]

Grarup N*, Moltke I*, Andersen MK*, Bjerregaard P, Larsen CVL, Dahl-Petersen IK, Jørsboe E, Tiwari HK, Hopkins SE, Wiener HW, Boyer BB, Linneberg A, Pedersen O, Jørgensen ME, Albrechtsen A, Hansen T: Identification of novel high-impact recessively inherited type 2 diabetes risk variants in the Greenlandic population. *Diabetologia* 2018, 61:2005-2015

[9]

Grarup N*, Moltke I*, Andersen MK, Dalby M, Vitting-Seerup K, Kern T, Mahendran Y, Jørsboe E, Larsen CVL, Dahl-Petersen IK, Gilly A, Suveges D, Dedoussis G, Zeggini E, Pedersen O, Andersson R, Bjerregaard P, Jørgensen ME, Albrechtsen A, Hansen T: Loss-of-function variants in *ADCY3* increase risk of obesity and type 2 diabetes. *Nat Genet* 2018, 50:172-174

* Denotes equal contributions on the specific paper.

ACKNOWLEDGEMENTS

The work of this doctoral dissertation was performed in the period 2007-2018 at Steno Diabetes Center, at Hagedorn Research Institute and since 2011 at Novo Nordisk Foundation Center for Basic Metabolic Research at University of Copenhagen.

The research presented here was among others funded by the Danish Clinical Intervention Research Academy, the Danish Council for Independent Research, Medical Sciences, the Lundbeck Foundation and the Novo Nordisk Foundation. I am very grateful for the continued support from these sources.

I would like to express my appreciation to Oluf Borbye Pedersen and Torben Hansen. Oluf and Torben have over the years been my mentors and sources of inspiring and motivation. They have taught me many important things about conducting and disseminating scientific research. I admire their continued enthusiasm and dedication to research and I am grateful for their influence on my research and career.

Over the years, I have worked together with many researchers. Some of my key colleagues have been Trine Boesgaard, Trine Nielsen, Anette Gjesing, Anders Albrechtsen, Ida Moltke, Camilla Sandholt, Mette Andersen and Thomas Sparsø. I am grateful for your involvement, interaction and positive spirit. Many thanks for hard work and good times. I would also like to thank all the technicians in the research group – especially Annemette Forman and Tina Lorentzen.

I would like to thank all national and international collaborators and co-authors of the studies on which this dissertation is based. A special acknowledgement goes to the key collaborators Torben Jørgensen, Allan Linneberg, Daniel Witte and Marit Jørgensen, who have been crucial in building the resources on which my research depends.

None of this would have been possible without support and devotion from my family, Carolina, Cecilia, Alberte, Hugo and Alexandra. You have been great in keeping my balance and showing me what is important in life.

I dedicate this dissertation and the achievement to my deceased father, Olaf, who was a steady and full support with a keen interest in my work and me.

Niels Grarup
Farum, December 2018

SUMMARY

This doctoral dissertation is based on nine scientific papers published in the period 2007-2018. The overall aim of the present research has been to contribute to the discovery and characterisation of genetic variation contributing to the risk of type 2 diabetes (T2D) and related metabolic traits.

T2D is a disease with involvement of processes in multiple tissues and organs. Prominent characteristics are obesity, insulin resistance and insufficient insulin secretion. Risk factors for development of T2D involve both lifestyle factors and a genetic predisposition. Research within the past decade has elucidated a large number of genomic loci associated with T2D and obesity primarily found through investigations of the European population through genome-wide association studies. These genetic variants are generally common in the population and individually inflict modest increases in risk of disease. Evidence points to a large number of risk variants below genome-wide statistical significance having an impact on disease risk, together contributing to the high degree of genetic heterogeneity of T2D, obesity and complex traits in general.

The nine papers included in the doctoral dissertation fall into three overall categories. Papers 1-4 had the overall common aim to investigate the more detailed physiological impact of genetic variation associated with T2D or glycaemia. These papers showed that a number of genetic variants associated with T2D and glycaemia have an intermediary effect on the ability of the pancreas to secrete an appropriate amount of insulin, a fact which at that point in time was a new realisation. These studies teach important lessons on the biology and pathophysiology of T2D and serve as a starting point for more detailed mechanistic investigations into specific disease-associated genetic loci.

Papers 5 and 6 had the common aim to apply large-scale nucleotide sequencing to discover genetic risk elements for metabolic traits under the hypothesis that rare and low-frequency protein-coding genetic variation has an impact on these phenotypes. In Paper 5, a large-scale exome sequencing study in T2D was performed, which through three stages of discovery and replication analyses detected three loci associated with one or more of 12 selected metabolic traits. The sequencing study was performed at a stage where large-scale sequencing was unproven, which influenced the conclusions of the study in terms of the effect of rare variants on disease risk. As such, the study serves as a forerunner for later studies on an even larger scale. In paper 6, the data generated in paper 5 were used in combination with data from DeCode Genetics to elucidate genetic factors behind the variation in circulating concentrations of vitamin B_{12} and folate. Here a number of loci were associated with variation in these vitamins in the general population highlighting that the identified genetic factors lie within proteins with a known role in vitamin B_{12} and folate metabolism.

Papers 7-9 aimed to investigate the impact of genetic variation on T2D and obesity in a small and isolated population. The small and historically isolated population of Greenland displays specific genomic characteristics, which make this population an alternative to the much-studied European population in the search for genetic risk factors for metabolic disease. In paper 7, we identified a common nonsense variant in *TBC1D4* with a high recessive impact on hyperglycaemia, hyperinsulinemia and T2D caused by post-prandial insulin resistance. Two further loci associated with T2D under a recessive genetic model were identified in paper 8. In paper 9, we zoomed in on putative loss-of-function variation and identified a low-frequency splice variant in *ADCY3*, which displayed a high impact on adiposity and prevalence of T2D. Together with accompanying papers, these findings point to novel obesity-related biology. In general, papers 7-9 have revealed aspects of the genetic architecture of T2D and obesity in the Greenlandic Inuit population and displayed important biological and clinical subtypes of these diseases. Thereby these studies have shown the potentials of investigating the genetic contribution to metabolic traits in isolated populations.

Genetic studies of Europeans and other populations have shed light on the genetic architecture of T2D, obesity and other metabolic diseases and have shown differences between populations. These findings have huge implications for the possibility of applying genetic information in precision medicine of metabolic diseases. Further expansion of the mechanistic insights is leading to translation of association signals into clinical useful knowledge, especially in development of

novel drugs, but this process has proven difficult and tedious. In the future, studies of mega-size cohorts and biobanks with deeply phenotyped samples coupled to clinical information will probably disentangle more clinically relevant omics-related knowledge with an impact on future precision medicine in metabolic disease. Furthermore, an increased focus on translational and mechanistic research to make sense of the hundreds of loci associated T2D and obesity will inevitably reveal biological knowledge of huge importance for disease understanding and future drug development.

BACKGROUND AND AIMS

Introduction

The last decade has brought enormous advances in the understanding of the genetic influence on complex disease. Genomic discoveries in type 2 diabetes (T2D), once called "the geneticist's nightmare", and related traits have been tremendous leading to hundreds of genomic loci firmly associated with risk of disease or inter-individual variation in metabolic phenotypes. However, these discoveries have also made it evident that the genetic architectures of complex metabolic traits are highly complex with a huge level of genetic heterogeneity. There has been a growing sense that genomic technologies and discoveries will revolutionise the practice of clinical medicine and bring about a paradigm shift in the way we handle health care for the individual in the modern society. Specifically, it has been anticipated that these advances will eventually lead to a new model of health care focussed on disease prevention and on disease treatments that are tailored to smaller subsets of patients – so-called precision medicine.

Type 2 diabetes – a common disease with a complex aetiology

The prevalence of obesity and T2D are increasing dramatically on a global scale. About 10% of the global population already has T2D or is likely to develop it, and ~40% of adults are overweight or obese. In 2013, the International Diabetes Federation estimated that 382 million adults aged 20–70 years worldwide had T2D, with 80% of those affected living in low- and middle-income countries. This number is expected to rise to 592 million by 2035 [10]. China and India are areas particularly affected by T2D, where the prevalence of T2D has increased dramatically despite the relatively low prevalence of obesity [11].

T2D is a disease with a complex pathophysiology including multiple different organs and processes. T2D is associated with major disturbances in several physiological responses: insulin secretion from pancreatic beta cell is impaired, fasting plasma glucagon secretion from pancreatic alpha cells is increased and does not suppress normally after a meal, basal hepatic glucose production is increased, muscle and adipose tissue glucose uptake is impaired, fasting plasma fatty acid levels are increased and there is resistance to the stimulatory effect of incretion hormones, glucagon-like peptide 1 (GLP1) and gastric inhibitory polypeptide, on insulin secretion. Further pathophysiological characteristics of T2D are related to increased renal glucose absorption, changes in the response of the brain to hormones such as insulin and GLP1 combined with adipose tissue inflammation and systemic low-grade inflammation [12, 13]. Hyperglycaemia, determined from plasma glucose or HbA_{1C}, defines the diagnosis of diabetes and is generally seen as a consequence of a complex interplay between insulin sensitivity and insulin secretion, with a failure of pancreatic beta cells to compensate appropriately for the increased insulin requirement induced by insulin resistance [14]. Whether insulin resistance or insufficient insulin secretion represent the primary defect in the pathogenesis of T2D has been a matter of debate [15, 16]. Since glucose tolerance is achieved by the combination of an appropriate insulin secretion and sufficient insulin action, increases in glycaemia can present when one component decreases and there is no concomitant improvement in the other, illustrating the fact that in a healthy individual any decrease in insulin sensitivity is compensated by an increase in insulin secretion [17-19] (Fig. 1). The fact that many individuals who are obese and insulin resistant never develop T2D shows the need for a beta cell defect for overt hyperglycaemia to manifest. Thus, both derangements are necessary, but not sufficient to reach the levels of hyperglycaemia that yield a clinical diagnosis.

Lifestyle and environmental factors are crucially important for the development of obesity and T2D. Important risk factors for obesity are physical inactivity, excessive energy intake, depression, sleep disorders and low socio-economic status while major risk factors for T2D include obesity, especially a visceral fat deposition, physical inactivity, smoking, male sex, high age, sleep deprivation, urbanization, low-socio-economic status and ethnicity [20-23].

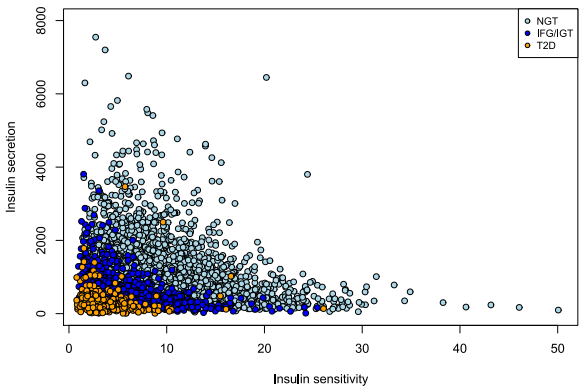

Figure 1. Illustration of the relationship between insulin sensitivity and insulin response based on data from an oral glucose tolerance test in 5331 individuals of the Inter99 cohort. Insulin sensitivity was estimated by the Matsuda insulin sensitivity index [24] and glucose-stimulated insulin response by the corrected insulin response [25]. Diabetes status is indicated as normal glucose tolerance (NGT), impaired fasting glycaemia (IFG), impaired glucose tolerance (IGT) or type 2 diabetes (T2D).

Multiple lines of evidence support the view that genetic component plays an important role in the pathogenesis of T2D and the major T2D risk factor, obesity. Evidence for the genetic importance for the development of T2D comes from studies showing familial aggregation [26] and from twin studies [27-29]. In general, 35-50% of the risk of T2D is estimated to come from genetic factors. Also, for T2D-related quantitative phenotypes, a substantial part of the inter-individual variation is influenced by genetic factors. This is the case for both insulin secretion and insulin sensitivity [30-32] and for body mass index (BMI), for which 50-70% of the variation is estimated to be driven by genetic determinants [33-35].

Aims and hypotheses

The overall aim of the present research has been to contribute to the discovery and characterization of genetic variation contributing to the risk of T2D and related metabolic traits. The doctoral dissertation is based on nine original papers, which fall in three overall categories. Papers 1-4 had the overall common aim to investigate the more detailed physiological impact of genetic variation associated with T2D or glycaemia. These studies teach important lessons on the biology and pathophysiology of T2D. Papers 5-6 had the common aim to apply large-scale nucleotide sequencing to discover genetic risk elements for metabolic traits under the hypothesis that rare and low-frequency coding variation has an impact on metabolic traits. Papers 7-9 aimed to investigate the impact of genetic variation on T2D and obesity in a small and isolated population showing aspects of the genetic architecture of T2D and obesity in Greenlandic Inuit and displaying important biological and clinical subtypes of these diseases.

In the following, I will review the status of genetic research in T2D and obesity in light of the developments during the past decade in relation to the nine papers. The review falls into five parts and the topics which will be covered include 1) the current knowledge of T2D genetic risk factors in the European population, 2) a discussion of the physiological impact of T2D-associated genetic variants, 3) the influence of sequencing-based studies on current knowledge of genetic landscape of T2D and related traits, 4) the genetics of T2D in isolated populations and finally 5) an overall recapitulation of the past decade of genetic research on the major research objectives of this research including a discussion of the future perspectives of this research.

PART 1. GENETIC IMPACT ON TYPE 2 DIABETES AND RELATED TRAITS IN THE EUROPEAN POPULATION

Opening the human genome

In 2001, the initial sequencing and analysis of the human genome was published in the Nature journal as a huge paper based on the collective work of the Human Genome Project [36] with a companion paper reporting a map of more than 1 million genetic variants [37]. Furthermore, a parallel private sequencing initiative was also published [38]. These projects and papers formed the basis of a revolution in the study of genomics and of the genetic impact on human diseases and phenotypes. It was the expectation that the completion of the Human Genome Project would mark the beginning of a new era of genomic medicine, in which new approaches to discovery research, disease prediction and treatment would develop from an improved understanding of the genetic risk factors of human disease. However, the journey towards these ultimate goals has been longer and filled with more obstacles than initially expected [39]. The translation from simple nucleotide sequence and sequence variation, via the complex time- and cell-dependent transcriptome and proteome to biological and clinical disease breakthroughs is enormously complex.

Studying the genomic impact on risk of type 2 diabetes and metabolic traits

Over the years, different approaches have been taken to the search for genetic determinants of T2D and other complex diseases. The technological advances in laboratory and in computer methods have largely directed the development in this research field, while also other factors such as lessons learned from early approaches have formed the basis for the status of this research field. Before the existence of genome-wide association studies (GWAS), the investigation of genetic determinants of the polygenic inheritance of common metabolic diseases and phenotypes was primarily done by the candidate gene approach or by genetic mapping by linkage analysis. These two early approaches were largely unsuccessful in identifying the genetic determinants of T2D. The linkage analysis approach seeks evidence of co-segregation between genomic markers and as such screens the entire genome for genomic susceptibility regions using a limited number of highly polymorphic microsatelitte markers [40]. This approach was rather successful in detecting genes involved in monogenic Mendelian diseases but had low statistical power in complex diseases dominated by low penetrance of risk alleles [41]. Yet, the T2D susceptibility locus, *TCF7L2* was discovered by typing of microsatellite markers in a region previously identified by linkage [42, 43]. The association between the *TCF7L2* locus and T2D has since been widely replicated and it remains the strongest common risk locus in the European population with an odds ratio (OR) of 1.4 per allele [44-47].

In contrast to the linkage approach, the candidate gene approach is not agnostic but relies on an *a priori* hypothesis that a certain genetic variant, gene or genomic region is involved in disease pathogenesis. Overall, the candidate gene approach has not been successful in finding validated associations between genetic variants and T2D or related traits; however, a few positive examples exist. For T2D, two loci have been found based on this method. Association between the p.Pro12Ala variant in the *PPARG* locus and T2D was demonstrated and has subsequently been replicated in large-scale GWAS [46, 48-50]. *PPARG* was initially investigated as a candidate gene substantiated by being the receptor of the thiazolidinediones class of antidiabetic medicine. Furthermore, the common p.Glu23Lys variant in *KCNJ11* has been shown to increase risk of T2D in association studies [51, 52] and replicated in GWAS [53]. As anticipated from the function of the gene product, this variant influence insulin secretion from the pancreatic beta cell [51, 52]. Despite such examples, the number of negative or false positive studies are vast and for a number of years they have contaminated the scientific literature. Several reasons for the general failure of this approach exist and while some are related to the method itself, others are more a reflection of the era of genetic research in which this approach was widely used, i.e. limitations of genetic association studies in the early phases of molecular genetic epidemiology. First, a major limitation of the candidate gene approach is the fundamental need to have a detailed knowledge of the disease of interest to be able to select a reasonable

candidate gene. Second, at the time in which most candidate gene studies were performed, the rather low sample sizes applied in genetic association studies meant that these studies were vastly statistically underpowered to detect associations of the effect sizes we now know that common variants infer. While some of this can be overcome by applying meta-analysis, as illustrated by the study of the *PPARG* p.Pro12Ala variant [49], other biases, such as publication bias, may influence the outcome of literature-based meta-analysis. Third, the genomic coverage of especially early candidate studies was inadequate and mostly gene-centric; while evidence from GWAS shows that, the majority of associated variants are found in non-coding genomic content [47]. Forth, insufficient correction for multiple testing was a major problem, which has contributed to false positive and non-replicated findings. Fifth, genetic association studies performed in a case-control design including only a single or few variants cannot be corrected for population stratification or other sources of inflation increasing the risk of false positive conclusions. Contrary, this is adequately corrected for in GWAS design [54, 55]. The research field investigating T2D genetics needed a change.

The era of genome-wide association studies

With the development of array-based genotyping technology, thousands of genetic variants could be genotyped in a single experiment leading to the design of GWAS. The design and pursuit of GWAS rely on the "common disease – common variant" hypothesis which states that the genetic variation responsible for most of the disease risk in the population are shared across members of the population [56-58]. The powerfulness of the design of GWAS relies on resolution of many of the shortcomings described above for the candidate gene approach. As such, GWAS is an agnostic approach in the sense that no prior hypothesis is put forward for any specific variant or gene. Furthermore, a genome-wide significance level of 5×10^{-8} has been adopted to account for the high multiple testing burden [59] and together with the possibility to correct for population stratification this has minimized spurious findings. Meanwhile, sample sizes of genetic association studies have increased to obtain statistical power to draw firm conclusions, because of establishment of large human cohorts, falling prizes of array-based genotyping and due to increasing willingness to collaborate internationally in large-scale meta-analyses (Fig. 2).

Genomic coverage of GWAS has increased over the years due to more dense genotyping arrays and since studies have started applying genotype imputation. Large-scale sequencing efforts have been essential for this development, since genotyping as such can only measure already known variation. In genotype imputation, non-genotyped genetic variation is inferred from reference panels such as HapMap [60], 1000 Genomes Project [61], UK10K [62] or most recently the Haplotype Reference Consortium (HRC), which includes whole-genome sequence data from more than 30,000 individuals establishing an unprecedented depth in genome sequence variation data [63].

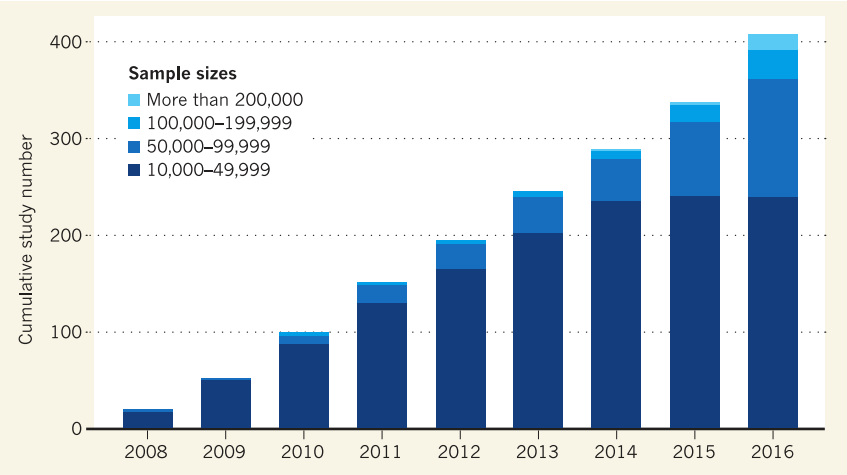

Figure 2. Illustration of the number of GWAS publications stratified on sample size during the years 2008-2016. Source: GWAS catalogue at https://www.ebi.ac.uk/gwas/. Modified from ref. [64] with permission.

Increasingly large and dense sequencing-based reference panels and improvement in statistical methods for inference have led to increasing coverage of variation leading to a situation in which a modern GWAS evaluates in the order of 10-30 million genetic variants with minor allele frequency down to 0.5% or maybe even 0.1% [63] (Fig. 3). For even rarer variants, genotype imputation is still inaccurate and does not provide sufficient coverage [63], and in this frequency space, sequencing is the method of choice.

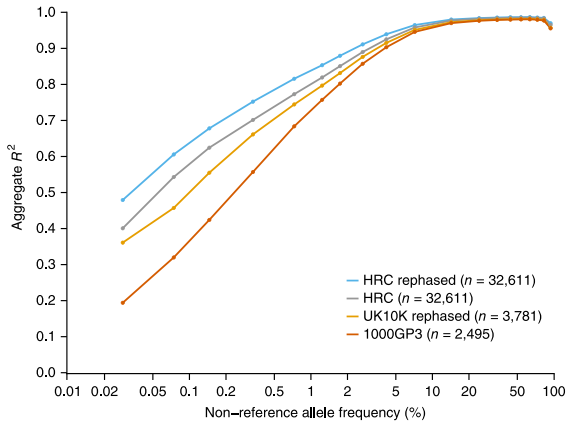

Figure 3. Imputation quality measured as R^2 dependent on minor allele frequency for different reference panels. Modified from ref. [63] with permission.

With the advent of large common resources such as the UK Biobank, which holds ~500,000 participants with rich phenotype data and genome-wide genotyping [65, 66], modern GWAS are reaching what would previously have been seen as an extreme sample size approaching 1 million individuals [47, 67, 68]. As described below, this combination of increased sample size and genomic coverage is currently leading to a new wave of studies and discoveries of specific genetic risk variants, which leads to a description of the genetic architecture of complex diseases at unprecedented levels of detail and to large-scale evaluation of the genetic relationship between different diseases and traits.

By convention, loci are named for the gene closest to the lead single-nucleotide polymorphism (SNP), which is not to say that the gene in question is necessarily the functional gene through which the disease mechanism is working. Similarly, the identified most strongly associated variant at a locus, often referred to as the lead variant, is not necessarily the causal variant, but may merely be correlated to a measured or unmeasured functional variant. For the majority of disease-associated loci, the causative variant and gene are unknown.

GWAS discoveries of the genetics of type 2 diabetes, obesity and metabolic traits

During the first years, GWAS was characterised by relatively low sample size in the discovery data investigating a relatively low number of genetic variants. For instance, the first published GWAS of T2D was performed in ~1400 cases and controls with information on ~400,000 variants, yet was nevertheless able to detect novel loci genuinely associated with T2D [69]. Besides replicating the previously identified *TCF7L2* locus [43], the study identified risk variants in *HHEX* and *SLC30A8* loci [69]. This study was quickly followed by a number of other studies [70-74]. The first GWAS based on genotype imputation and meta-analysis of several studies was published in 2008 being a landmark paper as the first step in future developments over the next decade [50]. Here six novel loci were identified, among others *JAZF1* [50] and since a number of GWAS of T2D have been reported finding an increasing number of associations bringing the number of association signals above 400 [47, 50, 53, 69, 73, 75]. This is in striking contrast to the three loci (*PPARG*, *KCNJ11* and *TCF7L2*), which were found prior to the introduction of GWAS and it must be seen as a great methodological success.

Most findings for T2D have been performed within the Diabetes Genetics Replication and Meta-analysis Consortium, which has published a number of papers including an increasing number of individuals and an increasing genomic coverage [50, 53, 75]. The most recent and largest GWAS thus far for T2D was published in 2018 [47]. The data included here were from ~900,000 individuals with genotype data imputed to the HRC panel testing ~27 million variants, hence being the most thorough search for T2D risk variants thus far, representing a ~45,000-fold increase in number of individual genotypes over the first GWAS from 2007. In the paper, 243 loci, including 403 distinct signals, associated with T2D were reported (Fig. 4) and this seminal paper can be used to describe the genetic landscape in the European population. These T2D risk variants are predominantly common with minor allele frequency >5% (323 of 403 distinct variants), which impose rather low risk increments with odds ratios up to 1.37 per allele. Furthermore, a number of low-

frequency and rare variants were identified as contributors to genetic risk T2D. The topic of low-frequency and rare variants in common diseases as T2D will be discussed in more detail in Part 3. Another characteristic of the identified loci is that at many loci, more than one distinct association signal exists. At 151 loci, a single signal was identified while two independent signals were identified at 57 loci and the remaining 35 loci contained three or more distinct signals (Fig. 4). This implies that genomic hotspot genes carry a number of T2D risk variants and may point to these loci and genes being biologically more important in T2D pathogenesis. Furthermore, since a number of the secondary signals at these loci are represented by low-frequency or rare variants, these loci also form a bridge between common variants with low impact via low-frequency or rare variants with intermediary effects to monogenic diabetes-causing variants imposing high penetrance in affected families. For instance, this is observed for the *HNF1A* locus, which will be discussed in more detail in Part 5. Finding such key genes related to several types of diabetes are of great importance, since it directly implicates the specific gene in the disease pathogenesis.

Another interesting effort undertaken in many modern GWAS is to use statistical methods to try to determine which variant in a given locus may be the causal variant. This is of great interest both for understanding the genetic architecture and for further studies into biological mechanisms from association to an impact on disease risk. In the most recent T2D GWAS [47], statistical fine-mapping efforts demonstrated a posterior probability >80% for the lead variant at 51 loci and at 18 loci the probability for the lead variant being causal was >99% and the credible set of putative causal variants contained just the lead variant. Similarly, for approximately half of the 403 distinct association signals, the genetic credible sets of variants include less than 51 variants. Although this is a major improvement over previous fine-mapping efforts [76], the majority of loci are still without a clear causal candidate variant. Hence, a lot of research is still needed to discover causal genes and mechanisms in the associated loci. While the number of genomic loci and variants associated with T2D is substantial, together they explain around 17.4% of phenotypic variance – a number that has with the recent study increased substantially from earlier reports [53, 75, 76].

One of the expectations to the achievements of elucidating the genetic determinants of diseases such as T2D were related to the possibility to use genetic information to predict future disease, which could enhance early screening for T2D and allow for targeted prevention. However, studies applying the early findings from GWAS to predict T2D were of limited success. For instance, Meigs et al. investigated prediction from 18 variants in 2377 participants of the Framingham Offspring Study and showed that the genotype score only provided a slightly better prediction of T2D than that of classical risk factors alone [77]. These findings are in line with other studies of the same question [78, 79]. It is expected that identification of more common variants with very low effect sizes would only increase predictive ability slightly. With novel methods applying genome-wide polygenic risk scores on densely imputed data sets, it is possible to evaluate the overall contribution of genome-wide variation on disease under the assumption that heritability can be ascribed to a long tail of variants of individually very tiny effects. A polygenic risk score is most commonly calculated as a weighted sum of the number of risk alleles carried by an individual, where the risk alleles and their weights are defined by the loci and their estimated effects as detected by GWAS [80]. In the most recent GWAS of T2D, genetic prediction was tested for a polygenic risk score of more than 130,000 variants. Here a maximum prediction with a C-statistic of 66% was found and individuals at the 2.5% upper extreme of this score had a 3.4-fold increased risk of T2D compared to the median and a 9.4-fold increased risk compared to the lower 2.5% of the score [47].

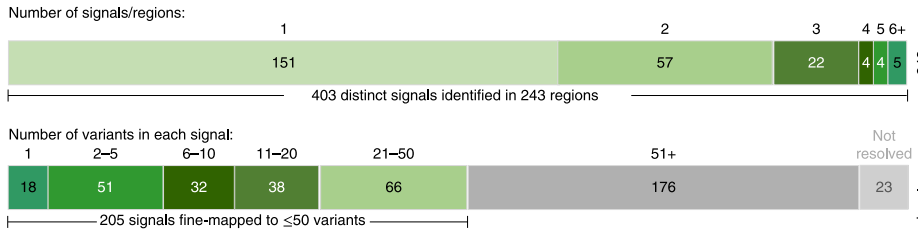

Figure 4. Identification and fine-mapping of genetic variants and loci in T2D. Number of distinct association signals in 243 loci associated with T2D in GWAS (top) and number of variants in genetic credible sets for 403 distinct signals (bottom). Modified from ref. [47].

These risk estimates translate to estimated absolute lifetime risks of T2D of 51% and 5.5% for these extremes, respectively, thus a very sizeable difference. Along the same lines, a recent paper investigated polygenic risk scores for several metabolic phenotypes and defined a polygenic risk score of 6.9 million variants with non-zero effects on T2D. This score defined 3.5% of the population being at 3-fold increased risk compared to the remaining population [81] (Fig. 5). Of interest, in the tails of the distribution of the polygenic risk score, the risk estimates deviated from the allelic additive effect. In essence, it seems as if applying polygenic risk scores may increase the predictive ability for T2D, which could be of clinical relevance especially in defining a smaller subset at particularly increased risk.

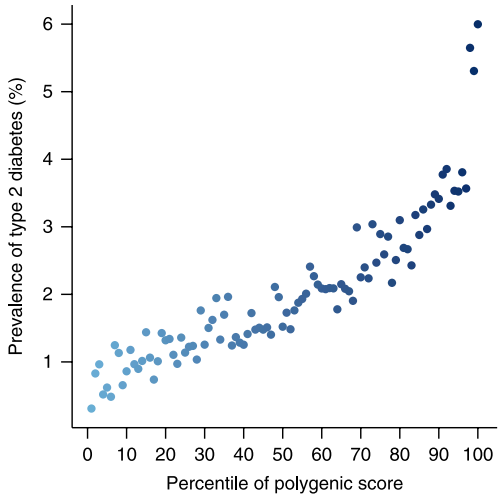

Figure 5. Polygenic risk score in relation to prevalence of T2D in the 294,831 UK Biobank participants. From ref. [81] with permission.

These risk predictions can theoretically be made at an early time in life, yet it remains to be seen if risk estimates are of similar magnitude in such a setting. In addition, a more specific evaluation of the predictive potential of genome-wide polygenic risk scores compared to other clinical parameters must be performed and their preventive and therapeutic benefits remain unproven [82]. In addition, to put this into perspective, a prognostic marker with an OR of 3.0 that correctly identifies 80% of persons who will develop T2D would incorrectly classify 60% of persons who will not develop T2D [83] and this degree of discrimination is by itself not clinically useful [84]. Comparable results have been found for cardiovascular disease (CVD) [81, 85] potentially opening for clinical genotyping to detect subsets a particularly high risk [86].

Besides GWAS of T2D, other sources of identification of loci of importance for T2D have been through studies of basic quantitative diabetes-related traits. These efforts have discovered more than 70 loci associated with quantitative traits reflecting glucose homeostasis, i.e. fasting glucose, fasting insulin, 2-hours glucose during an oral glucose tolerance test (OGTT) and HbA_{1C} [87-89]. Many of these loci do also associate with T2D, yet the overlap between loci for these traits is not extensive as illustrated in Figure 6 for the loci discovered until 2014 (Fig. 6). These findings indicate that some gene variants may impose modifying effects on fasting glucose levels in the general population while others have specific thresholds at which the genetic effect sets in thereby inflicting risk of T2D without modifying levels of fasting glucose at the population level.

For other complex metabolic traits, many aspects of the development have been very similar to that of T2D. Initial studies for of BMI and obesity picked up the variants with highest effect sizes such as the *FTO* [90] and *MC4R* [91] loci and the *FTO* variant remains the common variant with the highest impact on BMI [90, 92]. Since then, a number of rounds of GWAS have enlightened the genetics of both BMI and body fat distribution, measured as waist-hip-ratio adjusted for BMI and identified a high number of associated obesity-related loci [92-97]. The most recent GWAS combined the results from studies performed by the GIANT consortium for BMI [92] and body fat composition [95] with results of association analyses of participants of the UK Biobank ending at sample size of more than 700,000 individuals. As is the case for T2D, these studies have provided unprecedented detail in the description of the genetic landscape of genetic disposition to obesity, while they have also revealed a number of statistical caveats of analysing such a large set of individuals for instance related to difficulties in separating the effect of subtle population substructure and the effect of a high degree of polygenicity [98]. For BMI, the novel analysis increased the number of associated variants to 716 consisting of 450 primary associations and 266 secondary associations [98].

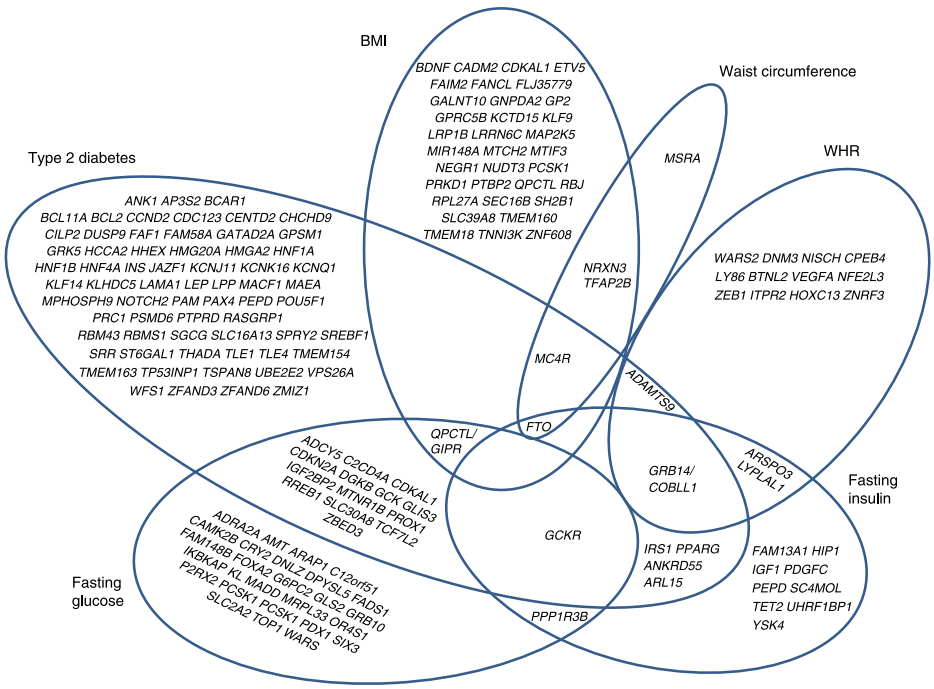

Figure 6. Overlap between genomic loci associated with T2D, glycaemic traits and obesity-related traits discovered until 2014. From ref. [99].

Yet, this set of genome-wide statistically significant variants still explains a mere 5% of the variance of BMI, which is far from the expected genetic contribution of 50-70% [33-35]. Applying a more liberal significance threshold included more variants into the associated set of variants and increased the explained variance in BMI to approximately 10% [98]. Yet, a previously published analysis of BMI estimated an explained proportion of variance from common (minor allele frequency (MAF) >1%) genome-wide variants of 27%. Considering the imperfect tagging of genotype imputation and the possible overestimation of heritability from twin- or family-based studies, this finding makes it likely that genome-wide genotype data may explain most of the genetic contribution to BMI [100]. For waist-hip-ratio adjusted for BMI, as an indicator of body fat distribution, 463 association signals were detected in 346 loci in almost 700,000 individuals [68]. Despite the high number of variants associated with waist-hip ratio, these variants combined only explained ~4% of the variation in the trait. Fat distribution shows a strong sex-dimorphism in humans and genetic studies have shown that this is at least partly genetically determined. Of the 463 variants associated in the sex-combined dataset, 105 showed sex-dimorphism – 97 of which had a stronger effect in women [68]. Analyses of BMI and T2D have not shown similarly strong sex-dependent effects [92, 98]. Of high interest, studies have indicated that variance in BMI is generally genetically regulated by genes in the brain [92], while body fat distribution is generally attributed to genes expressed in the adipose tissue [95], however, the new studies do not shed further light on this aspect.

Although T2D and obesity are highly interrelated from both epidemiological and pathophysiological viewpoints, the shared genetic aetiology imposed by hitherto identified common variants is limited. Figure 6 shows the status as per 2014 and of the 90 loci associated with T2D and 56 loci associated with standard measures of adiposity, merely five loci are shared (*FTO*, *MC4R*, *ADAMTS9*, *GRB14/COBLL1* and *GIPR*). While some data indicate that a major reason for this lack of genetic locus overlap may be lack of statistical power to identify the minute effects inflicted on T2D by BMI-associated variants [99], other approaches may shed light on this matter. The recent GWAS discovery analysis of T2D was performed with and without adjustment for BMI and for most loci there was only minimal difference in effect size estimates and significance of T2D-associated variants between these models (Fig. 7). However, at 41 of 403 distinct T2D association signals there were BMI-dependent associations – 26 variants showed attenuation of association after adjustment for BMI while 15 showed strengthening of the association after BMI adjustment

(Fig. 7) [47]. Among signals showing an adiposity-mediated effect on T2D, were *FTO*, which was actually initially identified due to this feature [71, 90], and *MC4R*. On the other hand, among variants with stronger effects in BMI-adjusted analysis were some loci with effects on insulin secretion, for instance *TCF7L2* and *JAZF1* [2, 101]. These findings indicate that only a rather small subset of genetic effect on T2D is mediated through BMI and adiposity. Another line of analysis to clarify this issue is genetic correlation analysis, which uses the full set of genome-wide association results to estimate the shared genetic contribution between two traits using methods such as linkage disequilibrium (LD)-score regression [102] often accessed through the LD hub webpage [103]. In such analyses, genetic correlation between T2D and a number of cardio-metabolic traits was seen including positive correlation with adiposity-related measures, fasting glucose, fasting insulin and fasting triglyceride [47]. The magnitude of these genetic correlations (r_g) are in the order of 20-60% indicating a significant proportion of shared genetic predisposition to these traits when estimating this based on genome-wide variation [47, 102, 103].

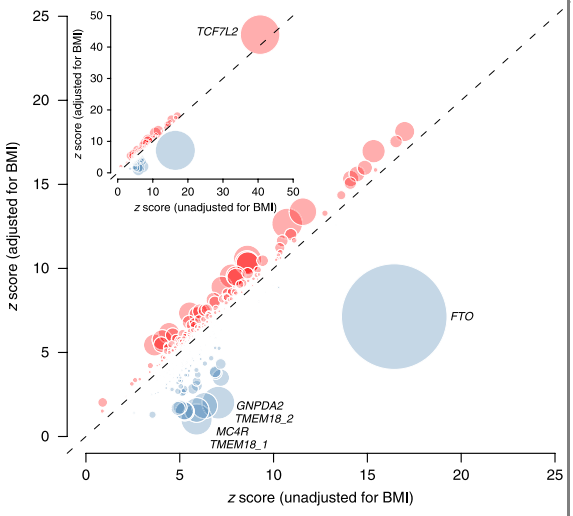

Figure 7. The effect size of T2D-associated variants on T2D with or without adjustment for BMI. Variants with a higher T2D effect size in BMI-adjusted analysis are showed in red and variants with a higher T2D effect in BMI-unadjusted analysis are shown in blue. Size of the circle is proportional to -log10 of the heterogeneity *P*-value. From ref. [47].

PART 2. THE PHYSIOLOGICAL IMPACT OF VARIANTS ASSOCIATED WITH TYPE 2 DIABETES AND GLYCAEMIA

As described above, most T2D risk variants have been detected by the agnostic GWAS approach and therefore the knowledge of the underlying biological mechanism and physiological phenotype leading to T2D or altered glucose levels is for most loci initially very limited. Subsequent studies can then be performed with the objective to unravel the intermediary phenotype of these loci in order to obtain a more detailed biological knowledge, initially seeking to break the diabetes-related phenotype for each locus into the major components of T2D pathogenesis. Genetic variants can influence risk of T2D by many different phenotypic alterations of which obesity, insulin resistance and decreased pancreatic beta cell function are the main elements. Elucidation of the intermediary mechanisms, initially in broad categories and if possible, pinpointing more distinct defects in human physiology, is important in order to understand the nature of disease mechanism but may also help to differentiate T2D in more homogenous subcategories with specific primary defects. This may influence both the specific treatment of the individual patient and potentially facilitate preventive initiatives based on genetic subgrouping of patients with T2D. In addition, these investigations will open for biological functional studies of the identified loci. Papers 1-4 of this dissertation are all studies of the endophenotypes of genetic variants found in GWAS of T2D or fasting glycaemia.

Insulin secretion deficiency as the primary genetic defect in type 2 diabetes

The project presented in paper 1 took its starting point in T2D risk variants found in the very first round of GWAS analysis, which were published in the spring and summer of 2007 [69-73]. The project investigated four of the variants initially associated with T2D in relation to T2D, insulin resistance and insulin secretion. For three loci (*HHEX*, *CDKN2A/B* and *IGF2BP2*), we replicated the impact on T2D with allelic OR of 1.10-1.30 per allele, which at this early stage of GWAS was an important proof for the validity of this novel approach. Of interest, for the variants in *HHEX* and *CDKN2A/B* loci we found associations with serum insulin levels at 30 minutes during an OGTT and with the insulinogenic index and BIGTT-acute insulin response (AIR) indices of insulin secretion in ~5,700 non-diabetic individuals of the Inter99 cohort. These findings indicate that these variants primarily increase risk of T2D due to a decreased ability of carriers to secrete the sufficient amounts of insulin in response to increases in plasma glucose concentrations. This was among the first reports to show that the GWAS-identified T2D risk variants primarily affect insulin secretion and was in agreement with other studies published at that time [104-107]. This general observation of the physiological effect of T2D risk variants was further substantiated in the second paper of the dissertation [2], which was published in 2008. In this study, we investigated the impact of six novel T2D loci identified in the first large-scale GWAS meta-analysis for T2D [50] on diabetes-related intermediary phenotypes in the Danish population. Here we found associations with indices of insulin secretion for index variants at three T2D loci in *JAZF1*, *CDC123* and *TSPAN8*, suggesting an impaired pancreatic beta cell function in risk allele carriers. Yet, the effect sizes were of a smaller scale than for the initially reported loci [1] and of borderline statistical significance. Along the same lines, papers 3 and 4 report the diabetes-related intermediary phenotypes for index variants associated with T2D or fasting glucose concentrations discovered through GWAS. Paper 3 deals with a number of variants shown to associate with fasting glucose in a GWAS of fasting glucose and fasting insulin [3]. Paper 4 presents the replication analysis of variants found to be associated with T2D in a Japanese study and here we find a strong and highly significant effect of the *VPS13C* rs7172435 variant on glucose-stimulated insulin secretion (GSIS) in individuals from the general Danish population [4].

While these findings were published in separate papers subsequent to the primary identification of the associated loci, a more general appreciation of the impact of T2D risk alleles on the two major intermediary phenotypes, insulin secretion and insulin sensitivity can be found by illustrating the effect of all T2D-associated variants on these traits. In 2010, we published a summary of the effect of the first 36 variants associated with T2D in relation to glucose-stimulated insulin release and insulin sensitivity in the

Danish Inter99 population [101], showing that the majority of T2D risk alleles have their major effect on GSIS. An updated analysis of 355 of the 403 distinct European T2D risk variants which were published recently [47] is shown in Figure 8. The figure shows the effect of the T2D predisposing allele on the Matsuda insulin sensitivity index [24] and the corrected insulin response (CIR) [25] based on analyses of 7671 non-diabetic individuals from the Danish population (Fig. 8). This analysis confirms the pattern observed in the earlier reports showing that many more T2D risk alleles are associated with decreased glucose-stimulated insulin secretion than there are alleles associated with decreased insulin sensitivity. In this analysis, 21 variants are associated with decreased GSIS at $P<0.001$, yet this sample size is insufficient to detect minor effects. However, there are some notable exceptions from this general pattern, since some insulin resistance variants have been identified. Among those loci are *PPARG*, *IRS1*, *GCKR* and *ADAMTS9* [49, 108, 109].

The published studies of some of the first GWAS-identified variants associated with T2D or fasting glucose together with updated analysis of all T2D variants shown in Figure 8, have taught us that a substantial part of the inherent genetically induced susceptibility for T2D relates to the extent to which pancreatic beta cell function can be maintained [101, 110]. This opposes the longstanding viewpoint that the majority of T2D genes would inflict a state of insulin resistance with the beta cell simply failing to respond to the lifelong state of increased insulin secretion demand [14]. Given that the T2D risk variants have been detected by the agnostic GWAS approach, the knowledge of the underlying phenotype predisposing to T2D is largely unbiased from biological hypotheses. Nevertheless, it is possible that T2D risk variants with impact on the beta cell generally have higher effect sizes, possibly due to less interaction with environmental factors, leading to increased statistical power to detect the beta cell variants both in the GWAS discovery phase and in the physiological follow-up studies. However, it is also likely that it is a genuine overall feature of the genetic predisposition to common T2D.

Of 355 distinct T2D variants investigated, 332 do not associate (at $P<0.001$) with either insulin sensitivity of insulin secretion in the analysis of 7671 individuals presented in Figure 8, which indicates that this analysis is not sufficient to capture the primary physiological impact of all T2D risk alleles. Reasons for this may be many including low statistical power in physiological follow-up studies compared to GWAS discovery studies, inability of the OGTT-based indices to capture the physiological impact of specific variants or that they work through other intermediary phenotypes. Comparisons of the effect on T2D and the effect on insulin secretion and sensitivity display that there is some correlation between these effect sizes (Fig. 9).

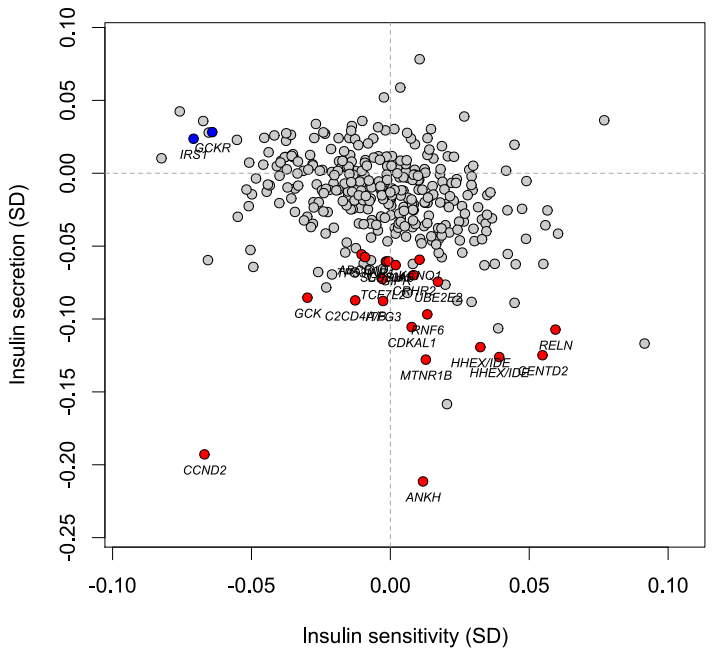

Figure 8. The effect of T2D-associated alleles on insulin sensitivity and insulin secretion. 355 T2D-associated variants [47] with MAF >2% were included. Insulin secretion and sensitivity were estimated from an OGTT by the corrected insulin response [25] and the Matsuda insulin sensitivity index [24], respectively. All effects are for the T2D-risk increasing allele. Analyses were performed in 7671 non-diabetic individuals from the Inter99, Health2008, 1936 Birth Cohort and ADDITION-Pro cohorts. Red and blue colours are shown for variants with $P<0.001$ for insulin secretion and insulin sensitivity, respectively.

This finding indicates that low statistical power may be the reason for the apparent lack of effect on intermediary phenotypes, since variants with a small impact on T2D are then expected to also have a small effect on intermediary phenotypes. Yet, other variants show a relatively high impact on T2D but no effect at all on insulin secretion or insulin sensitivity, implying that other intermediary mechanisms are at play in the connection between the genetic variant and risk of T2D. Although being at the centre of T2D pathogenesis, changes in insulin-related biology is only one of the possible intermediary mechanisms. As such, other processes related to adipose tissue biology, liver function, brain function or other biological pathway might be crucial for specific T2D risk variants [13, 111, 112].

The search for the physiological effects of genetic variants related to T2D and glycaemic traits has several limitations. The data used in these studies are a compromise between obtaining a large sample size and getting depth in the characterisation of the phenotype. Most data applied in papers 1-4 were obtained from an OGTT, based on which estimates of insulin action and GSIS were constructed. These estimates are constructed based on correlation with or modelling of more detailed physiological measures, such as intravenous glucose tolerance test (IVGTT) or hyperglycaemic clamp for insulin secretion and the hyperinsulinemic-euglycaemic clamp for insulin sensitivity [24, 113, 114]. In OGTT-based investigations, it is possible to achieve a relatively high sample size and retain a high similarity with the intended physiological parameter. Even higher sample sizes are possible if estimates of insulin response and action are based on fasting values instead of OGTT data.

Depending of the number of individuals studied, the correlations between different insulin secretion estimates from the OGTT and golden-standard clamp or IVGTT tests are in the range 0.5-0.8 [115, 116] showing that a great deal of the variation in the gold-standard index is not explained by the OGTT indices. Still, when applying OGTT-based estimates to evaluate the impact of genetic variation it is currently not possible to obtain sufficient statistical power to detect the minute effects probably inflicted by many of the T2D-associated variants. This lack of statistical power is for instance indicated by the fact that most proven variants with an impact on beta cell function were among the first identified by GWAS – the low hanging fruits – because of their relatively high effect on T2D and intermediary traits. The simplest indices are solely based on fasting concentrations of glucose and insulin and hence it is possible to obtain a higher sample size in cohorts applying such measures.

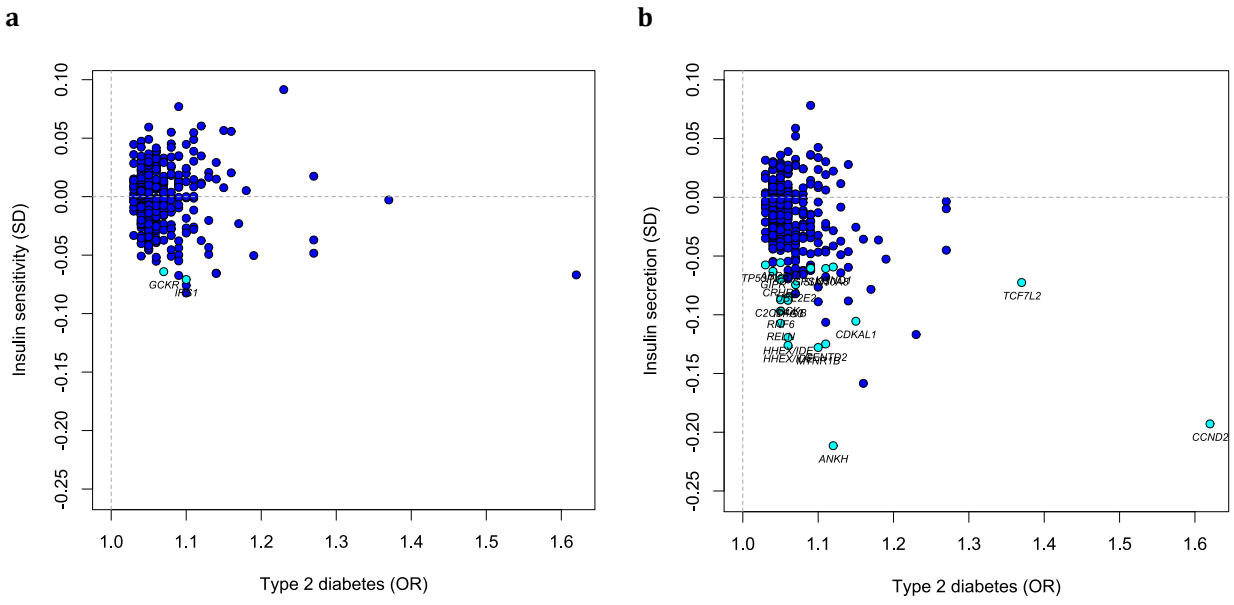

Figure 9. Relationship between the effect of T2D risk alleles on T2D and on estimates of a) insulin sensitivity and b) insulin secretion. All effects are shown for the T2D-risk increasing allele. Effects on insulin secretion and insulin sensitivity were calculated as described in Figure 8. T2D effect sizes are in OR and were obtained from ref. [47].

However, while the homeostasis model assessment (HOMA) of insulin resistance index is a rather good surrogate for hepatic insulin resistance [117], HOMA-B is not a very accurate or useful tool for estimating insulin response or beta cell function.

This limitation in statistical power of OGTT-based indices is also reflected in the rather limited success in using more refined measures of insulin secretion or insulin sensitivity in discovery analyses of genetic contributions to these phenotypes in the general population. Such studies aim at defining the genetic determinants of insulin biology. For insulin sensitivity, a GWAS applying the Stumvoll insulin sensitivity index [118, 119] found two loci, *BCL2* and *FAM19A2*, associated with variation in insulin sensitivity in individuals without diabetes [120]. This GWAS included 16,753 individuals in the discovery phase with replication studies in up to 13,354 individuals. Similarly, a GWAS of direct measures of insulin sensitivity, such as euglycemic clamp or insulin suppression test, in up to 5500 individuals without diabetes discovered a missense variant in *NAT2* with an impact on insulin sensitivity [121]. NAT2 potentially causes insulin resistance via NAT1-dependent mitochondrial dysfunction associated with increased ectopic lipid deposition [122, 123]. In 2009, a GWAS of GSIS was published finding a single novel associated variant at the *GRB10* locus after discovery analysis of 10,831 individuals [124]. Knock-down of *GRB10* in human pancreatic islet showed reduced insulin and glucagon secretion pointing to a potential but complex mechanism [124]. In addition, a GWAS of the insulin response to an IVGTT in 5567 individuals has recently been published [125]. The report evaluated genetic contributions to both peak insulin levels during an IVGTT and insulin secretion rate estimated from serum C-peptide measurements, the latter providing an estimate of the rate of insulin secretion independent of hepatic insulin clearance. While this study did not detect novel associations with GSIS, the data confirmed that decreased insulin secretion is a key mechanism for many T2D-associated loci. In addition, the study was able to compare the effect of T2D-associated variants on estimates of GSIS coming from IVGTT with that of GSIS during an OGTT. Such a comparison can point to possible effects on the incretin system if variants show a dissociative impact on GSIS after oral and intravenous stimuli, yet only serves as an indicator of such effects. For instance, the common variant inflicting the highest T2D risk in the European population at *TCF7L2* has been shown to cause incretin resistance [126-130]. Nevertheless, this was not immediately evident from these results, although the effect on peak insulin levels during IVGTT seemed to be lower than expected compared to the effect size on T2D [125]. The rather meagre findings of novel loci associated with insulin biology and T2D coming from these studies are most likely the consequence of the relatively low sample size and hence statistical power in studies of estimates of insulin sensitivity and insulin response.

Applying diverse OGTT-based estimates of beta cell function to refine physiological knowledge

As described, several different estimates of insulin secretion and insulin sensitivity exist and to a large extent the applied method is a trade of between detail and accuracy of the method and feasibility. For many purposes, the OGTT provides a reasonable compromise between these conflicting interests. Applying an OGTT makes it possible to estimate insulin sensitivity and insulin secretion by a range of different models. Since, these models have been derived in different ways they are not similar and not mutually interchangeable. Whether differences between indices also describe different biological mechanisms has not been fully elucidated. However, it is plausible that for instance an index modulated after IVGTT data is not reflecting the same insulin biology as an index modelled after other insulin response tests or simply being based on the incremental relative increase in serum insulin in relation to the increase in plasma glucose during the first part of an OGTT. Studies of the phenotypic and genetic correlations of various GSIS indices show that the level of shared genetic background varies between surrogate measures of insulin release. Although indices share genetic determinants to a large degree, a subset of variation is explained by non-shared genetic factors [131]. The CIR and insulinogenic index shared the majority of their genetic backgrounds, with genetic correlations of 0.80–0.99, while the BIGTT-AIR, which is modelled after an IVGTT [116], differed slightly more from the latter with genetic correlations of 0.78–0.87. Hence, there may be different genetic determinants of these indices and difference in associations with various indices for risk variants may point to more specific processes in insulin release. Figure 10 shows the normalised effects of T2D risk alleles on the two different OGTT-based GSIS indices – BIGTT-AIR and CIR based on the analyses in 7671 individuals.

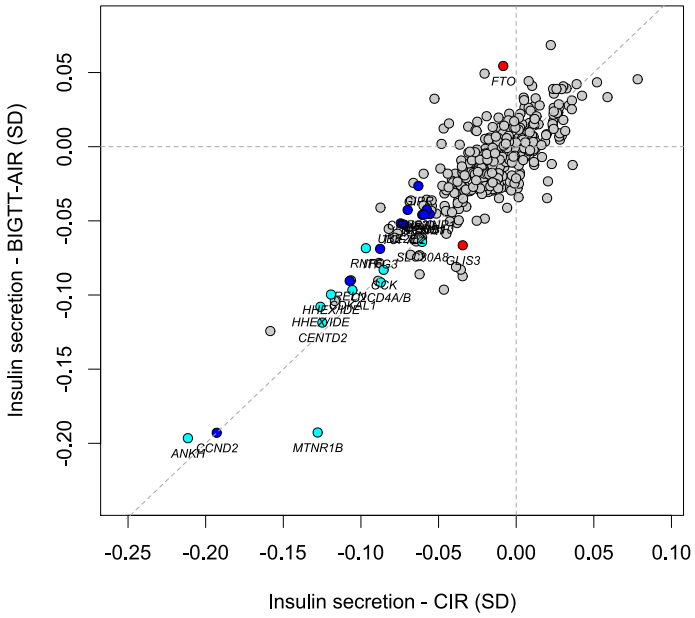

Figure 10. Relationship between the effect of T2D risk alleles on two different indices of glucose-stimulated insulin response. 355 T2D-associated variants [47] with MAF >2% were included. Insulin secretion was estimated from an OGTT by the corrected insulin response (CIR) [25] and the BIGTT-acute insulin response (BIGTT-AIR) indices [116]. All effects are for the T2D-risk increasing allele. Analyses were performed in 7671 non-diabetic individuals from the Inter99, Health2008, 1936 Birth Cohort and ADDITION-Pro cohorts. Red and blue colours are shown for variants with *P*<0.001 for CIR and BIGTT-AIR, respectively, while light blue designates *P*<0.001 for both traits.

It is evident that most T2D risk alleles fall on the diagonal line as expected in case of an equally strong impact on the two different insulin secretion estimates. However, for instance the rs10830963 variant at the *MTNR1B* locus deviates somewhat from the line displaying a stronger impact on the BIGTT-AIR index. Along the same lines, this variant showed a stronger effect on GSIS after an IVGTT than after an OGTT in data from Wood et al. [125] and it shows a higher than expected effect on BIGTT-AIR given its effect on T2D (Fig. 9 & 10) [132]. The *MTNR1B* locus was originally discovered via GWAS of fasting glucose in people without diabetes and was subsequently shown to be associated with T2D [133-135] and with GSIS estimated from OGTT data [133, 136]. Therefore, these studies have implied that the deleterious effect of the *MTNR1B* locus on T2D risk is likely to be due to a dysfunction of pancreatic islets and beta cells. *MTNR1B* encodes one of the two receptors of melatonin, a neurohormone involved in circadian rhythms. However, a conflict between the biological mechanism and direction of effect exists since reports have both shown that rare loss-of-function variants increase T2D risk [137] and that the common T2D risk allele of rs10830963, which is likely the causal variant at the locus [47, 138], correlates with higher *MTNR1B* expression in human pancreatic islets [134] and increased FOXA2-bound enhancer activity in human islets [138]. Whether decreased or increased melatonin signalling is the pathway to increased risk of T2D thus remains to be clarified and is currently intensely discussed [139, 140].

A cluster-based view on physiological impact of type 2 diabetes-associated variants

An alternative approach to follow-up on T2D variants with estimates of proinsulin processing, insulin secretion and insulin sensitivity was published in 2014. This was one of the largest effort investigating these traits thus far collecting association results for 37 T2D-associated variants [75] in up to 17,237 individuals with dynamic measures and 58,000 participants with fasting samples [141]. Based on these association results, variants were clustered into groups with similar association patterns and hence similar predicted physiological mechanism. Five major groups of loci were formed depicting insulin sensitivity loci (*PPARG*, *KLF14*, *IRS1* and *GCKR*), reduced insulin secretion and fasting hyperglycaemia (*MTNR1B* and *GCK*), defects in insulin processing (*CARAP1*) and defects in insulin processing and secretion without a strong effect on fasting glycaemia (*TCF7L2*, *SLC30A8*, *HHEX*, *CDKAL1* and *CDKN2A*/B). Finally, 20 risk loci showed no clear grouping or associations [141]. Highly comparable results were found when analysing extended sets of T2D-associated variants [76, 142]. In these analyses, a relatively high fraction of the associated loci could not be assigned to a specific cluster. Recently, another study performed soft clustering of 94 T2D-associated variants with 47 T2D-related traits and identified five distinct clusters of variants [143]. These five clusters were 1) beta cell loci,

2) loci with an impact on proinsulin, 3) loci associated with obesity, 4) loci associated with lipodystrophy-like insulin resistance and 5) loci associated with liver and lipids. Of these, clusters 1 and 4 were the largest containing 30 and 20 loci, respectively [143]. The finding of a cluster of variants associated with lipodystrophy-like insulin resistance is supported by previous reports [144, 145] and supports the notion that limited storage capacity of peripheral adipose tissue is an important etiological component in insulin-resistant cardio-metabolic disease. These studies provide important information on the major mechanistic pathways from genetic susceptibility to T2D.

Refined endophenotypes for type 2 diabetes

Papers 1-4 seek to shed light on a number of loci primarily found to be associated with T2D or levels of fasting glucose. Most of these studies were done in samples with OGTT and therefore primarily evaluated GSIS based on an oral stimulus. Furthermore, it is clear that solely analysing the crude insulin response and insulin action estimates will not bring knowledge on the more detailed processes involved the relationship between genetic risk variants and T2D pathogenesis. A more detailed estimation trying to disentangle processes related insulin synthesis, secretion and processing is needed to get closer to the biological mechanisms. However, studies of these more detailed phenotypes are generally small in size and therefore with little statistical power to pinpoint the modest effects inflicted by common genetic variants in humans. For instance, a study from 2010 applied a 2-hr hyperglycaemic clamp and in a subset a test of response to GLP1 and arginine during an extended clamp. While the investigations were very thorough, the number of individuals studied was rather limited ranging from 123 to 336 depending on the test [146]. The study found associations with diabetes risk alleles at *CDC123*, *THADA*, *ADAMTS9*, *BCL11A* and *MTNR1B* with various specific aspects of beta cell function; however, the small sample size deterred any strong associations and hence weakened the conclusions.

As mentioned, measurement of incretin hormones can reveal an impaired incretin effect, as shown for the *TCF7L2* risk allele (Table 1). Similarly, measurement of serum C-peptide levels instead of serum insulin and establishment of secretion indices based on this can possibly delineate effects, which are otherwise hidden due to the relative instability of the insulin assay and due to variability in the first-pass effect by the liver. Another way to investigate the effect of the first-pass effect by the liver is to estimate insulin clearance. As for insulin secretion and sensitivity, this can be done by surrogate measures, mostly the fasting C-peptide to fasting insulin ratio, based on the assumption that C-peptide is secreted from the beta cell in equimolar amounts with insulin, but is not subjected to first-pass metabolism by the liver. Alternatively, insulin clearance can be estimated by more refined clamp techniques [147]. A paper published in 2012 showed that several T2D-associated loci (*CDKAL1*, *DGKB*, *JAZF1*, *GLIS3*, *FADS1* and *IGF1*) have an effect on insulin clearance as estimated from hyperinsulinaemic–euglycaemic clamp procedure in a sample of ~1300 individuals [148]. Possibly differences in insulin clearance may compromise the insulin response estimates obtained from an OGTT, since hepatic insulin clearance by the liver removes around 50% of insulin at first pass [149]. At the same time, it is likely that at least some T2D risk alleles have an impact on insulin clearance as part of the diabetes susceptibility mechanisms.

Another related measurement is proinsulin, which is a precursor of mature insulin and C-peptide. Higher circulating levels of proinsulin are indicative of impaired beta cell function, beta cell stress or abnormalities in insulin processing or secretion [150]. A number of large-scale studies have applied proinsulin measurement to differentiate the genetic impact on different aspects of insulin biology. A GWAS from 2011 investigated 10,701 individuals without diabetes in the discovery phase with replication in further 16,378 individuals and found a number of genetic loci associated with circulating proinsulin levels [151]. Interestingly, some of these were T2D risk loci known at that time or which have been established since (*TCF7L2*, *SLC30A8*, *MADD*, *ARAP1* and *C2CD4A)*, at which the T2D risk alleles of all but *ARAP1* were associated with increased proinsulin levels [151]. Together with other studies of the relationship between T2D risk variants and circulating levels of proinsulin, this study indicates that carriers of a number of T2D risk variants have impairments in the distal insulin processing and secretion pathways.

Table 1. Summary of association with endophenotypes for 108 GWAS-identified variants associated with T2D.

Locus name	Variant	Chr	Position	RAF	OR	Endophenotype	References
MACF1	rs3768321	1	40035928	0.20	1.09	-	
FAF1	rs58432198	1	51256091	0.88	1.07	-	
NOTCH2	rs1493694	1	120526982	0.11	1.09	-	
FAM63A	rs145904381	1	151017991	0.99	1.19	-	
PROX1	rs340874	1	214159256	0.56	1.07	Reduced GSIS during OGTT and IVGTT	[3, 141, 152]
TMEM18	rs62107261	2	422144	0.95	1.12	Obesity	[153]
GCKR	rs1260326	2	27730940	0.61	1.07	Insulin resistance	[87, 141, 154]
THADA	rs80147536	2	43698028	0.90	1.13	Reduced beta cell function	[141, 146]
BCL11A	rs243024	2	60583665	0.46	1.06	-	
CEP68	rs2249105	2	65287896	0.63	1.10	-	
RBMS1	rs3772071	2	161135544	0.71	1.05	-	
GRB14/ COBLL1	rs10195252	2	165513091	0.59	1.07	Lipodystrophy-like insulin resistance	[87, 145]
IRS1	rs2972144	2	227101411	0.64	1.10	Lipodystrophy-like insulin resistance	[108, 141, 145, 155]
PPARG	rs11709077	3	12336507	0.88	1.14	Lipodystrophy-like insulin resistance	[48, 145]
UBE2E2	rs35352848	3	23455582	0.79	1.07	Reduced GSIS during OGTT	[156]
KIF9	rs11926707	3	46925539	0.63	1.27	-	
PSMD6	rs3774723	3	63962339	0.84	1.07	-	
ADAMTS9	rs9860730	3	64701146	0.70	1.06	Insulin resistance	[109]
ADCY5	rs11708067	3	123065778	0.77	1.09	Reduced GSIS during IVGTT, beta cell dysfunction	[125, 141]
IGF2BP2	rs6780171	3	185503456	0.31	1.14	Reduced GSIS during OGTT and IVGTT	[1, 125]
ST6GAL1	rs3887925	3	186665645	0.55	1.07	-	
LPP	rs4686471	3	187740899	0.61	1.06	-	
MAEA	rs56337234	4	1784403	0.50	1.06	-	
WFS1	rs10937721	4	6306763	0.59	1.06	Reduced GSIS during OGTT	[157, 158]
FAM13A	rs1903002	4	89740894	0.50	1.04	Lipodystrophy-like type insulin resistance	[145]
TMEM154	rs7669833	4	153513369	0.70	1.06	Reduced GSIS during OGTT	[159]
ACSL1	rs58730668	4	185717759	0.86	1.07	-	
ANKH	rs146886108	5	14751305	0.99	1.41	Reduced GSIS during OGTT	
ARL15	rs702634	5	53271420	0.69	1.05	Lipodystrophy-like insulin resistance, adiponectin	[145, 160, 161]
ANKRD55	rs465002	5	55808475	0.74	1.11	Lipodystrophy-like insulin resistance	[145]
POC5	rs2307111	5	75003678	0.60	1.05	-	
ZBED3	rs4457053	5	76424949	0.30	1.06	-	
PAM	rs115505614	5	102422968	0.05	1.19	Reduced GSIS during OGTT	[162]
JADE2	rs329122	5	133864599	0.43	1.04	-	
RREB1	rs9379084	6	7231843	0.89	1.11	-	
CDKAL1	rs7756992	6	20679709	0.27	1.15	Reduced GSIS during OGTT and IVGTT, insulin clearance	[70, 104, 124, 125, 141, 148]
MHC	rs601945	6	32573415	0.18	1.06	-	
VEGFA	rs6458354	6	43814190	0.29	1.05	-	
TFAP2B	rs3798519	6	50788778	0.18	1.06	-	
CENPW	rs11759026	6	126792095	0.23	1.07	-	
SLC35D3	rs9494624	6	137300960	0.29	1.04	-	
SLC22A3	rs474513	6	160770312	0.52	1.04	-	
DGKB	rs10228066	7	15063569	0.54	1.07	Reduced GSIS during OGTT, insulin clearance	[3, 141, 148]
JAZF1	rs1708302	7	28198677	0.51	1.10	Reduced GSIS during OGTT, insulin clearance	[2, 148]
GCK	rs878521	7	44255643	0.24	1.06	Reduced GSIS during OGTT	[124, 163]
KLF14	rs1562396	7	130457914	0.32	1.06	Insulin resistance	[141]
MNX1	rs6459733	7	156930550	0.67	1.06	-	
LPL	rs10096633	8	19830921	0.88	1.07	Lipodystrophy-like insulin resistance	[144]
ANK1	rs13262861	8	41508577	0.83	1.07	Reduced GSIS during OGTT	[124, 164]
TP53INP1	rs10097617	8	95961626	0.48	1.04	-	
SLC30A8	rs3802177	8	118185025	0.68	1.11	Reduced GSIS during OGTT and IVGTT, insulin processing	[105, 107, 125, 141, 151]
GLIS3	rs10974438	9	4291928	0.36	1.05	Reduced GSIS during OGTT insulin clearance	[3, 148]
CDKN2A/B	rs10811660	9	22134068	0.83	1.27	Reduced GSIS during OGTT and IVGTT	[1, 125, 141]
TLE4	rs17791513	9	81905590	0.93	1.10	-	
TLE1	rs2796441	9	84308948	0.59	1.07	-	

Locus name	Variant	Chr	Position	RAF	OR	Endophenotype	References
ABO	rs505922	9	136149229	0.33	1.05	-	
GPSM1	rs28505901	9	139241030	0.75	1.09	-	
CDC123	rs11257655	10	12307894	0.22	1.09	Reduced GSIS during OGTT	[2, 146]
ZMIZ1	rs703972	10	80952826	0.53	1.07	-	
HHEX/IDE	rs10882101	10	94462427	0.59	1.06	Reduced GSIS during OGTT	[1, 104, 124, 141]
TCF7L2	rs7903146	10	114758349	0.30	1.37	Reduced GSIS during OGTT and IVGTT, incretin dysfunction	[125, 127, 129, 130, 141, 151, 165]
PLEKHA1	rs2280141	10	124193181	0.52	1.05	-	
INS/IGF2	rs4929965	11	2197286	0.38	1.07	-	
KCNQ1	rs2237895	11	2857194	0.43	1.12	Reduced GSIS during OGTT and IVGTT	[125, 166, 167]
KCNJ11	rs5213	11	17408404	0.36	1.07	Reduced GSIS during OGTT	[168]
HSD17B12	rs1061810	11	43877934	0.29	1.05	-	
MAP3K11	rs1783541	11	65294799	0.20	1.06	-	
ARAP1	rs77464186	11	72460398	0.84	1.11	Reduced GSIS during OGTT and IVGTT, insulin processing	[125, 141, 151, 169]
MTNR1B	rs10830963	11	92708710	0.28	1.10	Reduced GSIS during OGTT and IVGTT	[124, 125, 136, 141]
CCND2	rs76895963	12	4384844	0.98	1.62	Reduced GSIS during OGTT	[170, 171]
KLHL42	rs10842994	12	27965150	0.80	1.08	-	
HMGA2	rs2258238	12	66221060	0.10	1.10	-	
TSPAN8	rs1796330	12	71522953	0.57	1.05	Reduced GSIS during OGTT	[2]
WSCD2	rs1426371	12	108629780	0.74	1.05		
HNF1A	rs56348580	12	121432117	0.69	1.05	Reduced GSIS during IVGTT	[125]
MPHOSPH9	rs4148856	12	123450765	0.78	1.05		
RNF6	rs34584161	13	26776999	0.76	1.05		
KL	rs576674	13	33554302	0.17	1.05	-	
SPRY2	rs1359790	13	80717156	0.72	1.09	-	
NRXN3	rs17836088	14	79932041	0.22	1.06	-	
RASGRP1	rs34715063	15	38873115	0.12	1.10	-	
C2CD4A/B	rs8037894	15	62394264	0.57	1.05	Reduced GSIS during OGTT and IVGTT, insulin processing	[3, 4, 124, 125, 151]
PTPN9	rs13737	15	75932129	0.76	1.05	-	
HMG20A	rs1005752	15	77818128	0.72	1.08	-	
AP3S2	rs4932265	15	90423293	0.27	1.07	-	
PRC1	rs12910825	15	91511260	0.36	1.05		
FAM234A	rs6600191	16	295795	0.82	1.06	-	
FTO	rs1421085	16	53800954	0.42	1.13	Obesity	[90]
NFAT5	rs862320	16	69651866	0.58	1.04	-	
BCAR1	rs72802342	16	75234872	0.92	1.17	-	
CMIP	rs2925979	16	81534790	0.30	1.05	-	
ZZEF1	rs1377807	17	4045440	0.31	1.05	-	
GLP2R	rs7222481	17	9785187	0.32	1.04	-	
HNF1B	rs10908278	17	36099952	0.48	1.08	-	
MLX	rs34855406	17	40731411	0.28	1.05	-	
TTLL6	rs35895680	17	47060322	0.68	1.06	-	
BPTF	rs61676547	17	65892507	0.19	1.06	-	
LAMA1	rs7240767	18	7070642	0.38	1.04		
MC4R	rs523288	18	57848369	0.24	1.05	Obesity	[91]
BCL2	rs12454712	18	60845884	0.61	1.05	Insulin resistance	[120]
TM6SF2	rs8107974	19	19388500	0.077	1.10	-	
PEPD	rs10406327	19	33890838	0.52	1.04	Lipodystrophy-like insulin resistance	[145]
APOE	rs429358	19	45411941	0.85	1.08	-	
GIPR	rs10406431	19	46157019	0.56	1.05	Reduced GSIS during OGTT	[124]
HNF4A	rs1800961	20	43042364	0.035	1.18	-	
MTMR3	rs6518681	22	30609554	0.91	1.09	-	
PNPLA3	rs738408	22	44324730	0.23	1.05	Liver function	[172, 173]
PIM3	rs1801645	22	50356850	0.28	1.04	-	

T2D associated variants were obtained from Scott et al. [76]. Effect sizes in OR on T2D and risk allele frequencies (RAF) were obtained from Mahajan et al. [47].

Besides biologically mechanisms strictly related to glucose homeostasis such as described in the above sections, many other T2D-related pathophysiological mechanisms could be behind the association of genetic variants and T2D. Some risk variants increase risk of T2D by increasing risk of obesity while quite a lot have unknown intermediary mechanisms (Table 1) and any other T2D-defining mechanisms could be in play for these variants. For instance, the T2D-associated p.Ile148Met variant at *PNPLA3* is associated with markedly increased risk of progressing into the entire spectrum of non-alcoholic fatty liver disease [172]. The mechanisms underlying this strong effect seem to be related to direct changes in hepatocyte and hepatic stellate cells lipid droplet biology [174, 175]. How this variant leads to an increased risk of T2D is unknown but it does not seem to be via insulin resistance, however, the variant paradoxically associates with lower risk of coronary artery disease [176].

Biology of loci associated with beta cell dysfunction

The identification of intermediary endophenotypes for T2D-associated loci represents an important step forward in the disclosure of the specific nature of the relationship between genetic variation and the increased risk of T2D. However, a lot of work remains to map the biological details in these relationships in biological terms. For some of the loci I have studied in papers 1-4, very little progress has been made in this regard, while for others studies have offered insights into the more specific mechanisms behind these associations.

The *HHEX*/*IDE* locus on chromosome 10 was one of the first GWAS findings of variants associated with T2D [69] and is among the T2D risk alleles with the strongest impact on GSIS [1, 104, 141]. For this locus, biological knowledge is relatively limited. It is currently not known, which of the genes in the region is the causal gene. The intergenic rs1111875 variant was the initially identified lead variant, however the latest GWAS of T2D meta-analysis pinpointed three independent association signals in the region with intergenic rs10882101 being the primary lead variant [47]. Fine-mapping analysis to identify the causal variant of this association signal has led to a rather narrow 22 kb region likely to house the causal variant. This region spans *HHEX* and in the densely imputed data from Mahajan et al. [47], this region contains only 10 variants associated with T2D, however, whether the causal variant has its effect on *HHEX* or possibly *IDE* is currently unknown. While *HHEX* is the closest gene, *IDE* is a reasonable candidate gene due to its role in insulin biology. Insulin-degrading enzyme (IDE) is a ubiquitous peptidase, which was initially discovered as the enzyme responsible for insulin catabolism. However, it also has the ability to degrade several other polypeptides, such as beta-amyloid, amylin, and glucagon [177]. There is increasing evidence that improper IDE function, regulation or trafficking might contribute to the aetiology of metabolic diseases and IDE inhibitors have been evaluated in animal models for their potential to treat T2D with some positive indications of this being a successful future treatment option [178, 179]. A study of human islets showed that islets carrying the rs1111875 T2D risk allele had significantly decreased number of docked insulin granules and a tendency to reduced insulin exocytosis [180]. Although this is of interest, the mechanism for the effects on beta cell biology, GSIS and T2D observed for variants in this locus remains largely unexplained 10 years after the initial discovery.

Another interesting locus, which was identified in some of the first GWAS of T2D, is *CDKN2A/B* [72]. Here rs10811660 was associated with increased risk of T2D and subsequently with decreased GSIS during OGTT [181] and IVGTT [125]. In the most recent GWAS of T2D, rs10811660 with a relatively high OR of 1.27, and five additional distinct signals were found in the locus [47]. For rs10811660, the 99% credible set only spans 1.5 kb and contains five variants, making it likely that rs10811660 is causal. Moreover, a variant in the locus has been shown to associate with coronary artery disease and myocardial infarction [74, 182, 183], however, this variant and the T2D-associated rs10811661 variant are not correlated (r^2<0.01). The *CDKN2A* and *CDKN2B* genes encode $p16^{INK4A}$ and $p15^{INK4B}$, respectively, which regulate proliferation, oncogenesis, senescence and ageing [184]. Both *CDKN2A* and *CDKN2B* are expressed in pancreatic beta cells [185] and are implicated in pancreatic islet regenerative capacity [186, 187]. Of interest, human carriers of rare *CDKN2A* loss-of-function mutations, which are a cause of familial melanoma, displayed increased insulin secretion, impaired insulin sensitivity, and reduced hepatic insulin clearance compared to non-carriers [188], which together with functional data [186, 188], points to *CDKN2A* being the T2D-related effector transcript of the region. Functional genomics data from the recent GWAS of T2D indicated

that regulatory mechanisms in human islet link the genetic variation with beta cell function [47].

Of interest, other of the variants and loci investigated in papers 1-4 also map to regulatory sequence in human islets indicating similar genomic mechanisms as for *CDKN2B*. This is the case for the *CDC123*, *IGF2BP2*, *JAZF1* and *GLIS3* loci [47, 138, 189, 190] indicating that T2D-risk variants are disproportionately located in DNA sequences involved in the regulation of islet-gene transcription, further highlighting the central role of the regulation of insulin secretion in genetic risk of T2D.

Future directions within this field

It remains an important objective to seek to elucidate the intermediary pathophysiological processes influenced by individual T2D risk variants, as this can be seen as a first but important step in finding the biological and physiological relationship between genetic variants and risk of T2D. However, the modest effect size imposed by most GWAS-identified variants found in large-scale studies impedes the statistical power to find intermediary effects since sample sizes for studies of more detailed phenotypes are generally much lower than for case-control studies of T2D. Larger studies — by incorporating more meta-analysis of cohort data and by establishment of larger sample sets — are needed to overcome this challenge. However, the resources needed to build such resources, which are severely larger than making large biobanks holding fasting or random samples, will always limit these efforts. For instance, UK Biobank is a major current and future resource for epidemiological and genetic-epidemiological research and does not include any OGTT data or other detailed physiological data relevant for T2D. Theoretically, it may be possible to substitute quantitative estimates of endophenotypes such as insulin sensitivity and insulin secretion by surrogate biomarkers from biomarkers from serum or urine metabolomics or proteomics data and by this enable application in a higher number of individuals and large-scale analysis.

As will be discussed in Part 5, the implementation of genetic data in the improvement of precision medicine may be relying on quantitative measures and the genetic determinants of quantitative T2D-related measures to ensure success. A reason for this hypothesis is that risk of T2D-related complications seems to be inflicted by a graded, linear increase in physiological variables, and not to be driven by thresholds arbitrarily defining categorical disease entities. For instance, this may well be the case for T2D-related risk of CVD, for which studies have shown a continuous or U-shaped relationship between fasting and post-OGTT glucose levels and risk of CVD, which is apparent below thresholds for diagnosis of diabetes [191-193]. Therefore, a molecular understanding of quantitative diabetes-related measures may well prove to be crucial for finding causes and individualised preventive approaches for important clinical outcomes.

PART 3. MASSIVE NUCLEOTIDE SEQUENCING IN SEARCH FOR RARE SUSCEPTIBILITY ALLELES

The design of GWAS has inherent limitations. First, by GWAS, it is only possible to investigate known variation and at least in the first number of years of GWAS this was a major limitation. However, after completion of larger sequenced project, such as the 1000 Genomes Project [194] rather exhaustive scaffolds for imputation exist. Second, genome-wide array based genotyping only covers common variants with MAF above 5%, although that newer and larger imputation reference panels have pushed the lower boundary of genotype imputation to enable coverage of low-frequency variants with MAF between 0.5% and 5%, and possibly even to a MAF of 0.1%, depending on the specific population group, the genotyping array and the imputation panel chosen [63] (Fig. 3). However, for rare variants (MAF <0.5%) a simple extension of the GWAS paradigm is probably not sufficient since genotype imputation is inaccurate and since single variant tests are underpowered for the detection of such variants. To circumvent the lack of statistical power of single marker tests several collapsing or burden methods that simultaneously analyse multiple rare variants are applied [195-198]. Because of the limitations of GWAS, especially in the early days with relatively sparse imputation panels available, a number of studies have aimed to include sequencing data in genetic association studies to discover novel variants and to include rare variants with MAF below 0.5% in the evaluation, although the costs of such studies are markedly higher than for genotype-based studies.

Rare variants in common disease?

As described, one of the ambitions of performing sequencing in complex disease is to evaluate the role of rare variants in common complex disease. While it is well accepted and empirically proven that common variants are important in common disease [199], the role of rare variants has been discussed for years [200]. The rare variant hypothesis suggests that rare alleles with relatively high penetrance are the primary drivers of common disease [201, 202], and this hypothesis has received renewed attention in the last few years. Rare variants are common in the sense that they severely outnumber common variants in the human genome [194, 203, 204]. Furthermore, evolutionary theory predicts that disease-causing alleles should be rare since even a minute reduction in fitness will keep allele frequencies low due to negative selection [205]. These variants are predicted to have stronger effects on disease (allelic OR>2) [198, 206], underlying arguments for their study as an addition to GWAS [204, 207]. While these arguments have previously been theoretical, we have in the past decade moved to a stage where we can empirically and systematically investigate this hypothesis. Deep whole genome or exome sequencing in large sample sizes allow for the interrogation of the full spectrum of genetic variation, including the many variants, which are too rare to be accurately studied through current GWAS and imputation strategies. This has been possible trough advances in sequencing technologies over the past decade leading to higher performance and lower costs [208]. The results are of more than academic interest, as an understanding of the genetic architecture of a disease is a crucial and powerfully tool to achieve knowledge for the application of precision medicine.

Exome-sequencing based discoveries in metabolic traits

The ultimate sequencing-based design is to apply deep whole-genome sequencing but it also carries the highest costs and the biggest bioinformatic workload. An alternative approach is to perform exome sequencing [209], which lowers the sequencing burden to ~1-2% of whole genome sequencing and hence reduces costs. Exome sequencing would seem as a logical first conquest, relying on the hypothesis that functional disease-associated variation resides in the coding regions of the genome. Strong-effect coding variants may be molecular guides into the pathological relevance of a gene and potentially establish a direct causal link between gene gain- or loss-of-function and disease risk [142, 210, 211]. This can especially be the case when there is evidence of multiple independent variant associations (an “allelic series”) within a gene [210-212]. Exome sequencing has proven valuable in the search for mutations responsible for Mendelian diseases [213, 214].

While both whole-genome and exome sequencing are now routinely done on a large scale in thousands of samples, this was not the case when we initiated a large-scale exome sequencing project in 2008. We established the *Lundbeck Foundation Centre for Applied Medical Genomics in Personalised Disease Prediction, Prevention and Care* and initiated an exome sequencing study of 1000 T2D patients and 1000 controls without T2D at a sequencing depth of 8× with the initial aim to investigate the role of rare variants with high impact on risk of T2D. The main results are included as paper 5 of the dissertation. At the point in time when the project was initiated, sequencing technology was still not at a stage in which it was possible to investigate 2000 samples efficiently and the sequencing in the project took more than 1½ years on Illumina Genome Analyzer II technology at BGI, Shenzhen, China. In 2000 individuals, we detected 70,182 variants with a frequency above 1%. Association analyses of metabolic traits were performed in a three-stage design (Fig. 11).

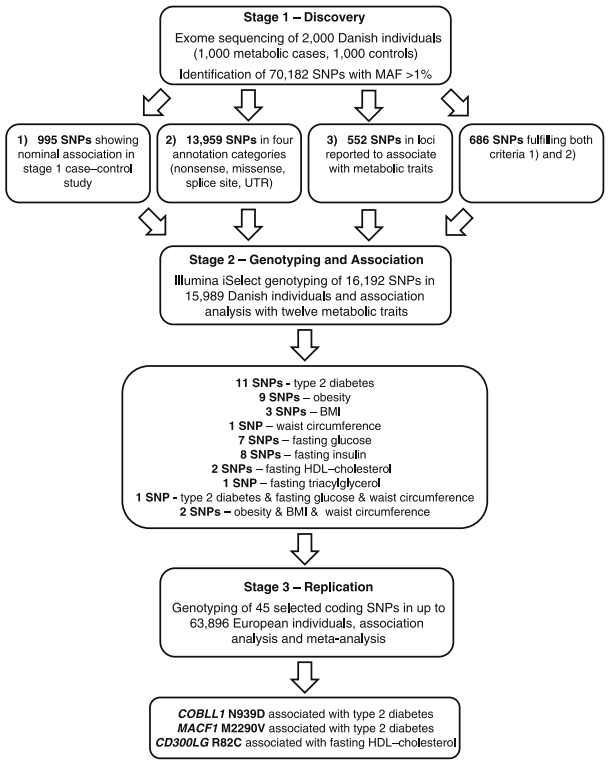

Figure 11. Flow chart of the discovery and replication phases of the LuCamp project. From paper 5.

In stage 1, we performed association analysis in the exome sequenced individuals, which did not reveal statistically significant associations. In stage 2, we genotyped 16,192 selected coding variants in ~16,000 Danish individuals and performed association testing with 12 metabolic traits. Based on the association results, we found 45 SNPs, which we selected for replication in more than 60,000 independent European samples. Finally, we found three missense variants in *MACF1* p.Met2290Val, *COBLL1* p.Asn939Asp and *CD300LG* p.Arg82Cys, which were associated with a metabolic trait. Two variants were associated with T2D, while *CD300LG* p.Arg82Cys was associated with circulating levels of high-density lipoprotein (HDL)-cholesterol and triglyceride. Although the strength of sequencing-based association studies in general would be to detect low-frequency and rare variants, two of the three identified variants were common, while *CD300LG* p.Arg82Cys had a MAF of 3.5%. All three associations have since been replicated in independent data [142, 176, 197, 215-217].

The association of the *CD300LG* p. Arg82Cys variant with decreased levels of serum HDL-cholesterol and increased serum triglycerides in the general population represents a true sequencing-based finding of that era and has since been replicated in studies applying the exome chip [176, 197, 215] or sequencing-based imputation of GWAS data [217]. While one of the promises of exome sequencing has been postulated to be a greater ability to pinpoint biological mechanism from association of coding variants, the relationship between *CD300LG* p.Arg82Cys and lipid metabolism has not been elucidated. The protein encoded by *CD300LG* belongs to the CD300 family of membrane-bound molecules, which have broad and diverse immunological actions, including the ability to recognise and interact with extracellular lipids [218-220]. The CD300LG protein is expressed in a broad range of tissues with highest expression in the placenta, adipose tissue, and skeletal muscle [5, 221]. The biological functions of CD300LG in adipose tissue and skeletal muscle are largely unknown.

Variants in the *COBLL1/GRB14* region had already previously been associated with fasting insulin levels [87, 88] and the impact on T2D was simultaneously with our discovery found in a classical GWAS design [53] and was later also found in South Asians individuals [222]. Furthermore, variants in the locus partly correlated with p.Asn939Asp are also associated with waist-hip ratio [95, 223], HDL-cholesterol and triglyceride concentrations [224]. Recent papers have pointed to that this locus is likely to influence T2D risk through a decreased capacity for fat storage in peripheral adipose tissue [144, 145]. Subsequently, it

has been discussed what the causal element is in this locus. *GRB14* is an attractive biological candidate gene for T2D since studies of *Grb14*-deficient mice show improved glucose homeostasis despite lower circulating insulin levels and enhanced insulin signalling in liver and skeletal muscle [225]. The latest statistical evidence, derived from an exome chip-based analysis of more than 450,000 individuals shows that the association signal of this locus is likely not driven by the *COBLL1* p.Asn939Asp variant but by non-coding variants at the locus (Fig. 12, page 31), yet the causal variant and biological mechanism have not been determined [142].

The T2D-associated variant in *MACF1* is common (MAF 23.4%) and had probably not been detected previously due to the incomplete genomic coverage of HapMap-based imputation of GWAS of the time. *MACF1* is a very large gene spanning 400 kb, consisting of a large number of exons and is expressed in several different isoforms. Recent studies have shown that additionally four coding missense variants (p.Ile39Val, p.Lys1625Asn, p.Met1424Val, p.Ala3354Thr) associate with T2D [142]. All these variants are common and correlated with each other and thus far, these association signals have been statistically undistinguishable. The recent paper in which these four variants were discovered also made an analysis to estimate the probability that coding variants were driving the observed association signal as compared to non-coding variants. For the *MACF1* locus, the overall probability attributed to coding variants for association with T2D for this locus is compatible with a partial role for coding variants [142] (Fig. 12). *MACF1* encodes a protein that forms bridges between different cytoskeletal elements, by stabilising and guiding microtubule growth along actin filaments. Loss-of-function studies using knockout mouse models have shown pivotal roles of MACF1 in embryonic development, skin integrity maintenance, neural development and bone formation [226]. As such, there is no obvious biological connection between MACF1 and T2D.

Sequencing studies of type 2 diabetes and related phenotypes

The described exome sequencing study published in paper 5 was an early attempt to implement sequenced-based association studies but suffered from a number of shortcomings. First, the sequencing was rather shallow at 8× depth leading to problems in accurately calling rare variants in individual samples. Furthermore, due to immature sequencing technology, the sequencing process was performed over a long period and shifts and developments in the assays introduced bias and difference in sequencing depth between cases and controls making association analysis difficult and forcing us to focus on variants with a MAF above 1% thereby removing the analysis of rare variants [5]. As such, this study can be seen as an initial steppingstone on the path to larger and more detailed studies.

Other early sequencing studies with a focus on rare variants and the risk of T2D have focussed on single candidate genes selected based on prior association between common variants and T2D in GWAS. A study of *MTNR1B* – a locus at which common variants in 2009 were found to be associated with fasting glucose and T2D [133, 135] – performed sequencing of 7632 individuals and demonstrated a burden of rare variants associated with increased risk of T2D. Performing functional studies of all variants and grouping them based on functionality increased the effect on T2D to an odds ratio of 5.7 [137]. Along the same lines, a study of *SLC30A8*, a locus at which the common coding T2D-associated p.Trp325Arg variant was discovered in GWAS [69], performed sequencing or genotyping in 150,000 individuals and discovered a burden of putative loss-of-function variants with a protective effect on T2D [211]. This study suggests that inhibition of ZnT8 may be a therapeutic strategy for T2D. As described in Part 2, the common p.Arg325Trp variant in *SLC30A8* is associated with T2D and with decreased GSIS [69, 107] and with increased proinsulin to insulin ratio [227], which is in line with studies of mice [228]. *SLC30A8* encodes a zinc transporter, ZnT8, which is expressed in the endocrine pancreas and has an emerging role in glucose homeostasis due to the requirement for zinc in the crystallisation of insulin within secretory granules. However, a conflict between the interpretations of the common variant associations compared to the T2D-protective effect of rare loss-of-function variants, which also suggests that the effects of both ZnT8 activators and inhibitors will need to be examined in suitable models prior to clinical trials in human [229].

Later attempts to pinpoint T2D risk variants through whole exome or genome sequencing have been technically more successful than the study described in paper 5, yet have not produced a wealth of positive findings of association between individual rare variants

or a gene-based burden of rare variants with T2D. While exome sequencing of 2000 samples at a medium depth was an enormous task in 2008, the development in sequencing technologies has since been fast and the amounts of data we produced for paper 5, are now routinely produced in a fast manner [208]. We repeated the exome sequencing experiment in the same cases and controls, but at a higher depth of 56× enabling evaluation of the contribution of rare variants. We performed single variant analysis and collapsing test of gene-based bins of rare variants, however, did not detect novel variants or gene-based burden of rare variants associated with T2D [230]. Our study was underpowered to detect modest genetic effects, but if much of the heritability of T2D is explained by variants in a modest number of genes, we should have detected at least one associated locus at our Bonferroni significance threshold. Thus, the empirical results, combined with the statistical power simulations, suggested that when clustered in fewer than 20 genes, coding variants of moderate effect do not account for much of the missing heritability of T2D. These data have recently been included in collaborative T2D exome sequencing study to increase sample size.

While exome sequencing remains an important intermediate, much focus has been oriented towards whole genome sequencing. In 2014, DeCode Genetics published the first whole-genome sequencing study of T2D including whole genome sequencing of 2,630 individuals and imputation into 11,114 patients with T2D and 267,140 individuals without diabetes. This study discovered four novel low-frequent or rare variants associated with the disease [170]. Of interest, among these findings was the discovery of a protective variant in *CCND2* [170], which follows the identification of rare protective variants in *SLC30A8* [211] indicating that rare variants potentially leading to loss-of-protein function may not necessarily lead to increased risk of disease. Furthermore, two low-frequent missense variants in *PAM* were identified. We replicated these findings in Danish samples [170]. In 2016, the GoT2D-T2DGenes Consortium published the analysis of an huge amount of data consisting of whole genome sequencing in 2657 individuals with or without T2D with imputation into 111,548 samples, which was supplemented with exome sequencing of 12,940 individuals and exome chip genotyping in ~80,000 individuals [216]. In these massive amounts of sample, very few novel susceptibility genes were discovered. The genome analysis replicated most findings previously done by DeCode Genetics, but did not pinpoint novel risk alleles with MAF <5%, while the exome-sequencing focused analysis found a coding susceptibility variant in East Asians located in *PAX4*. Furthermore, both empirical and simulated data from this study suggested that low-frequency and rare variants contribute much less to T2D heritability than do common variants [216]. Another large-scale international collaborative sequencing effort was performed by the UK10K consortium and the main results of association analyses with metabolic traits were published in 2015 [62]. The consortium performed whole genome sequencing of 3781 individuals and in line with studies described above the findings of novel association signals were relatively sparse, however, included rare variants in *APOC3* and *LDLR* associated with serum triglyceride and low-density lipoprotein (LDL) cholesterol, respectively [62]. The *APOC3* variant was simultaneously found in other efforts [231-233]. In fact, within lipid traits, some very early sequencing studies were pioneering for targeted sequencing of a single gene with high prior likelihood of being involved in disease or sequencing individuals at the extreme of the lipid phenotype distribution [210, 234, 235]. The rather strong impact of certain coding variants on circulating lipid levels has also generated positive findings in studies of whole exomes or genomes besides the *CD300LG* variant discovered in paper 5. Among such studies are the massive UK10K sequencing study [62] and investigations of the DeCode Genetics resource [217]. Recently, Dewey et al. published the study of the exome sequence data of 50,726 individuals combined with lipid levels from electronic health records [236]. The study confirmed the association of genetic variants in genes that are drug targets (*NPC1L1* and *PCSK9*) and identified variants or a gene-based burden of variants associated with plasma lipids among others at *APOC3*, *LDLR* and *APOB* [236]. As for other studies [62, 237] this study observed the expected inverse relationship between allele frequency and effect size.

These initial massive whole-genome sequencing-based papers taking the first dive into the landscape of rare variants in T2D susceptibility have all been characterised by relatively few novel findings. With exception of picking of the lowest hanging fruits on the tree of rare variants, although we cannot exclude the role of rare variants in T2D susceptibility, the individual effect sizes of these variants are relatively low. Therefore, where initial studies were rather small it is clear that similar or even higher than GWAS-like sample sizes are needed for the individual detection and

validation of such variants using the best of methods developed for individual and combinations of rare variants [198, 238]. Nevertheless, exome sequencing, and possibly genome sequencing, are important in future genetics studies of complex diseases but may due to higher costs currently be best placed to complement array-based GWAS by producing allelic series of rare variants, which aids experimental gene characterisation and points to the effect of modulating the target by drugs.

From sequencing back to chip-genotyping to discover the association of coding rare and low frequency variants

As discussed previously, low-frequency variants and especially rare alleles are poorly characterised by GWAS genotyping arrays. Sequencing across the exome or genome can directly assay low-frequency and rare variants but such approaches are currently still rather costly and studies of tens to hundreds of thousands of individuals are limited. One proposed method for testing low-frequency and rare variants is to first sequence the exome to discover the variation and then to genotype the discovered variants in a larger number of individuals from the same or similar populations to test for association with phenotype. The Illumina Human Exome genotyping array (exome chip) was designed based on this principle and contains ~200,000 coding sequence variants discovered from sequencing the exomes of ~12,000 individuals [239].

By applying the cheaper array genotyping technology, is has been possible to reach higher sample sizes which has contributed with both discovery of novel variants in the rare and low frequency areas and with interesting contributions to evaluation of the importance of coding variation in complex diseases. Such studies have been performed for a number of different human diseases and phenotypes. In the largest study of T2D published in 2018, we performed exome array genotyping in 81,412 patients with T2D and 370,832 individuals without diabetes and by this 40 coding variant signals across 38 loci were discovered, hereof 16 novel and five driven by a variant with MAF <5% [142]. All had modest effect sizes with OR below 1.5. To seek to identify the causal variant responsible for the association signals, and thereby seek to determine if coding variants are causal for the coding lead variants, a credible set of variants accounting for 99% of the posterior probability of driving the association was performed. This approach incorporated an annotation-informed prior model of causality, which boosts the posterior probability of driving the association signal that is attributed to coding variants. Of interest, evidence for a causal signal coming from a coding variant was observed in only 16 of 38 loci and in 13 of 38 loci, the association signal was clearly driven by non-coding variation (Fig. 12) [142]. Hence, the basic idea that coding variant signals would directly lead to biological insight in complex disease is challenged, and great care is needed to identify causal contribution for such coding disease-associated variants. However, this may especially be the case since most coding lead variants identified thus far are common in the population. While common variants are typically in LD with a number of other coding or non-coding alleles, rare variants tend to be in LD with very few variants and hence the likelihood that rare coding variants associated with disease are causal and change the function of the gene in question is higher.

While this analysis sheds light of the relative contribution of coding and non-coding variants at specific loci identified through a coding variant, it is biased in relation to the relative contribution of coding and non-coding causal variants to the general genetic architecture of T2D. The recent large-scale GWAS based on HRC-imputed data in a large sample size [47] described earlier may shed some light on the overall relative contribution of coding and non-coding variants to inheritance of T2D. Of 51 loci with strong indications of only a single causal variant, eight of these lead variants were coding missense variants. This finding indicates that coding causal variants are actually over-represented compared to their genomic abundance, which is somewhat in contrast to the initial findings from GWAS [240], however this conclusion may be biased towards coding variants. In essence, the genomic complexity is huge and there is no certainty that finding a coding variant association signal thereby also identify the causal variant and gene – when a coding variant is lead it may be due to a non-coding causal variant and the other way around. What is needed is full genomic coverage in large numbers of individuals to narrow the possible local causal variant(s) by statistical and functional means.

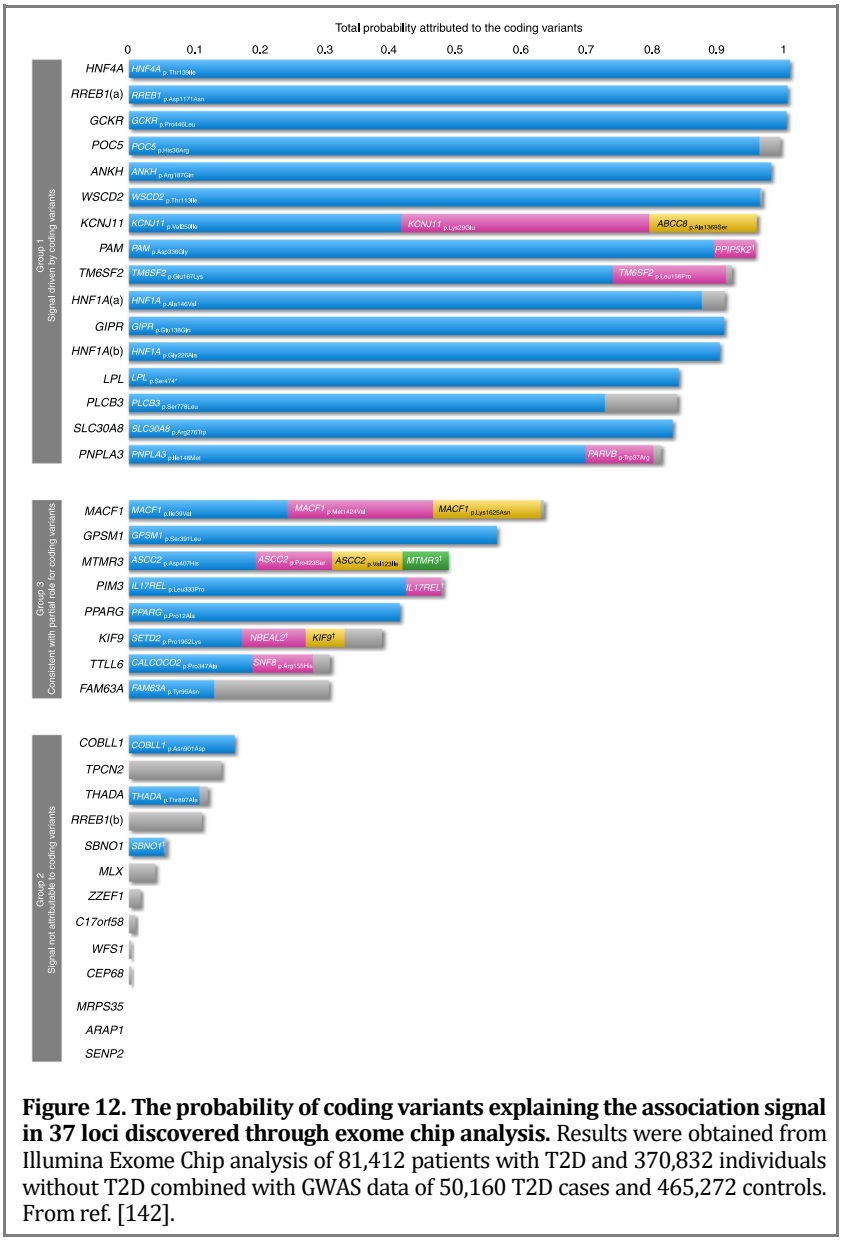

Figure 12. The probability of coding variants explaining the association signal in 37 loci discovered through exome chip analysis. Results were obtained from Illumina Exome Chip analysis of 81,412 patients with T2D and 370,832 individuals without T2D combined with GWAS data of 50,160 T2D cases and 465,272 controls. From ref. [142].

What is also evident is that the sample size of studies is of outmost importance in the attempt to identify variants with lower frequency. This is illustrated by the recent GWAS of T2D, which found an unprecedented number of low-frequency and rare variants associated with T2D [47]. This effort identified 56 low-frequency and 24 rare variants in 60 loci, of which 6 were known T2D loci. Of interest, 14 of these 80 variants had an OR above 2 and these 80 risk variants explained 1.1% of phenotypic variance in T2D. These findings are a tremendous progress in discovery of risk variants with lower frequency but are still in agreement with the conclusion that low-frequency and rare variants contribute much less to T2D heritability than do common variants. The obtained progress largely stems from the large sample size of 900,000 individuals combined with the dense imputation applying the HRC reference panel.

Also, for other metabolic traits, data coming from the exome chip has contributed substantially to the understanding of coding variants in the genetic contribution to variation in these phenotypes. A major study of more than 300,000 individuals found 444 independent variants in 250 loci associated with circulating lipids (total cholesterol, LDL-cholesterol, HDL-cholesterol and triglyceride) [176]. The paper highlighted the different origin of lipid-changing

variants since associated variants show different pattern of association with related diseases and phenotypes such as T2D, coronary artery disease, age-related macular degeneration and liver fat and the study serves as an important paper to qualify targets for medical intervention based on genetic information [176]. The primary analysis of BMI included 526,508 individuals and identified 14 coding variants in 13 genes. Most notably for this analysis, some rare alleles were identified with large impacts on BMI. The strongest impact was for a variant with a MAF of 0.01% in *MC4R*, which showed an effect of 0.54 SD of BMI per minor allele equivalent to 7 kg body weight per BMI-increasing allele [67]. Along the same lines, analyses of waist-hip ratio (WHR) adjusted for BMI identified 41 independent association signals, hereof a few rare variants [241]. Thus, the latter obesity-related paper of WHR identified fewer association signals than the former paper investigating BMI, which may be due to lower sample size, lower phenotype precision, sex-dimorphism, different genetic architecture or other factors.

While data generated from the exome chip have contributed quite a lot to understanding the genetic and biological contribution to metabolic traits, this exact chip can only be seen as an intermediate step on the way to higher resolution of rare variants in a high number of samples. As described, the exome chip is based on exome sequence data from 12,000 samples predominantly from Europeans limiting the range of rare alleles to what has been found in these samples. In European samples, 81.6% of protein-altering variants with MAF >0.5% are captured using the Exome Chip, while this numbers decreases to 25% when considering all MAFs [216]. These numbers are based on ~4500 European samples and the fractions covered by the Exome Chip will decrease when comparing with higher sample sizes and when considering individuals of non-European origin. It is thus evident that a better capture of rare alleles still requires exome or genome sequencing. Furthermore, accurate calling of genotypes from array-based technologies is challenging the lower the number of alternative allele carriers possibly setting a lower boundary for the MAF of variants being interrogated by this approach.

Applying genome and exome sequencing to study a complex phenotype

Paper 6 serves as an example of using and combining whole genome and exome sequence-based data to elucidate the genetic architecture of a complex phenotype. The study was performed in collaboration with DeCode Genetics in Iceland. At DeCode Genetics, a major genetic resource has been built applying the principle of collecting samples from as many as possible from the Icelandic population. A subset of these samples has been whole genome sequenced to discover variation in the Icelandic genome [242]. At the time when we carried out the project described in paper 6, 1176 individuals had been whole-genome sequenced – a number which has since increased substantially. The whole-genome sequence data were subsequently used to impute into all samples, which had been genotyped by genome-wide arrays. In the Icelandic population, genotype imputation is made more accurate through long-range phasing [243]. Additionally, the Icelandic genealogical database allowed for further propagation of the sequence information, applying genealogy-based imputation, into relatives of the chip-genotyped individuals [244]. To capitalise on the data generated previously [5], we used the genotype data set of 16,192 coding variants produced in stage 2 (Fig. 11), to study the impact of coding variation on the levels of the water-soluble B vitamins, vitamin B_{12} and folate, in the general population. Vitamin B_{12} is solely produced by bacteria and archaea and the only natural source of this vitamin for humans are through food items of animal origin [245]. Vitamin B_{12} and folate are enzyme cofactors or substrates in one-carbon metabolism, a process whereby folate transfers one-carbon groups in a range of biological processes including DNA synthesis [245, 246]. Several epidemiological studies have observed associations between lower circulating vitamin B_{12} levels and adverse cardio-metabolic health profiles, with insulin resistance, adiposity and cardiovascular disease [247-251].

In the study presented in paper 6, we performed meta-analyses of results from ~8400 Danish individuals with the results from a high-coverage GWAS from the Icelandic population to obtain a sample size above 30,000 individuals. In this analysis, we found 13 genetic loci, which were associated with variation in the levels of circulating vitamin B_{12} or folate. Of the 11 loci associated with serum vitamin B_{12}, five were novel and six were previously reported either in populations of European or East-Asian ancestry [252-255] (Fig. 13).

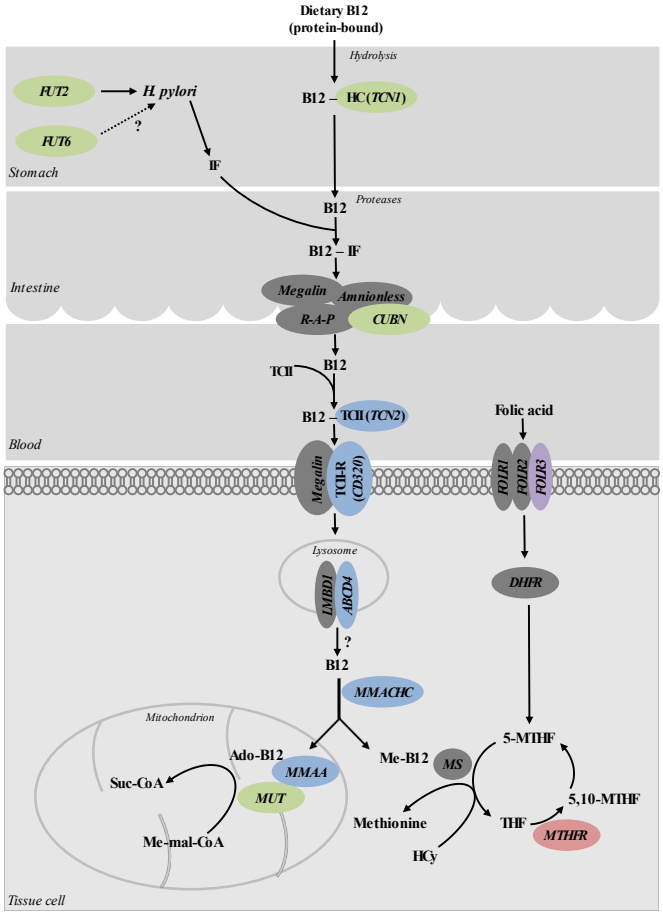

Figure 13. Genes that associate with serum vitamin B_{12} and folate levels. Genes previously identified to possess variants associated with serum levels of vitamin B_{12} are shown in green. In blue are novel genes identified in paper 6. In red, genes with variants previously associated with serum folate and in purple are novel genes for serum folate. Abbreviations: HC: heptocorrin (TCN1); IF: intrinsic factor; R-A-P: receptor-associated-protein; CUBN: cubilin; TCII: transcobalamin II (TCN2); TCII-R: transcobalamin II receptor (CD320); MMACHC: methylmalonic aciduria (cobalamin deficiency) C type; MMAA: methylmalonic aciduria; ABDC4: ATP-binding cassette, sub-family D, member 4; LMBD1: LMBR1 domain containing 1; FOLR1– 3: folate receptors 1–3; Ado-B12: adenosyl-cobalamin; Me-B12: methyl-cobalamin; Me-mal-CoA: methyl-malonyl-coenzymeA; Suc-CoA: succinyl-coenzymeA; MUT: methylmalonyl-CoA mutase; H. pylori: helicobacter pylori; DHFR: dihydrofolate reductase; MS: methionin synthase; THF: tetrahydrofolate; 5,10-MTHF: 5,10-methyl-tetrahydrofolate; Hcy: homocysteine MTHFR: 5-methyl-tetrahydrafolate reductase. From paper 6 with permission.

Association analyses with serum folate yielded one novel locus, *FOLR3*, and confirmed the reported *MTHFR* locus (Fig. 13). As discussed above, the mere finding of a coding variant associated with the trait or disease does not necessarily imply that this is the causal variant. In this study, the meta-analysis was biased towards finding coding variants, since the Danish data contained only coding variation discovered in exome sequencing, thus resulting in higher combined sample size for coding variants [6]. To evaluate each locus for a stronger non-coding association signal, we used the Icelandic whole-genome imputed data. Interestingly, the strongest signal at 10 of the 11 loci associated with serum B_{12} in the Icelandic genome-wide data corresponded to a coding variant with only the *FUT6* locus having a stronger non-coding signal. Although we did not perform advanced modelling, as previously presented for T2D, this is still indicative of the coding variants as the most likely causal variant at these loci.

Biologically, the identified coding variants fall in genes encoding proteins known to be involved in the metabolism and signalling pathway of vitamin B_{12} and folate (Fig. 13). This is a rather unique scenario within sequencing- or GWAS-based discoveries in complex traits where findings generally tend to fall in unknown biological disease mechanisms [99, 256]. The known biological effects of these genes also further strengthen the likelihood that the identified coding variants are causal.

One of the interesting perspectives of finding genetic variants associated with a phenotype or disease is that it may enable studies of the causal relation between phenotypes. Mendelian randomisation is a method using variation in genes of known function to examine the causal effect of a modifiable exposure on disease in observational studies and has become a method of choice to strengthen causal inference in observational research. Mendelian randomisation is based on the realization that a genetic variant associated with an exposure can be used as an instrumental variable to estimate the causal effect of the exposure on an outcome of interest [257, 258]. Such studies have also been performed to shed light on the causal relationship between circulating vitamin B_{12} levels and cardio-

metabolic outcome. These studies are based on genetic findings of paper 6 and other GWAS of vitamin B_{12} and folate levels [6, 252-255]. Two studies found no causative effect of vitamin B_{12} levels on BMI [259] or levels of circulating lipids and blood pressure [260]. In agreement, a recent study applying a 2-sample Mendelian randomisation design and thereby reaching a rather large sample size found no evidence of causal effects on BMI and other adiposity-related phenotypes, fasting insulin or lipid levels, yet did observe indications of a causal effect of vitamin B_{12} on fasting glucose levels [261]. In general, these studies do not support the causal role of vitamin B_{12} levels in relation to cardio-metabolic traits and question the proposed cardio-protective effect of general vitamin B supplementation [262, 263].

The future of sequencing in complex metabolic disease

The two papers included in this section of the dissertation [5, 6] both takes their starting point in an exome sequencing of 2000 Danes. From our study and from later published and emerging studies, it is clear that the effect sizes inflicted by most disease-associated rare and low-frequency variants are rather low although there are a few exceptions [47, 137, 142, 170, 211, 216, 230]. The consequence of this is clearly that more samples are needed to pinpoint rare variants in T2D and other complex diseases [198]. While genome imputation from array genotyping has improved by application of large reference panels, such as Haplotype Reference Consortium [63] or UK10K [62] or the coming TOPMed reference panel (https://www.nhlbiwgs.org/), a lower limit on MAF still exists. This brings forward the need for sequencing in large populations to capture the range of range variants, which continues to be an important task, since rare susceptibility or protective variants, as exemplified by the genes mentioned above, point directly to biology, targets for novel drugs and potentially as tools in precision medicine.

In the future, collaboration and big resources are needed to obtain thousands of samples, which are needed to obtain statistical power. A number of large projects are generating such data and collaboration across different projects is increasingly taking place. Some of these initiatives will be driven by national scale initiatives. In UK, the "Genomics England" sequencing initiative has the ambition to perform whole-genome sequencing of 100,000 individuals. Other initiatives are driven by the private sector, for instance, Regeneron aims to whole-exome sequence the entire 500,000 individuals of the UK Biobank. As well as sequencing large populations of European ancestry, it is also important to apply sequencing to more diverse populations. This topic will be discussed in more detail in Part 5.

PART 4. METABOLIC GENOME RESEARCH IN ISOLATED POPULATIONS

The great majority of genomic research within T2D and metabolism has been performed in the European population. As described, major GWAS meta-analyses have been performed in this population leading to the identification of hundreds of genomic loci [47, 76]. GWAS have also been performed in other large populations such as the South Asian and East Asian population [222, 264, 265]. In general, the genetic heterogeneity between these populations in relation to T2D risk is relatively low showing that the effect of lead SNPs from European GWAS are transferable to other large populations in spite of allele frequencies being very different [266]. Nevertheless, other populations than the European population are important for studying T2D and metabolic genetic risk factors partly since differences in LD between populations can help defining causal variants for each locus across populations [46]. Furthermore, other populations may also harbour unique genetic variation, which brings people at risk of disease. These variations are often rare or absent in the large European population or only put people at risk in specific given combination of genetics and environment. On the other extreme than large, open populations as the European are isolated populations. Isolated populations are living in limited geographical regions separated from other populations and inhabitants therefore share environment to a much larger extent than inhabitants in large outbred populations, such as the European population. At the same time, diseases as T2D are more likely to have a specific genetic aetiology in an isolated population [267]. One such isolated population is the Inuit population living in Greenland.

Genomic characteristics of isolated populations – what are the implications for genetic association studies?

Different isolated populations have many features in common even though the extent of the implications of being an isolated population depends on depth of the genetic bottleneck, length of isolated period, specific environmental factor and other factors. The current description is based on the historically isolated population of Greenland, which is the basis of papers 7-9. Due to its inhospitable environment, Greenland was one of the last areas on Earth to be populated and today remains one of the least populated countries in the World with a population of less than 60,000 individuals. Historically, Artic people have populated Greenland for thousands of years but the current Inuit population has lived less than 1000 years in Greenland [268, 269]. The Inuit people has lived isolated in the Arctic for thousands of years and migrated to Greenland from the northern part of Canada [268, 270]. Historically the Inuit population has been small and living in isolation and during these many years the Inuit people has adapted culturally and genetically to the extreme environmental conditions of the Northern Arctic. Starting with the arrival of the priest Hans Egede in 1721, Europeans have entered Greenland, which has led to genetic admixture. Studies of the modern Greenlander has shown an average proportion of European ancestry of 25% and that 80% have some degree of European genetic ancestry [270]. The degree of European ancestry varies a lot across Greenland and is by far less in the isolated areas in the north and east and in the small villages in the south of Greenland [270] (Fig. 14).

Historically isolated populations, such as the Greenlandic Inuit population, have several distinct genomic characteristics, which have implications for their use in genetic disease mapping. First, LD is generally much higher in Greenlanders than in the European population or other large populations, although the recent European admixture in Greenland has reduced the LD significantly compared to the ancestral Inuit population [270]. This feature makes indirect association mapping, which is the principle of GWAS, more efficient, since each genotyped marker tags variation in a larger genomic region. Second, the Greenlandic individuals have less genetic variability than found in larger open populations. Especially they have fewer rare variants than seen in large population changing the frequency spectrum of this population towards more common variants [270].

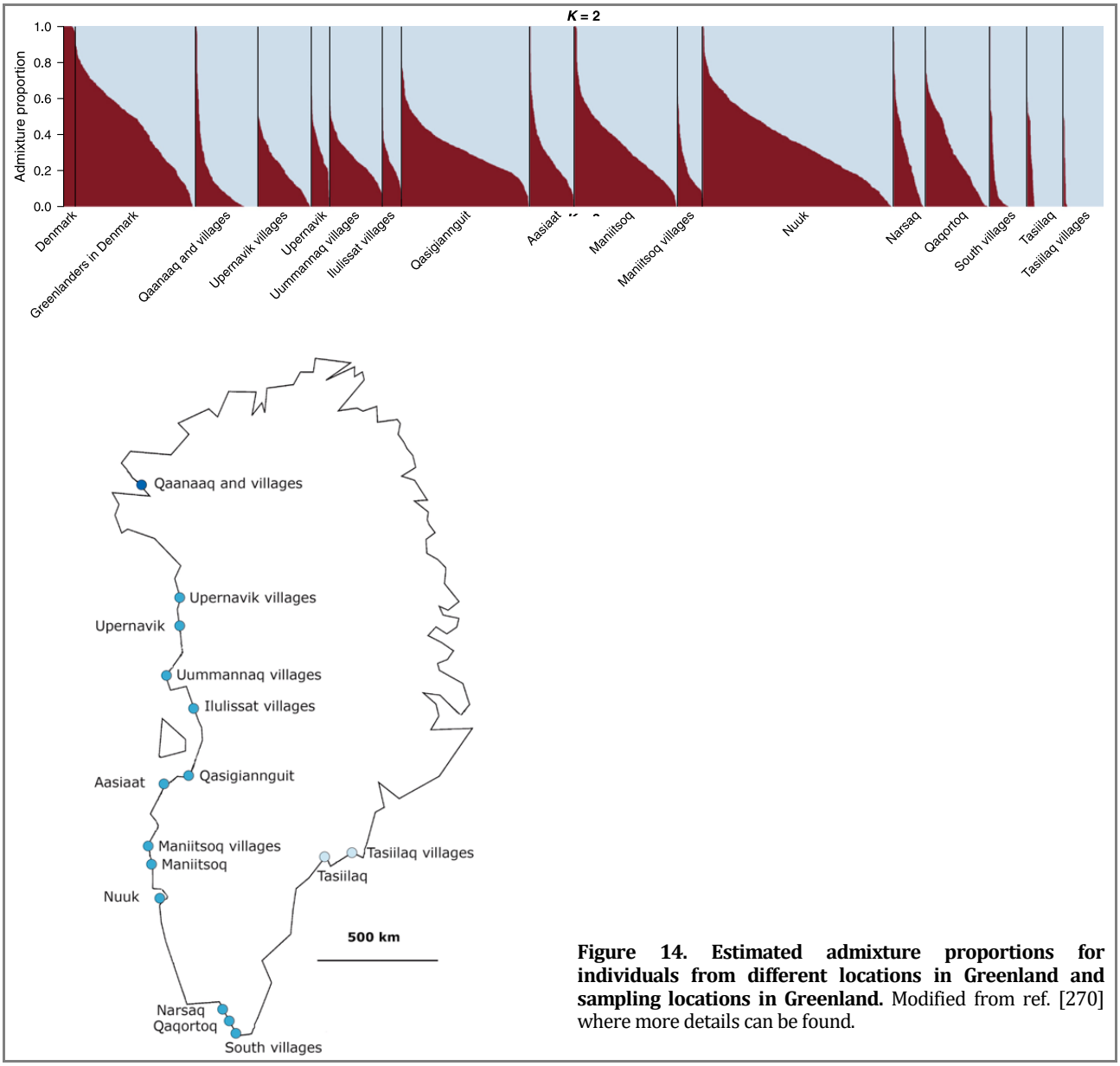

Figure 14. Estimated admixture proportions for individuals from different locations in Greenland and sampling locations in Greenland. Modified from ref. [270] where more details can be found.

The reasons for this are the founder effect – i.e. loss of genetic variation that occurs when a new population is established by a very small number of individuals – and the increased power of genetic drift caused by a long-standing isolation and small population size. Hence disease-causing variants, if not absent in Inuit, will have a higher probability of being common, which leads to increased statistical power to be detected by association mapping. As such, the history of the Greenlandic population provides several advantages in genetic association studies, which are to some extent shared with other isolated populations [271]. Yet, the strength of these features vary according to local factors such as the length and degree of isolation, the depth of the founder effect and the local environmental circumstances. However, the Inuit population is different from well-studied founder populations, such as the Finnish and the Icelandic populations, since it is not genetically close to any large population. Estimated by F_{st}, which is a commonly used measure for population differentiation, the European population and the Han Chinese population are genetically closer than the Greenlandic Inuit population and one of its genetically closest related large populations, the Han Chinese population [270]. A major reason for this is that the Inuit population has been small and isolated in the Northern Arctic for a much longer period than most other isolated populations.

Diabetes and obesity in the Greenlandic population

Although T2D prevalence is generally high and increasing worldwide, some specific populations and ancestry groups have different occurrences of T2D [272]. The classic example is the Pima Indian population in Arizona. This population has one of the highest prevalence of T2D, which has increased during a transition from a traditional lifestyle with limited food supply and high physical activity to a modern Westernised lifestyle. Their diabetes is characterised by features of the metabolic syndrome, i.e. obesity, insulin resistance and dyslipidaemia [273, 274]. In the Greenlandic population, the prevalence of T2D defined by World Health Organization 1999 criteria [275] has increased dramatically. In the early 1960-ties, a survey of the Greenlandic population was performed and found almost no cases of diabetes [276]. In contrast, the latest population surveys conducted in 1999-2010 have shown a prevalence of more than 10% [277, 278]. The prevalence of T2D is on the same level as in India and twice as high as in Denmark [10]. Obesity is the most important risk factor for diabetes and the prevalence of obesity, mostly abdominal obesity, is high among the Inuit population of the Arctic region [279] and has been increasing since the 1960-ties [280, 281]. Several studies have shown that the obesity-dependent risk of metabolic disease and complications is lower than in the European and other populations [279, 282, 283]. Furthermore, a study has indicated that Greenlanders have more subcutaneous fat tissue than Europeans for any given waist circumference, possibly partly explaining the lower cardiovascular risk associated with obesity in Greenlanders [284]. Importantly, the rapid societal changes likely play an important role in changing lifestyle and risk of obesity and T2D. Of interest, the prevalence of T2D in Greenland has been found to be higher in rural than in urban areas [285], which is in contrast to most other populations, where urbanization is a risk factor for T2D. In addition, the association between dietary intake and risk of T2D is also different than expected since a study has found higher risk in Greenlanders eating a traditional Inuit diet with a high content of marine mammals and fish [286]. The traditional way of living in Greenlandic is physically demanding. However, this has changed with the general change in lifestyle towards people being less physically active both at work and in their way of transportation. In data measuring physical activity by movement and heart monitoring, level of physical activity was not associated with T2D, but with estimates of peripheral insulin sensitivity, which is a risk marker of diabetes development [287]. Changes in traditional lifestyle risk factors therefore do not seem to explain the high T2D prevalence, although confounding by admixture may partly have influenced the results of these studies.

Genetic risk factors may cause some of the high risk of T2D in the Greenlandic population and may possibly explain some of the epidemiological findings. The ambition of the research in the Greenlandic Inuit historically isolated population included in this dissertation was to map genetic risk elements related to T2D and metabolism. The investigations presented in papers 7-9 were performed in up to 5,000 adults sampled as part of the two health surveys of the Greenlandic population, B99 and IHIT and from a cohort of Greenlanders living in Denmark [277, 278]. In these population surveys, T2D-related phenotypes were collected, including data from an OGTT in all individuals above 30 years of age in the two investigations performed in Greenland. This sample represents up to ~10% of the adult Greenlandic population and was sampled from all parts of Greenland (Fig. 14).

The *TBC1D4* p.Arg684Ter nonsense variant imposes a high impact on type 2 diabetes and insulin resistance

In paper 7, we described our initial genetic association analysis in the Greenlandic population [7]. We performed a genetic association study of variation on the MetaboChip [288] in relation to fasting and 2 hr plasma glucose and serum insulin during an OGTT. Here we discovered an intronic variant in *TBC1D4*, which was associated with plasma glucose and serum insulin 2 hr after an oral glucose load. By exome sequencing in a small subset of the population, we found a p.Arg684Ter nonsense variant in *TBC1D4.*

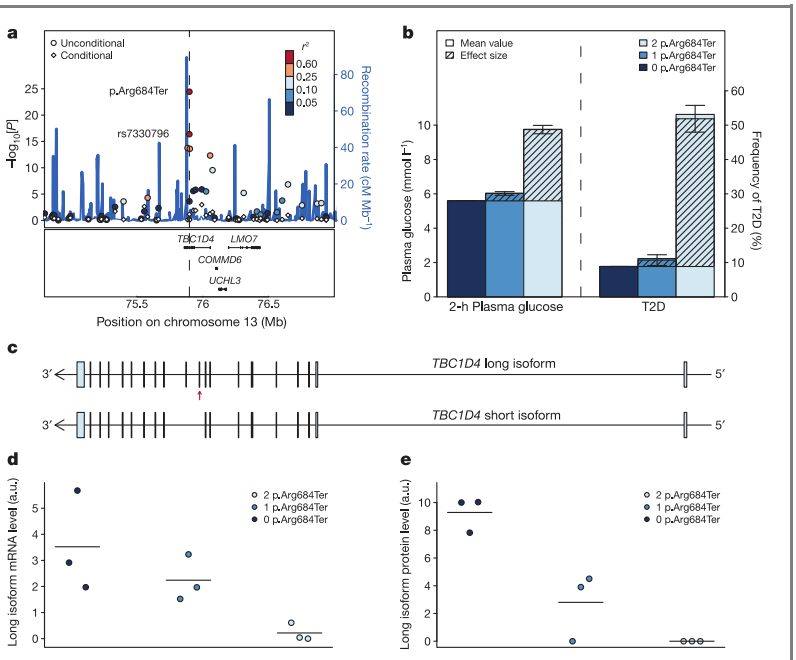

Figure 15. Effect of the p.Arg684Ter variant in *TBC1D4*. a, Regional association with 2-hr plasma glucose levels. **b,** The mean 2-hr plasma glucose and the frequency of T2D for three genotype groups. **c,** The two predominant isoforms of *TBC1D4* illustrating which exons are transcribed. **d,** The mRNA expression level of the long *TBC1D4* isoform in skeletal muscle from nine Greenlandic individuals. **e,** TBC1D4 protein long isoform in skeletal muscle from nine individuals. From paper [7].

This variant was likely to be the causal variant and it mainly has a recessive genetic effect (Fig. 15). Approximately 4% of the Greenlanders are homozygous carriers of p.Arg684Ter. On average, these individuals had 3.8 mmol/L higher 2-hr plasma glucose, 160 pmol/L higher 2-hr serum insulin during the OGTT, a 1 SD decreased insulin sensitivity and much higher risk of T2D (OR 10.3) as compared to the rest of the population (Fig. 15) and of interest, 88% of the homozygous carriers above 60 years have T2D. These effect sizes are 10- to 50-fold higher than reported for common variants on these traits in the European populations [46, 47, 87, 89]. Interestingly, homozygous carriers had the same level of obesity and circulating lipids as the rest of the population, supporting the very specific nature of this T2D subtype. The *TBC1D4* p.Arg684Ter nonsense variant is located in the long isoform of *TBC1D4*, which is not expressed in homozygous carriers of the variant (Fig. 15).

The findings from paper 7 were subsequently corroborated in a study of 1141 Inuit from Nunavit, Canada and Alaska, which were investigated by an OGTT [289]. Of interest, the *TBC1D4* p.Arg684Ter variant was present at comparable frequency in these populations indicating a presence throughout the North American Arctic Inuit populations.

TBC1D4 encodes a protein that acts as a mediator of insulin-stimulated cellular glucose uptake through increasing GLUT4 translocation [290]. *Tbc1d4* -/- knockout mice have decreased plasma glucose levels, have lower GLUT4 levels and have decreased insulin-stimulated glucose uptake in muscle and adipose tissue compared to wild-type mice [291, 292]. TBC1D4 may also play a role in the insulin sensitising effects of exercise [293, 294]. Subsequent research points to that the effect of *TBC1D4* p.Arg684Ter on 2-hr plasma glucose levels may be modified by the level of physical activity of the individual such that very physically active homozygous carriers have a smaller genetic effect on 2-hr plasma glucose [295]. This finding indicates that variant carriers would benefit from more specific intervention with physical activity to avoid T2D. In addition, these results point to TBC1D4-independent mechanisms between physical activity and glucose uptake.

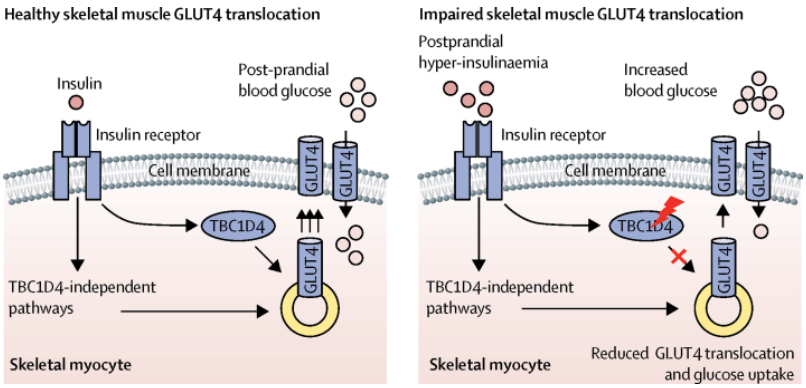

Figure 16. Simplified model of the role of TBC1D4 in insulin-stimulated glucose uptake in the skeletal muscle and the defect caused by the *TBC1D4* p.Arg684Ter variant. From ref. [296] with permission.

Taken together, our findings therefore suggest that *TBC1D4* p.Arg684Ter in homozygous carriers imposes a reduced insulin stimulated translocation of GLUT4 to the cell membranes of skeletal muscle, which leads to post-prandial hyperinsulinemia and hyperglycaemia (Fig. 16). As such, the variant causes isolated peripheral insulin resistance, which leads to T2D. Constitutive recycling of GLUT4 causing an elevated level of GLUT4 in the muscle cell membrane in the fasting state is probably responsible for the decreased fasting levels of glucose and insulin observed in homozygous carriers.

Common T2D is characterised by pathophysiologic defects of multiple organs, each contributing to the overt disease. However, homozygous carriers of the identified *TBC1D4* nonsense variant, have a specific subtype of T2D with what seems to be a tissue-specific molecular defect. Homozygous *TBC1D4* p.Arg684Ter carriers are almost exclusively diagnosed with T2D due to elevated 2-hr plasma glucose levels (Fig. 17) underlining the specific nature of the inflicted T2D subtype.

The finding of a genetically induced T2D subtype has several implications. First, the single molecular defect in patients with this T2D subtype opens for detailed studies of biological and physiological mechanisms in humans. Such studies will shed new light on mechanisms between a dysfunction of TBC1D4, insulin resistance and T2D. Second, epidemiological investigations of the long-term risks associated with this variant will help to elucidate the risk associated with isolated postprandial hyperglycaemia. As such, the *TBC1D4* variant will work in a Mendelian randomisation framework to shed light on the causal relationship between hyperglycaemia and endpoints such as T2D complications including cardiovascular disease. As opposed to most Mendelian randomisation studies, studies involving *TBC1D4* will have the advantage that the functional implications of the *TBC1D4* variant are rather clear. For instance, T2D is known to impose risk of cardiovascular disease and it has been disputed whether fasting plasma glucose levels, 2-hr OGTT plasma glucose or HbA_{1c} levels are better at predicting risk of CVD [297-299]. Yet, although the variant imposes a strong effect on glycaemia and T2D, the low number of homozygous carriers with incident CVD may impede statistical power of such analyses. Third, defining a subtype of T2D may have future implications for treatment and prognosis of T2D in this subgroup, although the specific implications and evidence is still not available. For instance, it seems logical that *TBC1D4* variant carriers would not benefit from treatment with exogenous insulin since they have a high endogenous insulin production.

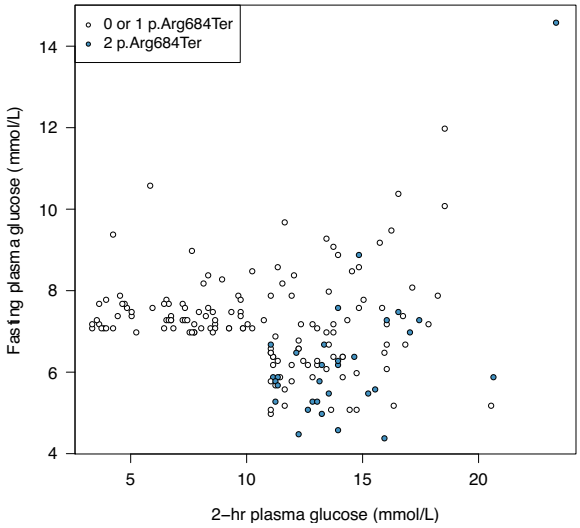

Figure 17. The fasting and 2-hr plasma glucose levels of individuals with T2D in the Greenlandic cohorts stratified on *TBC1D4* p.Arg684Ter genotype. From ref. [271].

Is the high effect size of *TBC1D4* due to genetic drift or positive selection?

The high effect on 2-hr plasma glucose, 2-hr serum insulin, insulin sensitivity and T2D of the variant in *TBC1D4*, calls for considerations on why the variant has a frequency as high as ~17% in the Greenlandic population. In fact, the MAF in the ancestral Inuit part is estimated to be 23% and it is present at comparable frequency in other populations in the Northern Artic [289]. In contrast, in ~123,000 individuals in data from the gnomAD browser, the variant was only found in the heterozygous form in two European individuals (http://gnomad.broadinstitute.org/variant/13-75898521-G-A). Several possible explanations for how the variant can have reached such high frequency in an isolated population, which has been through a genetic bottleneck. First, it is possible that the reason is random genetic drift, which in small populations like the Greenlandic Inuit population, has a strong impact on allele frequencies [300, 301]. Genetic drift is the random fluctuation of allele frequencies and in large populations, where detrimental variants are efficiently removed by selection. However, in small populations, selection is less effective and harmful variants may increase in frequency. As described, few variants segregate at low frequency in this population as variants are either lost or driven to high frequency due to the forces of genetic drift [270]. Alternatively, forces of positive selection favouring the alternative allele may be the reason behind the high frequency. This could be interpreted as the *TBC1D4* p.Arg684Ter variant previously was associated with a beneficial effect in the context of the traditional Inuit diet with a low intake carbohydrate [302]. Even though, in the modern Greenlandic society, it poses a high risk of T2D, this may be a relatively new feature. In paper 7, we presented suggestive evidence that this is actually the case by showing decreased variability at this locus; however, future studies of selection and of interaction between genotype and diet may shed further light on this interesting question. In fact, a specific combination of certain circumstances may be needed for such a variant to impose the effect we measure with the detailed OGTT-based phenotypes in the current data. These specific factors, which may all be necessary for expression of the phenotype, are homozygosity for the genetic variant, consuming a diet with a certain level of carbohydrate content and not being extremely physically active. At least theoretically, if any of these factors is removed, the phenotype will not manifest itself.

Applying a recessive genetic model to discover additional type 2 diabetes risk variants

The risk variant identified in *TBC1D4* was found by an additive genetic analysis although it proved to have the major effect in homozygous variant carriers under a recessive genetic model. This prompted us to repeat the genetic association analysis of T2D across all data and cohorts applying a recessive genetic model. As described in paper 8, the recessive model has a clearly improved statistical power over the additive model to detect a variant, which is genuinely imposing a recessive effect. The renewed analysis showed three loci which were associated with T2D at Bonferroni corrected study-wide significance level. These loci were the previously identified *TBC1D4* and the novel loci *ITGA1* and *LARGE1*. Of these, the variant in *LARGE1* did not reach genome-wide significance and while we sought replication in both Yup'ik Inuit individuals and in Danish samples, we were unable to validate this association further. This fact illustrates one of the difficulties, when performing genetic association studies in isolated populations such as the historically isolated Greenlandic population namely that the availability of cohorts and samples for studies and for replication is limited. This is of course the consequence of the small population size of the Arctic region and the absence of the specific genetic variants of interest in other larger populations.

The most convincing novel locus from this study was the *ITGA1* locus on chromosome 5. Of interest, the previously described recent GWAS of T2D identified three independent variants in this locus each with a distinct impact on risk of T2D [47]. However, none of these variants were the same as the Greenlandic lead variant and LD between variants is low in European population (all r^2<0.15). Furthermore, the variants identified in the GWAS were found applying an additive genetic model. For both populations, the causal variants responsible for the association signals have not been identified making comparisons difficult. This locus serves as an example on the complexity of some genomic regions. Future data in the Greenlandic population with a higher genomic resolution and recessive genetic analyses of European data may serve to shed light on the multiple signals of the region.

Characterising homozygous loss-of-function variants in humans

In studies presented in papers 7 and 8, we studied variants for which the major impact is inflicted in homozygous carriers. For the *TBC1D4* locus, we were able to pinpoint a causal variant, which was a variant leading to a loss-of-function of *TBC1D4*. Such homozygous loss-of-function variants are of special interest since they can be shortcuts to study human biology related to specific proteins. Investigating homozygous gene knockout model animals has been a key procedure in understanding the biological role of specific genes and hence proteins in health and disease including T2D [303]. This is still a necessary and important tool, yet differences between mice and men can interrupt the biological translation, limiting the translational influence of such models. Hence, there is a great interest both in academic science and from the pharmaceutical industry in identifying and characterising large sets of humans with homozygous loss-of-function alleles. Naturally occurring human genetic variants provide "experiments of nature" that can directly inform on the function of human genes. In addition to providing novel insight into human biology, they can aid in the identification and validation of genes that would be efficient and well-tolerated targets for therapeutic modification in both rare and common diseases. In light of this, finding humans with homozygous loss-of-function variants is very valuable, because detailed physiological investigations of such individuals can shed light on gene functions in humans potentially paving the way for novel drug discoveries for human diseases [304].

While identification of homozygous loss-of-function variant carriers is the goal, there are different ways to get there. Ongoing large-scale whole genome and whole exome sequencing of European individuals will reveal a high number of loss-of-function variants, however most of these possibly disruptive alleles are rare in Europeans [305], making investigations of homozygous human loss-of-function mutation carriers particularly challenging. As an example, 40,000 individuals will have to be investigated to identify one homozygous carrier of a loss-of-function allele that is carried by 1% of the population (MAF ~0.5%) and as the frequency of the loss-of-function allele drops, the required sample size to identify homozygous carriers increases exponentially. Investigation of heterozygous loss-of-function variants may be of high interest as well, as shown numerous times for instance in the described study of rare variation in *SLC30A8* in relation to T2D [211], however it adds complexity to the biological interpretation. A number of different initiatives are pursuing this approach. For instance, the DiscovEHR study, which is a private collaboration between Regeneron Genetics Center and Geisinger Health System, have applied large-scale exome-sequencing to discover loss-of-function alleles, which could be potential drug targets in T2D and CVD leading to several important findings [236, 306, 307]. Furthermore, two recent studies used data from the UK Biobank to evaluate the impact of loss-of-function variants across the genome on complex phenotypes. In these studies of between 337,000 and 406,000 individuals, a number of associations between rare or low-frequency loss-of-function variants in the heterozygous state and complex phenotypes were found [308, 309]. Examples of these are an association between rare *GPR151* variants and protection from obesity and T2D and between rare variants in *PDE3B* and elevated height, improved body fat distribution and protection from coronary artery disease [309]. However, the evaluation of loss-of-function variants in the homozygous state was mostly limited to very common variants such as rs601338 at *FUT2*.

Alternative ways to address this issue relate to specific populations with characteristics leading to changes in the genomic constitution. First, as described it is possible to use isolated populations, where the founder effects and genetic drift can make homozygous loss-of-function carriers more frequent [310]. This means that under a recessive model, the magnitude of the bottleneck and the level of isolation will determine the extent of the increase in homozygous genotypes. A study of rare homozygous loss-of-function carriers in the Icelandic population showed that 7.7% of genotyped individuals from the Icelandic population are homozygote or compound heterozygote carriers of loss-of-function variants with a MAF below 2% together covering more than 1000 genes [311]. In addition, studies of Finns have shown more homozygous low-frequency loss-of-function variants than in non-Finnish Europeans [312]. Given the length and depth of the genetic bottleneck in the formation of the present-day Inuit population [313], it is expected that the frequencies of such variants will be even higher in the Greenlandic population. Studies of the Greenlandic Inuit population have shown that this special population carries fewer deleterious variants but that these segregate at higher frequency than in other populations [314]. Furthermore, Greenlandic Inuit have

a higher genetic load compared to other populations [314].

Another option for studying homozygous loss-of-function variants to infer human biology and evaluate potential drug targets, relates to the special features of populations such as the Pakistani population or specific populations in the Middle East. Here, the marital traditions result in a population with a higher rate of consanguinity, resulting in increased frequencies of homozygous loss-of-function mutations. A recent paper used data from a cohort of 10,503 individuals from Pakistan with a high rate of consanguinity to identify a number of interesting associations between homozygosity at specific genes and complex phenotypes such as an association between *NRG4* and reduced fasting plasma insulin and C-peptide concentrations. Furthermore, homozygous carriers of loss-of-function variants in *APOC3* were identified strengthening and validating the association between *APOC3* and levels of triglycerides [315]. These findings in a relatively low sample size show the power of this approach.

The *ADCY3* locus – revealing novel biology in obesity pathogenesis

As described, the historically isolated Greenlandic Inuit population holds increased numbers of loss-of-function variants, which are not rare. Hence, a logical next step was to look specifically for loss-of-function variants, which segregates at sufficiently high frequency to enable the study of homozygous carriers. In the study presented in paper 9, we therefore used the exome sequencing data of 27 individuals generated previously [7], to detect common putative loss-of-function variants across the exome. To narrow the search, we intersected the identified variations with genes at known GWAS loci, since there is an increased chance that such genes harbour loss-of-function variants with a stronger impact on phenotype. The sole variant to emerge from this investigation was a splice variant in *ADCY3*, a locus previously associated with BMI in GWAS [316, 317]. In paper 9, we describe the relationship between this variant, c.2433–1G> A, which has a MAF of 2.3% in the Greenlandic cohorts, with metabolic phenotypes in the Greenlandic samples.

We show that the variant is associated with obesity, obesity-related phenotypes, glycaemia and T2D with the largest effect in a recessive genetic model. The seven homozygous carriers had 7.3 kg/m^2 higher BMI, 8.1% higher fat percentage and higher risk of T2D. To show that this variant actually results in loss of functional ADCY3, we performed RNA sequencing of blood cells in 17 individuals to display the aberrant splicing patterns caused by the variant. Here we showed that the variant results in two additional splicing patterns characterized by intron retention or skipping of exon 14 (Fig. 18). Thus, this mutation might impair the function of ADCY3 through shifting the expression of mRNA isoforms. However, we did not measure protein levels in a relevant tissue and it is thus unknown whether this variant modulates the structural conformation of the ADCY3 protein.

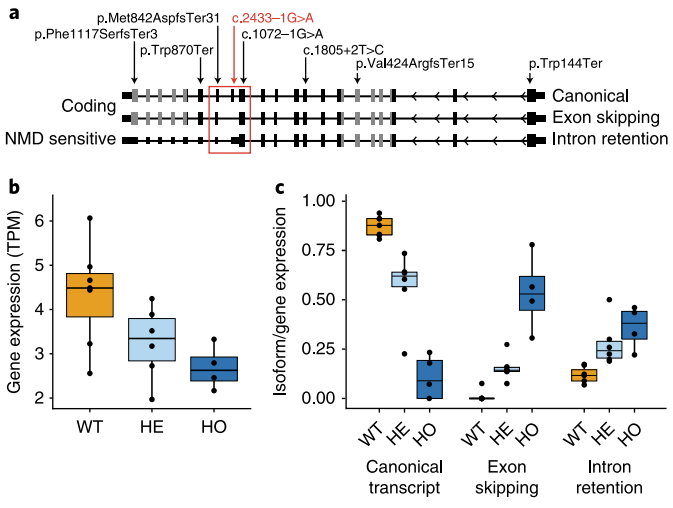

Figure 18. *ADCY3* isoforms, observed loss-of-function variants and functional consequences based on RNA. a, Illustration of *ADCY3* displaying the three relevant transcript isoforms with their predicted functional consequences. **b,** Normalized *ADCY3* expression, stratified according to *ADCY3* c.2433–1G> A variant genotype. **c,** ADCY3 transcript isoform fractions stratified according to *ADCY3* c.2433–1G> A variant genotype. From paper 9.

Of interest, our discovery of the splice variant in *ADCY3* was published in the same issue of Nature Genetics as another study finding a relationship between variation in *ADCY3* and obesity [318]. Saeed et al. investigated severely obese probands from consanguineous families from Pakistan by exome sequencing and found three homozygous or compound heterozygous, potential causative variants in *ADCY3* including a frameshift variant, a splice site variant and a missense variant [318]. The three variants were functionally characterised and shown to be loss-of-function variants that affect the catalytic activity of the encoded protein. As such, these findings corroborate our findings from the Greenlandic population and firmly establish *ADCY3* as the causative gene of this locus in obesity. At the same time, the two studies illustrate different approaches by which to arrive at homozygous loss-of-function variants associated with extreme phenotypes (Fig. 19).

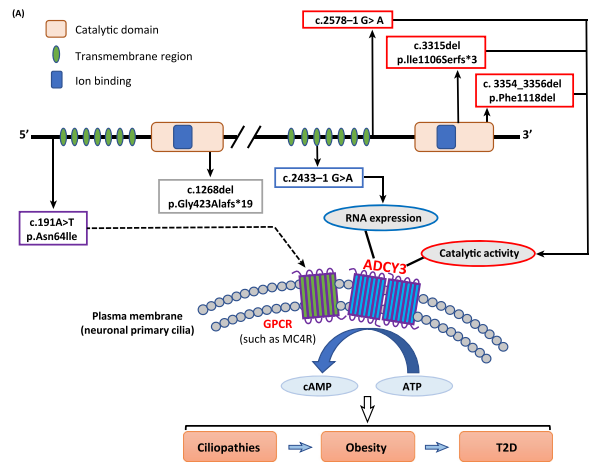

Figure 19. Novel *ADCY3* mutations [9, 318] and their possible functional consequences. Modified from ref. [319] with permission.

ADCY3 catalyses the synthesis of cyclic AMP (cAMP) from ATP. Studies of mice have associated loss of *ADCY3* function with impaired insulin sensitivity, dyslipidaemia, obesity and increased fat mass, hyperphagia, depression-like phenotypes and resistance to leptin [320-323]. Possibly, leptin resistance occurs through disrupted cAMP signalling in primary cilia in hypothalamus [322]. Interestingly, previously described syndromic forms of obesity, including Bardet–Biedl and Alström syndromes, have been found to be caused by altered function of primary cilia [324]. Along those lines, a study demonstrating that the MC4R-ADCY3 pathway in neuronal primary cilia is involved in monogenic obesity [325] was published back to back with the two genetic association papers [9, 318]. Interestingly, MC4R and ADCY3 were specifically co-localized in the primary cilia of a subset of neurons in the paraventricular nucleus of the hypothalamus. Furthermore, obesity-associated *MC4R* mutations impaired ciliary localization and specific inhibition of ADCY3 activity at primary cilia of MC4R-expressing neurons was sufficient to cause obesity in mice. These findings confirm the essential role of MC4R and ADCY3 at primary neuronal cilia in regulating body weight [325]. Together, these studies highlight a causal role of ADCY3 and provide new genetic associations and mechanistic insights in the aetiology of obesity and T2D. In conclusion, these recent studies reinforce a rationale for pursuing ADCY3 as an attractive therapeutic target for obesity and obesity-associated disorders.

Future research in the historically isolated population such as the Greenlandic Inuit

From initial studies of Greenlandic Inuit, it is evident that this is a fruitful avenue for identification of important variants and genes with biological impact on T2D and metabolic disease. Some of the reasons for this is connected to the specific genomic context of this population, as described earlier. Especially in the cases where we have been able to pinpoint causal variants as shown for *TBC1D4* and *ADCY3* make biological translation from association to disease mechanisms more straightforward than what is the experience from GWAS of the European population. While studies of isolated populations, like the Greenlandic Inuit, evidently have a number of advantages, there are also some limitations. First, collection of samples from a small population living spread over a large area in places hard to reach, is very difficult and with a small population, the sample size of a cohort will naturally be limited. Second, replication of findings from Greenlandic cohorts is difficult as evident by the study presented in paper 8. Sources of cohorts with similar characteristics are very limited.

The findings from Greenlandic Inuit presented in the current dissertation are based on what in a modern age of genomics seems as rather limited datasets consisting of MetaboChip [288] data for all individuals and exome sequencing data in 27 individuals and RNA sequencing in 17 individuals. More and deeper omics-data will

likely reveal more findings and we are currently generating extensive genomic data, including genome-wide array genotyping combined with whole-genome and RNA sequencing in a subset, in the Greenlandic population to evaluate the potential in more detail and to discover additional individuals with functionally disrupted genes. A putative approach is to re-examine such individuals and families by recall-by-genotype principles and perform physiological evaluation with statistical confidence, although such an approach poses practical challenges related to geography. Furthermore, given the high-impact genetic variants found in Greenlanders, it is likely that research in this population will serve as a forerunner for clinically translation of genetic findings.

PART 5. PERSPECTIVES ON GENOMIC DISCOVERIES IN TYPE 2 DIABETES – WHAT IS MISSING AND WHAT MAY THE FUTURE HOLD?

The genetic architecture of type 2 diabetes and complex metabolic phenotypes

The term genetic architecture in human genetics describes the characteristics of genetic variation that are responsible for the heritability of certain phenotypes [326], i.e. the complete understanding of all genetic contributions to a phenotype or disease and includes the number, effect size, allele frequency of variants and the possible interactions between the variants, or between variants and environmental factors. Within metabolic complex traits, the shape of the genetic architecture has been discussed for years. The last decade of research has shed light on specific elements in this composition and a more detailed picture across the allele frequency spectrum of genetic variation is currently emerging. As described, the question if the major disease liability is carried by common genetic variation or by rare variants unique to specific families or individuals has been debated [202, 204]. With the emergence of large-scale sequencing and densely imputation-based studies, we are obtaining growing empirical knowledge of the relative contribution of these types of variation. The current picture in large populations, such as the European, is clearly that the major part of the genetic risk of metabolic traits as T2D and obesity is explained by common genetic variants [47, 92, 97, 98, 216, 266]. While specific rare variants have been shown to change risk in small subsets of the population [137, 211], these variants do not contribute much to the genetic liability on the population scale. For instance, in the latest GWAS of T2D the identified common variants (MAF>5%, *n*=323) explained 16.3% of phenotypic variance, while the identified low frequent and rare variants (MAF<5%, *n*=80) only explained 1.1% of variance [47]. Still, rare alleles contributing to risk of T2D and obesity are being increasingly revealed with the growing sample sizes in studies applying sequencing or densely imputed genotyping [67, 142, 216, 241]. Since rare variants with higher effect size can provide valuable and rapid biological and clinical insights, these findings are of huge interest. Another longstanding debate relates to the extent of genetic heterogeneity in complex traits and to the effect sizes imposed by individual variants. The emerging picture for both T2D and BMI shows an extremely high degree of genetic heterogeneity with hundreds of proven risk variants and indications of the existence of many more common variants with low impact [46, 47, 76, 92, 95, 97, 98, 216, 266].

For risk of T2D and variation in BMI, GWAS have, as reviewed in Part 1, identified a plethora of associated genetic variation in the European population. The relationships between the MAF and effect size inflicted on risk of T2D or variation in BMI, respectively, for these variants are illustrated in Figure 20. In the European population (blue dots), the relationship between MAF and effect size lies mainly on hyperbolic curves with increasing effect sizes for lower allele frequencies. The curves almost co-localise with the results of statistical power analysis of study samples of 100,000 to 150,000 individuals for BMI and T2D, respectively (Fig. 20). Above this curve are alleles with higher effect size than expected from their frequency; however, relatively few of these seem to exist in large populations, with the obvious examples being the *TCF7L2* locus for T2D and the *FTO* locus for BMI. It is expected that many more yet undetected variants lie below the curve; however, limitations in statistical power currently impede their discovery.

The genetic architecture of common complex traits such as T2D and BMI emerging over the past decade is putting empirical data and evidence on years of discussions on the genetic model involved in the genetic predisposition but producing results from densely imputed genotyping and from sequencing-based studies. These findings point to a genetic model in which a very high number of risk alleles co-exist in a complex highly polygenic model (Fig. 20). This is also illustrated by findings of a whole genome sequencing study of T2D [216]. Both empirical and simulated data from this study suggested that low-frequency and rare variants contribute much less to T2D heritability than do common variants [47, 216]. The observed polygenic model is in contrast to earlier theories, which predicted a limited number of risk alleles including "major genes" with higher effects. However, a paper proposing an omnigenic model has recently challenged the polygenic model. In the omnigenic model, it is proposed that all variants affecting gene expression in disease-relevant cells contribute to the genetic susceptibility of a specific trait implying that a substantial fraction of all genes

contributes to the genetic predisposition to any given trait [327]. The omnigenic model would help to explain the high complexity of genetic architecture, which is currently emerging. Of interest, this model has resemblance to the infinitesimal model, which has been discussed for more than 100 years. In the infinitesimal model, a huge number or potentially all variants, have a non-zero but small role in phenotypic variation [328-331]. Again, these theories mirror recent development in studies of polygenic risk scores, which in some studies includes more than 6 million variants with non-zero effects [81]. The omnigenic model distinguishes between core genes with a biologically interpretable role in disease and other non-core genes; however, any gene may be just a few steps from a core gene and hence have a tiny but non-zero effect on the outcome. The model points to a thus far unseen complexity in the genetic architecture of complex traits and opens for some fundamental questions. One such question relates to the overall ambition of using the genetic mapping to gain knowledge on the biology behind specific diseases. What are we expected to learn about the specific disease biology if thousands of genes are involved in the genetic predisposition? Here the ambition may be to map and characterise the core genes specifically involved in disease. Since the core genes are expected to have the largest effects among thousands of genetic associations, it also implies that at some lower threshold of effect size, further expansion of sample sizes to find even smaller effects will not add much to the biological understanding. However, also tissue- or cell-specific effects and differences in effects through different time windows in life will complicate the simple distinction between core genes and non-core genes and may lead to new important playing fields for genetic research.

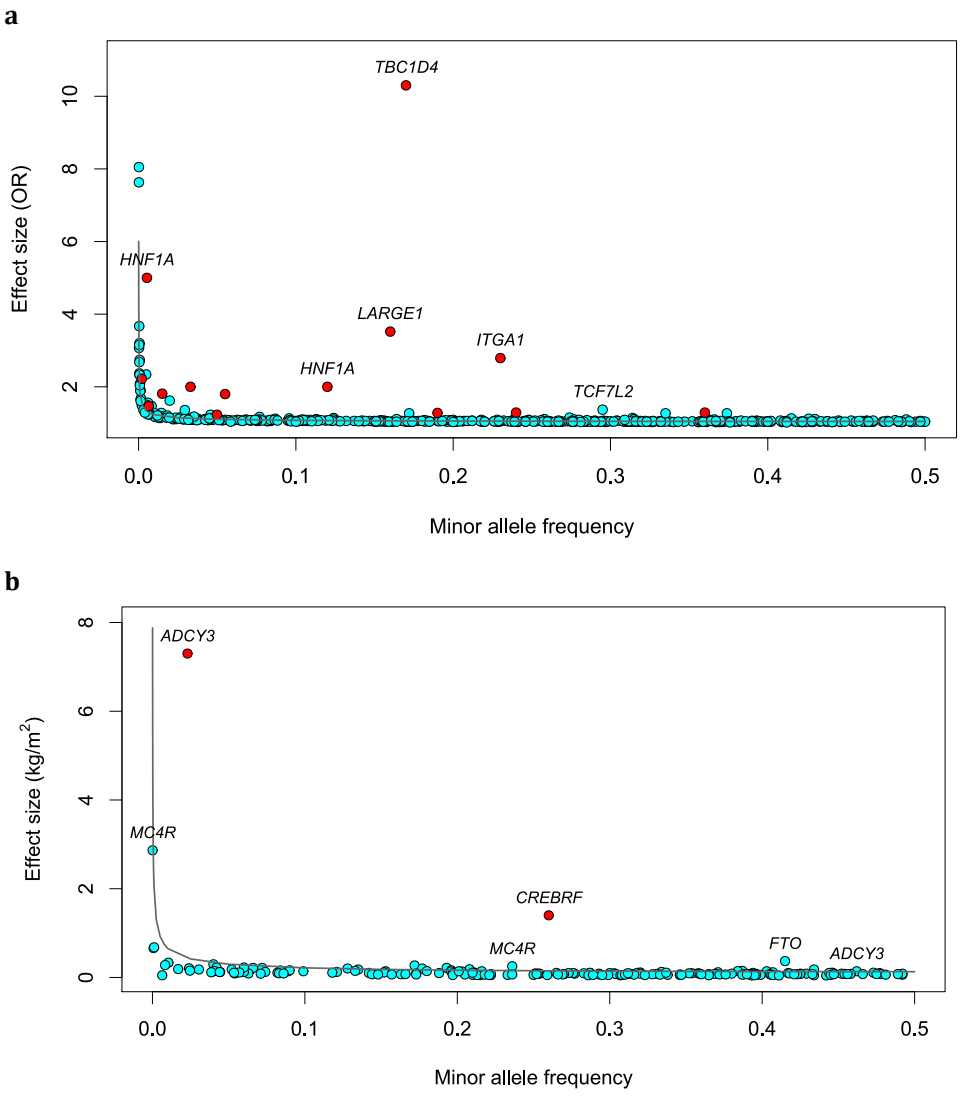

Figure 20. Relationship between minor allele frequency and effect size for variants associated with a) T2D or b) BMI in the European (blue) and isolated (red) populations. The relationship between MAF and effect size resulting in 80% statistical power is shown as the blue line. The statistical power analyses were based on simulations for sample sizes of 50,000 T2D cases and 100,000 controls for T2D and 100,000 individuals for quantitative analysis of BMI assuming a genome-wide significance level of 5×10^{-8} and a T2D prevalence of 8%. Association data for genetic variants were obtained from recent GWAS and other relevant publications [7-9, 47, 92, 170, 332-338].

The genetic architecture may vary between different phenotypes [339] (Fig. 20) and it may also be substantially different in other populations with specific population history and circumstances such as an isolated population. This is due the factors, such as population size and the influence of selection and genetic drift in an isolated population especially when living with stronger environmental pressure, as described in more detail in Part 4. The emerging picture of the genetic architecture of T2D and BMI in isolated populations thus appears rather different from in large open populations. Here rather common variants with frequencies above 1-2% have much stronger effects than seen in large populations as the European population. Among examples of these specific T2D-associated variants in isolated populations are the variants in *TBC1D4*, *ITGA1* and *LARGE1* [7, 8], but findings also include a number of variants identified in other isolated populations than the Greenlandic Inuit population. In the Mexican population, a number of studies have identified variants associated with T2D. In 2014, a missense variant in *HNF1A* (p.Glu508Lys) associated with T2D with an OR of 5 [332]. This variant was had a frequency of 2.1% in patients with T2D as compared to 0.4% in individuals without diabetes. As discussed in the next section, *HNF1A* is known for being the genetic cause of maturity-onset diabetes of the young (MODY) 3 and in addition, common variants in this locus are associated with common T2D. Furthermore, variants in *SLC16A11* and *IGF2* have been associated with T2D in Mexicans [333, 334]. While being more common in the Mexican population (MAF 24% and 19%, respectively), these variants also have more modest effect sizes (OR 1.29 and 1.28, respectively). Other interesting associations in isolated populations include a missense variant in *ABCC8* (p.Arg1420His) in the Pima Indian population (MAF 3.3%, OR 2.0) [335], a *HNF1A* missense variant (p.Gly319Ser) in the Canadian Oji-Cree population [336] and a nonsense deletion in *LIPE* associated with T2D (MAF 5.1%, OR 1.8) in the Older Amish population [337]. In Iceland, there is a large agreement with associations in Europeans, yet some rarer potentially population-specific variants in *PDX1*, *PAM* and *CCND2* have also been demonstrated [170].

Also, for BMI, some specific associations have been found in isolated populations; among them, the *ADCY3* loss-of-function [9]. Of interest, a common missense variant in *CREBRF* (p.Arg457Gln, MAF 26%) was associated with a 1.4 kg/m^2 higher BMI in carriers in a Samoan population [338]. For all these described variants associated with T2D or BMI in isolated populations, it is the case that the identified variant is absent or segregating at much lower frequency in large populations such as the European or Asian populations. Furthermore, for other metabolic traits several findings have been done in isolated populations in subsets of the Greek population [340, 341] or in Finns [342].

Bridging the monogenic to the polygenic genetic landscape of type 2 diabetes in specific hotspot loci

In general, the relationship between MAF and effect size for T2D-associated alleles (Fig. 20) shows that in a European population there is a tight boundary to the effect size for any given MAF. It also illustrates the emerging overall continuum of effects from very common variants via rare alleles segregating in the population towards the family-specific variation shown to cause monogenic forms of diabetes. In addition, when zooming in on specific loci, there are some interesting findings showing the continuity of effects at certain loci with allelic series of variants associated with diabetes.

There is a rather large overlap in loci between common variants associated with T2D and rare variants associated with monogenic diabetes such as maturity-onset diabetes of the young (MODY) or permanent neonatal diabetes mellitus, which goes for the *GCK, HNF1A*, *HNF1B, HNF4A*, *PPARG, KCNJ11, GLIS3* and *WFS1* loci [50, 75, 343-351]. This is equally true for a number of loci associated both BMI and with monogenic subsets of obesity (*MC4R*, *POMC, LEPR, BDNF, SH2B1, PCSK1* and *NTRK2*) [91, 92, 316, 352-357]. Elucidating an allelic series of variants in a locus associated with different severity and manifestations of the same phenotype can be important for several reasons. Finding both rare and common variants associated with comparable phenotypes in the same gene leads to a solid causal connection between the specific gene and disease pathogenesis. Such findings may also shed light on the pathophysiological connections between aspects of disease processes shared by these phenotypes.

A prime example of such a locus is *HNF1A*. *HNF1A* encodes a master transcription factor, which regulates genes expressed in liver, kidney and pancreas [358, 359]. Rare family-specific variants in this locus have long been known to cause MODY inflicting an insulin secretion deficiency [347, 360]. Because of its role in

MODY, this gene was an early candidate gene for common T2D and a missense variant was in 1999 associated with T2D in the Canadian Oji-Cree population [336] (Fig. 20). The *HNF1A* locus has since been firmly established as an risk locus for common T2D in Europeans [75] and in the latest GWAS of T2D, six distinct signals were found in the locus with p.Gly226Ala (rs56348580) as primary lead variant [47]. Furthermore, a rare missense variant in *HNF1A* has been shown to have a high impact on T2D in the Mexican population [332]. Several interesting aspects of possible clinical value stems from these observations. First, generally it is possible to treat patients with *HNF1A*-MODY with sulphonylurea class drugs instead of administrating exogenous insulin [361]. This makes it extremely important to genetically identify HNF1A-MODY patients, since they are often clinically misclassified. It is currently unknown whether the clinical implications observed in HNF1A-MODY patients also applies to patients with rare high-impact but not fully penetrant variants in *HNF1A*. A recent study of carriers of the *HNF1A* p.Glu508Lys variant found in the Mexican population indicated that clinical use of sulfonylureas seen in HNF1A-MODY patients does not translate to carriers of this variant [362]. This finding indicates that the clinical implications need to be proven for each specific class of associated variants or potentially for every specific variant, making the clinical translation extremely tedious or impossible. Another implication of such an allelic series of associations for a specific gene is that it highlights the importance but also the difficulties in assigning the proper molecular diagnosis for variants, which have not been intensively studied previously. As most families with HNF1A-MODY have family-specific variants, it can be extremely difficult to call a variant as disease causing in a suspected MODY patient. Here large-scale and systematic functional screening of potential mutations in important genes, such as previously done for the T2D-associated *PPARG* gene [363], will improve the interpretation. However, since functional in vitro screening is not a perfect tool, the results will need to be interpreted together with other clinical and molecular evidence. Building such algorithms for prediction of causative effects of specific variants is of high clinical importance [364] and is, together with increased knowledge on how to clinically care for patients with high-impact but non-monogenic variants, a major objective for future translational genetic research.

Within obesity, the major example of a gene involved in both common forms of obesity and monogenic subsets is *MC4R*. Here coding variants are a relatively frequent cause of familiar obesity [365] and common variants were associated with BMI in one of the early GWAS [91]. Recently, also rare coding missense variants have been implicated in variation in BMI at the population level showing higher effect than for the common variant [67] (Fig. 20).

Precision medicine in type 2 diabetes – a future prospect or an illusion?

In T2D, there is substantial heterogeneity in the response to medical treatment. One of the hopes of shedding light on complex trait genetics has been to be able to identify subgroups of patients with more similar disease pathogenesis and to use such subgrouping to direct specific treatments for groups of patients achieving higher efficacy and fewer side effects. As such, knowing the specific genetic risk factors would pave the way for personalised or stratified medicine such that prevention, diagnosis, treatment and prognosis would be informed by the genetic profile of the individual. Lessons from monogenic metabolic disease suggest that identification of genetically homogenous groups may lead to improvements in treatment of the patient [366]. Here genetic diagnosis has changed the clinical care of many patients with either permanent neonatal diabetes mellitus or MODY. This has for instance led to discontinuing of treatment for patients with mutations in *GCK* or change of treatment to drugs of the sulphonylurea class in MODY patients with mutations in *HNF1A* and patients with permanent neonatal diabetes mellitus with mutations in *KCNJ11* or *ABCC8* [349, 360, 366-368].

A number of mostly smaller studies have been performed seeking to associate common genetic variation with specific existing disease treatments in T2D, mostly metformin and sulphonylurea but also other classes of T2D medication, either by analysing validated T2D risk variants or by applying GWAS. As reviewed in 2014, there is some evidence for pharmacogenetic effects of the commonly used T2D treatments on immediate treatment outcome, mostly evaluated by the HbA_{1C} levels, but studies of sufficient size are lacking [369]. Of interest, in 2016 a GWAS investigated genetic variation in relation to response to metformin, which is the first-line treatment in T2D, measured by HbA_{1C} levels [370]. Here an intron variant

of *SLC2A2*, which encodes the glucose transporter, GLUT2, was associated with a 0.17% greater metformin-induced reduction in HbA_{1C} in 10,577 European participants [370]. The clinical impact of this association has not yet been fully investigated, but it is likely that genotyping results for such a variant in T2D patients may in the future be included in individualised treatment choice algorithms. This is especially the case in a future clinical scenario where full genotype or sequence data are available for all patients to inform clinical decision-making so that such variants do not need to be specifically selected and genotyped. In general, findings from pharmacogenetic studies have not been as solid as to offer real clinical value at present. However, a growing interest for the field of genetics in the pharmaceutical industry, an increasing tendency to make clinical trial data available for other researchers and the advent of large biobanks with register-based prescription and outcome information may change this within the coming years. From current findings, however, it may questionable if such genetic variants with a validated pharmacogenetic impact may be sufficiently strong to predict outcome of specific pharmacological treatment, but may add to prediction of effect coming from clinical disease-related data.

The emerging genetic landscape of T2D and metabolic diseases in large populations as the European and in isolated populations as the Greenlandic have huge consequences for the future prospects for the utility of genetic discoveries to improve clinical care of patients. In the Greenlandic population, the large effect size of identified genetic variants [7-9], may make a shortcut to precision medicine as single variants define specific subgroups of individuals and patients. The most evident case is for homozygous carriers of the *TBC1D4* p.Arg684Ter variant where future studies may well bring evidence for specific treatment and prognosis prediction in these individuals.

In large populations such as the European population, there is no obvious evidence for subgrouping and precision medicine based on genetic risk factors for the bulk of T2D patients, but there are several indications that genetics may be useful for smaller specific subsets of individuals. Such subsets may be patients with MODY, which accounts for ~1% of patients with T2D [371] or potentially individuals with rare variants with higher impact in genes such as *HNF1A* or *SLC30A8*, which may in due time show to be of clinical importance. In the individual, the specific pattern of hundreds or more likely thousands of genetic risk alleles, acts in concert with environmental exposures to determine individual risk of disease. As discussed, common genetic variants individually unanimously impose modest changes in the risk of T2D and adiposity. Although genome-wide evaluation indicates that the full set of variants in well-imputed genotype data explain a substantial amount of heritability [47, 100, 199], the high complexity and genetic heterogeneity makes genetic prediction of T2D or obesity and genotype-specific interventions extremely difficult.

In the process of determining the specific genetic vulnerabilities increasing risk of T2D, important lessons are being learnt about the pathogenic constitution of the disease. Although subgrouping of individuals based on major disease aetiologies is an obvious goal in order to achieve clinical translation, the genetic risk composition and the complexity seen in biological disentangling of these genomic loci, display the complexity and heterogeneity of T2D. The consequence of this is that it is overly optimistic to believe that the disease can be broken into a limited number of disease entities for which specific clinical management can be installed. Instead, as suggested in a recent commentary [372], the pathogenesis in T2D can rather be seen as liabilities and defects on a quantitative scale related to the major pathophysiological processes that contribute to diabetes risk and progression. For a minor subgroup of patients, one or few defects will dominate the set of liabilities and for such patients targeted clinical intervention may be possible. However, for the majority of patients, a number of parallel liabilities with modest impact on multiple disease processes are the root of the disease making the pathogenesis highly heterogenic [372]. Identifying and specifically targeting the multiple defects and processes are huge challenges for reaching precision medicine in the bulk of patients with T2D. Hence, it will be natural to start the hunt for precision medicine in T2D by seeking to identify the smaller subsets of patients with rather uniform disease processes. With the outlook of genetics, this may be done identifying subsets with high impact variants as seen in isolated populations [7] or by identifying patients with a combined load of specific variants. This combined impact could likely be alleles with an impact on insulin secretion, a group of patients which is imaginable to be able to target with other clinical approaches than the metformin treatment used as first line standard drug [373]. In addition, novel developments give some hopes to be able to use GWAS-based findings to get closer to precision medicine in T2D. First, the application of polygenic risk scores may, as described in Part 1, be clinically applicable to

pinpoint subsets at particularly high risk of disease [47, 77, 81, 374] and future translational studies may depict if these individuals will benefit from specific preventive initiatives or treatment. Here of special interest is the construction of polygenic risk scores for T2D-related endophenotypes since individuals at the extreme high end of such scores may have a more uniform T2D pathogenesis. Second, recent studies applying clustering of T2D-associated variants have indicated that discrete groups of variants can be formed and that up to 30% of T2D patients are in top 10% of the genetic risk score based on variants from a single cluster [143]. These findings support the use of genetics to deconstruct T2D heterogeneity at least in a subset of patients with a dominant disease process, as discussed above. These findings indicate that classification of patients by these genetic pathways may offer a step toward genetically informed T2D patient management.

The ability to individualise treatment based on genetic information will likely prove to be a continuum from monogenic subsets with highly individualised treatment, over genetic identification of pathophysiological specific subgroups of patients with rare but high-impact variants or accumulation of common risk alleles to a large group of highly heterogeneous patients in which knowledge of genetic profile will not add significantly to clinical care. Yet, with the ever-falling costs of genome genotyping and sequencing, the future will likely bring a situation where all patients have their full genome sequenced thereby opening for large-scale and more accurate real-life studies of genetic impact on treatment outcome in different strata of patients.

To get to precision medicine in the future it seems clear that more than genetic data are needed to be taken into consideration. A recent study took a step away from the omics techniques and used six rather basic clinical and biochemical phenotypes and applied clustering methods to seek to subgroup newly diagnosed T2D patients [375]. These analyses identified five distinct subtypes of T2D, which were characterised by certain phenotypic features assessed soon after diagnosis and possibly representing broad disease-related pathophysiological processes. These included early-onset severe autoimmune diabetes (6.7% of patients), severe insulin-deficient diabetes (17.5% of patients), severe insulin-resistant diabetes (15.3%), mild obesity-related diabetes (21.6%) and mild age-related diabetes (39.1%). Rates of diabetes-related complications differed between the subgroups [375]. While being of huge interest, one limitation of this approach resides in the use of phenotypic criteria for clustering that are ascertained at disease onset and are not necessarily generalizable to other stages of disease development. This is in contrast to a genetically driven strategy. As such, the clinical translation of these findings remains unclear; however, the study does provide a steppingstone on which to build future subgrouping strategies.

Other emerging sources of data possibly driving future T2D subgrouping with the potential to change clinical care come from other types of omics data. Such data, for instance transcriptomics, proteomics or metabolomics data sources, typically lie on the biological pathway between genetic markers and clinical phenotypes and hence are closer to phenotype. While this may provide greater precision to predict phenotype, for instance as seen in a recent paper based on data from the Framingham study [376], it also leads to more complex interpretation as results are related to the specific cells, tissues or organs, to time-dependent and context-dependent effects and to classical epidemiological challenges, such as confounding and reverse causation. Hence, as for the study by Ahlqvist et al. [375], it remains crucially important that findings represent reflections of true disease aetiology and are not merely different stages in disease progression. Yet, if achieved, the challenges may turn to be opportunities since offering dynamic tracking of disease progression via omics data.

Emerging biological translation of genetic findings

As reviewed, research during the past decade has revealed a large number of specific genetic variants associated with obesity or T2D. An obvious next goal is to aim to shed light on the biological mechanism behind the associated variants. Here genetics offer the possibility to test mechanistic relationship between gene variation and human phenotype directly at the cellular level. However, in general it has proven very difficult to elucidate the paths from associated marker to distal phenotype. A number of obstacles are making this process extremely cumbersome. First, finding the causal variant at a locus will be the first step to find the mechanisms where for some loci more than one functional variant will exist. However, since most of the identified signals map to noncoding regions of the genome, the identity of the genes through which they operate is often obscure [377]. Even for coding variants,

the translation from associated variant to causal variant may not be straightforward. Although association of rare coding variants may point directly to the causal variant and the causal transcript, yet, as discussed in Part 3, even for common coding variants it may not be straightforward to assign the functional variant. However, for predicted loss-of-function variants, as identified in papers 7 and 9, the evidence for functional variants and genes may be much easier to obtain.

Identifying the causal variant and the effector transcript in a locus is thus a quite huge task. At some loci, an obvious biological candidate will be present, but for most loci, this will not be the case. Non-coding disease-associated variants may affect gene expression through effects on transcription, splicing, or mRNA stability [378]. Statistical fine-mapping and genomics integration with publicly available functional genomics data, such as ENCODE or Roadmap Epigenomics data [379, 380], to investigate eQTLs in relevant tissues [381], chromatin accessibility or transcription factor binding, may help defining causal variants and transcripts [382, 383]. Even when a causative gene has been identified for a locus, there are still issues in functionally connecting the causal gene with the distal phenotype. Novel high-throughput techniques, such as 4C, HiC and ATAC-seq that can map chromatin interactions and accessibility [384], together with technologies, such as CRISPR-Cas9 genome editing tools [385] and single-cell RNA sequencing studies of how RNA expression differs between cells within tissues [386], provide drastically changed opportunities for disentangling the molecular functional impact of GWAS-identified genes.

Metabolic genetics in drug development

Another perspective of performing research into the specific genetic contribution to the risk of diseases as T2D is that these genetic discoveries will shed light on novel biological targets that can be manipulated to serve as future drug targets to treat the disease. The pharmaceutical industry is struggling to keep up efficiency in discovering and evaluating drug targets and the average cost of novel compounds is increasing [387]. Here genetic research can bring a validation of targets from discovery of previously unknown potential targets in the disease and by genetic association evaluate efficacy, pleiotropy and possible side effects. The identification and validation of novel drug targets is a major stated objective of genetic studies of T2D and obesity. It has been estimated that drug mechanisms with genetic support can succeed twice as often in the clinic as those without such support [388]. Increasing developmental success rates will result in fewer drug candidate failures for each successful drug and thereby a more cost-effective drug development pipeline. Even relatively modest improvements in success rates achieved through better target and indication selection based on genetic evidence, could have a substantial impact on productivity [389].

A number of important proof-of-principles especially related to development of lipid-lowering drugs have been achieved through discoveries of *PCSK9*, *APOC3* and *ANGPTL3* loci [233, 390] and medicine being developed based on genetic identification. Especially the *PCSK9* locus is a successful example where the identification of both gain-of-function variants causing familial hypercholesterolemia [391] and subsequently loss-of-function alleles associated with decreased LDL-cholesterol and protection against CVD in the general population [210, 392]. These findings stimulated the development of antibody-based inhibitors targeting PSCK9, which were approved for the market in 2015 [393, 394]. These drugs are a novel class of cholesterol lowering agents and are used as alternatives or adjuvants to treatment with statins. Although no such solid examples exist within T2D or obesity, there are indications that similar translational examples could emerge. For instance, as described in Part 2, three T2D susceptibility variants map to genes encoding drug targets for existing T2D drugs (*PPARG*, *KCNJ11* and *ABCC8*) and genetic studies after drug development have also provided at least some additional evidence related to GLP1 receptor agonists [395, 396]. With excellent examples coming from lipid research, one may speculate why no such findings and translations have been achieved in T2D or obesity. One reason for this may be the higher complexity in the pathogenesis of T2D and obesity compared to the efforts in treatment of dyslipidaemia. As such, targeting pancreatic islet, intestinal cells, adipocytes, skeletal muscle or brain cells may thus be more complicated than reducing the levels of lipids in the blood both in terms of achieving efficacy and to avoid side-effects. However, pending findings may hold promise for future application of genetics to inspire and inform drug development and a future example could be related to *SLC30A8* [211]. Variants with protective effect on disease may pose advantages for the development of drugs. In addition, identification of an allelic series of both gain-of-function and loss-of-function variants in relation to the

phenotype strengthens the knowledge being translated to modification of the target, as also seen for the *PCSK9* locus [397]. In addition, findings from studies included in this dissertation may in the future lead to development of drugs. For instance, major findings from the Greenlandic Inuit population [7, 9] may in the future lead to T2D medication targeting insulin resistance through TBC1D4 and drugs against obesity based on biology of ADCY3. In addition, the coding variant in *CD300LG* associated with triglyceride and HDL-cholesterol [5] may in turn into novel lipid treatments, although the lack of biological knowledge for this association is a complicating factor.

The future of genetic research in type 2 diabetes and complex disease

With the seemingly ever-falling prizes for sequencing and genotyping, a huge number of large initiatives are currently under way generating genetic data at an unprecedented scale. The UK Biobank is currently the largest biobank with ~500,000 participants in whom genome-wide genotyping data are available [66] combined with rather detailed phenotype characterisation. Within a few years, these samples will also be whole exome sequenced to gain insights into rare coding variants. Other large initiatives include the Genomics England Project aiming at sequencing 100,000 genomes of which ~70,000 are currently done [398], the US Million Veterans Program of 400,000 individuals [399] and the Chinese Kadoorie Biobank with ~500,000 participants [400]. On top of these are a number of national healthcare-related sequencing initiatives. The increasing willingness to form academia-industry partnerships and the increased tendency for public availability of results and data sharing, for instance through the Genetics Portals initiated through the Accelerating Medicines Partnership, will in the coming years drastically increase the quantity and depth of data available for researchers. While is naturally important to increase sample sizes since genetic effects are modest at best, it is also important to consider improved clinical and molecular phenotyping of the individuals. Many cohorts include at least subsets with rather deep phenotypes and a lot of them have the opportunity for register-based outcome analysis at follow-up which can prove to be a great strength [401]. Ideally, such huge collections of biobank samples can, besides register-based outcome data, also be merged with clinical data from digital hospital records opening for large-scale clinical genomics studies. These developments will in time probably lead to standard generation of detailed genetic data for every patient to be used when clinically relevant.

While SNPs have been at the centre of all disease trait mapping described in this dissertation, a range of other types of variation also exist. Systematic investigations of other types of genetic variation, for instance deletions, insertions and structural variations, remain relatively sparse and only a few major findings have been done [402, 403]. Large scale sequencing initiatives will probably make it possible to directly assess the role of these types of variation in metabolic disease, however they also need to be included in future large-scale imputation panels, which is not the case for the currently frequently used HRC panel [63]. Another uncertainty in genetic research is the extent to which interactions between genetic variation and environmental factors are important in the disease pathogenesis. While such interactions may seem logical, very few interactions have been firmly demonstrated in metabolic disease [404], with the interaction between *FTO* variants and physical activity on BMI being the most validated in metabolic traits [405, 406]. However, two recent genome-wide interaction studies of more than 200,000 individuals did identify novel interactions between genetic variants and physical activity or smoking on BMI [407, 408]; indicating that applying huge sample sizes may start to reveal genuine interaction effects.

Furthermore, the future will see other kinds of omics data, such as tissue transcriptomics and single-cell sequencing, proteomics, metabolomics data, being increasingly generated on a large epidemiological scale and integrated with genetic data. In addition, data from wearable devices that monitor behaviours, exposures and digital imaging technologies will in the coming years contribute the collection of big data. All of this has the potential to substantially improve the prediction, prevention and treatment of T2D starting with patients carrying high impact variants or with high polygenic risk related to specific disease processes. However, developments in the analytical approaches seem to be needed the get the full benefit of a more holistic and integrative system medicines approach taking [409].

CONCLUSIONS

The past decade has elucidated a large number of genomic loci associated with T2D and obesity. These loci are generally common and inflict only modest risk increments on disease. Evidence points to large number of risk variants below statistical significance, displaying the high degree of polygenicity of T2D, obesity and complex traits in general. These features have huge implications for precision medicine in metabolic diseases. Further exposures of the mechanistic insights are leading to translation of association signals into clinical useful knowledge especially in development of novel drugs, but this process has proven difficult and tedious.

Work included in this dissertation has contributed to this development in several ways. First, identification of T2D genetic variants associated with intermediary phenotypes have shown that the major T2D-related intermediary phenotype is decreased insulin secretion. Second, sequencing-based studies included in the dissertation has proved to be a starting point for future in-depth characterisation of low-frequency and rare variants in complex traits. Third, major studies of the Greenlandic Inuit population have identified a number of biologically and possibly clinically important associations of functional variants and thereby shown the potentials of investigating the genetic contribution to metabolic traits in such a population.

In the future, studies of mega-size cohorts and biobanks with deeply phenotyped samples coupled to clinical information will probably disentangle more clinically relevant omics-related knowledge with an impact on future precision medicine. Furthermore, an increased focus on translational and mechanistic research to make sense of the hundreds of loci associated T2D and obesity will inevitably reveal biological knowledge of huge importance for disease understanding and future drug development.

REFERENCES

[1] Grarup N, Rose CS, Andersson EA, et al. (2007) 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 56: 3105-3111

[2] Grarup N, Andersen G, Krarup NT, et al. (2008) 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 57: 2534-2540

[3] Boesgaard TW, Grarup N, Jørgensen T, et al. (2010) Variants at DGKB/TMEM195, ADRA2A, GLIS3 and C2CD4B loci are associated with reduced glucose-stimulated beta cell function in middle-aged Danish people. Diabetologia 53: 1647-1655

[4] Grarup N, Overvad M, Sparsø T, et al. (2011) The diabetogenic VPS13C/C2CD4A/C2CD4B rs7172432 variant impairs glucose-stimulated insulin response in 5,722 non-diabetic Danish individuals. Diabetologia 54: 789-794

[5] Albrechtsen A, Grarup N, Li Y, et al. (2013) Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. Diabetologia 56: 298-310

[6] Grarup N, Sulem P, Sandholt CH, et al. (2013) Genetic architecture of vitamin B12 and folate levels uncovered applying deeply sequenced large datasets. PLoS Genet 9: e1003530

[7] Moltke I, Grarup N, Jørgensen ME, et al. (2014) A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. Nature 512: 190-193

[8] Grarup N, Moltke I, Andersen MK, et al. (2018) Identification of novel high-impact recessively inherited type 2 diabetes risk variants in the Greenlandic population. Diabetologia 61: 2005-2015

[9] Grarup N, Moltke I, Andersen MK, et al. (2018) Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. Nat Genet 50: 172-174

[10] IDF Diabetes Atlas IDF Diabetes Atlas, Sixth Edition. Available from http://www.idf.org/diabetesatlas, accessed 06-01 2015

[11] Hu FB (2011) Globalization of diabetes: the role of diet, lifestyle, and genes. Diabetes Care 34: 1249-1257

[12] Defronzo RA (2009) Banting Lecture. From the triumvirate to the ominous octet: a new paradigm for the treatment of type 2 diabetes mellitus. Diabetes 58: 773-795

[13] DeFronzo RA, Ferrannini E, Groop L, et al. (2015) Type 2 diabetes mellitus. Nat Rev Dis Primers 1: 15019

[14] Ferrannini E (1998) Insulin resistance versus insulin deficiency in non-insulin-dependent diabetes mellitus: problems and prospects. Endocr Rev 19: 477-490

[15] Lillioja S, Mott DM, Spraul M, et al. (1993) Insulin resistance and insulin secretory dysfunction as precursors of non-insulin-dependent diabetes mellitus. Prospective studies of Pima Indians. N Engl J Med 329: 1988-1992

[16] Haffner SM, Miettinen H, Gaskill SP, Stern MP (1995) Decreased insulin secretion and increased insulin resistance are independently related to the 7-year risk of NIDDM in Mexican-Americans. Diabetes 44: 1386-1391

[17] Ferrannini E, Mari A (2004) Beta cell function and its relation to insulin action in humans: a critical appraisal. Diabetologia 47: 943-956

[18] Bergman RN, Ader M, Huecking K, Van Citters G (2002) Accurate assessment of beta-cell function: the hyperbolic correction. Diabetes 51 Suppl 1: S212-220

[19] Bergman RN, Phillips LS, Cobelli C (1981) Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose. J Clin Invest 68: 1456-1467

[20] Park Y, Zhu S, Palaniappan L, Heshka S, Carnethon MR, Heymsfield SB (2003) The metabolic syndrome: Prevalence and associated risk factor findings in the us population from the third national health and nutrition examination survey, 1988-1994. Arch Intern Med 163: 427-436

[21] Maier W, Holle R, Hunger M, et al. (2013) The impact of regional deprivation and individual socio-economic status on the prevalence of Type 2 diabetes in Germany. A pooled analysis of five population-based studies. Diabet Med 30: e78-e86

[22] Rathmann W, Scheidt-Nave C, Roden M, Herder C (2013) Type 2 diabetes: prevalence and relevance of genetic and acquired factors for its prediction. Dtsch Arztebl Int 110: 331-337

[23] Speakman JR, O'Rahilly S (2012) Fat: an evolving issue. Dis Model Mech 5: 569-573

[24] Matsuda M, DeFronzo RA (1999) Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. Diabetes Care 22: 1462-1470

[25] Sluiter WJ, Erkelens DW, Reitsma WD, Doorenbos H (1976) Glucose tolerance and insulin release, a mathematical approach I. Assay of the beta-cell response after oral glucose loading. Diabetes 25: 241-244

[26] Meigs JB, Cupples LA, Wilson PW (2000) Parental transmission of type 2 diabetes: the Framingham Offspring Study. Diabetes 49: 2201-2207

[27] Newman B, Selby JV, King MC, Slemenda C, Fabsitz R, Friedman GD (1987) Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. Diabetologia 30: 763-768

[28] Poulsen P, Ohm-Kyvik K, Vaag A, Beck-Nielsen H (1999) Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance - A population-based twin study. Diabetologia 42: 139-145

[29] Willemsen G, Ward KJ, Bell CG, et al. (2015) The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium. Twin Res Hum Genet 18: 762-771

[30] Elbein SC, Hasstedt SJ, Wegner K, Kahn SE (1999) Heritability of pancreatic beta-cell function among nondiabetic members of Caucasian familial type 2 diabetic kindreds. J Clin Endocrinol Metab 84: 1398-1403

[31] Gjesing AP, Hornbak M, Allin KH, et al. (2014) High heritability and genetic correlation of intravenous glucose- and tolbutamide-induced insulin secretion among non-diabetic family members of type 2 diabetic patients. Diabetologia 57: 1173-1181

[32] Vaag A, Henriksen JE, Madsbad S, Holm N, Beck-Nielsen H (1995) Insulin secretion, insulin action, and hepatic glucose production in identical twins discordant for non-insulin-dependent diabetes mellitus. J Clin Invest 95: 690-698

[33] Hemani G, Yang J, Vinkhuyzen A, et al. (2013) Inference of the genetic architecture underlying BMI and height with the use of 20,240 sibling pairs. Am J Hum Genet 93: 865-875

[34] Stunkard AJ, Harris JR, Pedersen NL, McClearn GE (1990) The body-mass index of twins who have been reared apart. N Engl J Med 322: 1483-1487

[35] Vogler GP, Sorensen TI, Stunkard AJ, Srinivasan MR, Rao DC (1995) Influences of genes and shared family environment on adult body mass index assessed in an adoption study by a comprehensive path model. Int J Obes Relat Metab Disord 19: 40-45

[36] Lander ES, Linton LM, Birren B, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409: 860-921

[37] Sachidanandam R, Weissman D, Schmidt SC, et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409: 928-933

[38] Venter JC, Adams MD, Myers EW, et al. (2001) The sequence of the human genome. Science 291: 1304-1351

[39] Lander ES (2011) Initial impact of the sequencing of the human genome. Nature 470: 187-197

[40] Teare MD, Barrett JH (2005) Genetic linkage studies. Lancet 366: 1036-1044

[41] Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265: 2037-2048

[42] Reynisdottir I, Thorleifsson G, Benediktsson R, et al. (2003) Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34-q35.2. Am J Hum Genet 73: 323-335

[43] Grant SF, Thorleifsson G, Reynisdottir I, et al. (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 38: 320-323

[44] Florez JC, Jablonski KA, Bayley N, et al. (2006) TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. N Engl J Med 355: 241-250

[45] Cauchi S, Froguel P (2008) TCF7L2 genetic defect and type 2 diabetes. Curr Diab Rep 8: 149-155

[46] Mahajan A, Go MJ, Zhang W, et al. (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 46: 234-244

[47] Mahajan A, Taliun D, Thurner M, et al. (2018) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 50: 1505-1513

[48] Deeb SS, Fajas L, Nemoto M, et al. (1998) A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. Nat Genet 20: 284-287

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

[50] Zeggini E, Scott LJ, Saxena R, et al. (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 40: 638-645

[51] Gloyn AL, Weedon MN, Owen KR, et al. (2003) 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 52: 568-572

[52] Nielsen EM, Hansen L, Carstensen B, et al. (2003) The E23K variant of Kir6.2 associates with impaired post-OGTT serum insulin response and increased risk of type 2 diabetes. Diabetes 52: 573-577

[53] Morris AP, Voight BF, Teslovich TM, et al. (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet 44: 981-990

[54] Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55: 997-1004

[55] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38: 904-909

[56] Lander ES (1996) The new genomics: global views of biology. Science 274: 536-539

[57] Reich DE, Lander ES (2001) On the allelic spectrum of human disease. Trends Genet 17: 502-510

[58] Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. Am J Hum Genet 90: 7-24

[59] Pe'er I, Yelensky R, Altshuler D, Daly MJ (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol 32: 381-385

[60] The International HapMap 3 Consortium (2010) Integrating common and rare genetic variation in diverse human populations. Nature 467: 52-58

[61] The Genomes Project Consortium (2015) A global reference for human genetic variation. Nature 526: 68-74

[62] UK10K Consortium, Walter K, Min JL, et al. (2015) The UK10K project identifies rare variants in health and disease. Nature 526: 82-90

[63] McCarthy S, Das S, Kretzschmar W, et al. (2016) A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 48: 1279-1283

[64] Manolio TA (2017) In Retrospect: A decade of shared genomic associations. Nature 546: 360-361

[65] Sudlow C, Gallacher J, Allen N, et al. (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 12: e1001779

[66] Bycroft C, Freeman C, Petkova D, et al. (2018) The UK Biobank resource with deep phenotyping and genomic data. Nature 562: 203-209

[67] Turcot V, Lu Y, Highland HM, et al. (2018) Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. Nat Genet 50: 26-41

[68] Pulit SL, Stoneman C, Morris AP, et al. (2018) Meta-analysis of genome-wide association studies for body fat distribution in 694,649 individuals of European ancestry. Hum Mol Genet 10.1093/hmg/ddy327

[69] Sladek R, Rocheleau G, Rung J, et al. (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445: 881-885

[70] Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. (2007) A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet 39: 770-775

[71] Zeggini E, Weedon MN, Lindgren CM, et al. (2007) Replication of Genome-Wide Association Signals in UK Samples Reveals Risk Loci for Type 2 Diabetes. Science 316: 1336-1341

[72] Diabetes Genetics Initiative of Broad Institute of Harvard and MIT Lund University and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, et al. (2007) Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels. Science 316: 1331-1336

[73] Scott L, Mohlke K, Bonnycastle L, et al. (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316: 1341-1345

[74] The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661-678

[75] Voight BF, Scott LJ, Steinthorsdottir V, et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42: 579-589

[76] Scott RA, Scott LJ, Magi R, et al. (2017) An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. Diabetes 66: 2888-2902

[77] Meigs JB, Shrader P, Sullivan LM, et al. (2008) Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med 359: 2208-2219

[78] Weedon MN, McCarthy MI, Hitman G, et al. (2006) Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. PLoS Med 3: e374

[79] Sparsø T, Grarup N, Andreasen C, et al. (2009) Combined analysis of 19 common validated type 2 diabetes susceptibility gene variants shows moderate discriminative value and no evidence of gene-gene interaction. Diabetologia 52: 1308-1314

[80] Chatterjee N, Shi J, Garcia-Closas M (2016) Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet 17: 392-406

[81] Khera AV, Chaffin M, Aragam KG, et al. (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 50: 1219-1224

[82] Torkamani A, Wineinger NE, Topol EJ (2018) The personal and clinical utility of polygenic risk scores. Nat Rev Genet 19: 581-590

[83] Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P (2004) Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol 159: 882-890

[84] Lyssenko V, Laakso M (2013) Genetic screening for the risk of type 2 diabetes: worthless or valuable? Diabetes Care 36 Suppl 2: S120-126

[85] Khera AV, Emdin CA, Drake I, et al. (2016) Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. N Engl J Med 375: 2349-2358

[86] Knowles JW, Ashley EA (2018) Cardiovascular disease: The rise of the genetic risk score. PLoS Med 15: e1002546

[87] Scott RA, Lagou V, Welch RP, et al. (2012) Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet 44: 991-1005

[88] Manning AK, Hivert MF, Scott RA, et al. (2012) A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet 44: 659-669

[89] Saxena R, Hivert MF, Langenberg C, et al. (2010) Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. Nat Genet 42: 142-148

[90] Frayling TM, Timpson NJ, Weedon MN, et al. (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 316: 889-894

[91] Loos RJ, Lindgren CM, Li S, et al. (2008) Common variants near MC4R are associated with fat mass, weight and risk of obesity. Nat Genet 40: 768-775

[92] Locke AE, Kahali B, Berndt SI, et al. (2015) Genetic studies of body mass index yield new insights for obesity biology. Nature 518: 197-206

[93] Lindgren CM, Heid IM, Randall JC, et al. (2009) Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. PLoS Genet 5: e1000508

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

[95] Shungin D, Winkler TW, Croteau-Chonka DC, et al. (2015) New genetic loci link adipose and insulin biology to body fat distribution. Nature 518: 187-196

[96] Sandholt CH, Grarup N, Pedersen O, Hansen T (2015) Genome-wide association studies of human adiposity: Zooming in on synapses. Mol Cell Endocrinol 418 Pt 2: 90-100

[97] Loos RJ (2018) The genetics of adiposity. Curr Opin Genet Dev 50: 86-95

[98] Yengo L, Sidorenko J, Kemper KE, et al. (2018) Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. Hum Mol Genet 27: 3541-3649

[99] Grarup N, Sandholt CH, Hansen T, Pedersen O (2014) Genetic susceptibility to type 2 diabetes and obesity: from

genome-wide association studies to rare variants and beyond. Diabetologia 57: 1528-1541

[100] Yang J, Bakshi A, Zhu Z, et al. (2015) Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet 47: 1114-1120

[101] Grarup N, Sparsø T, Hansen T (2010) Physiologic characterization of type 2 diabetes-related loci. Curr Diab Rep 10: 485-497

[102] Bulik-Sullivan B, Finucane HK, Anttila V, et al. (2015) An atlas of genetic correlations across human diseases and traits. Nat Genet 47: 1236-1241

[103] Zheng J, Erzurumluoglu AM, Elsworth BL, et al. (2017) LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 33: 272-279

[104] Pascoe L, Tura A, Patel SK, et al. (2007) Common variants of the novel type 2 diabetes genes CDKAL1 and HHEX/IDE are associated with decreased pancreatic beta-cell function. Diabetes 56: 3101-3104

[105] Staiger H, Machicao F, Stefan N, et al. (2007) Polymorphisms within novel risk loci for type 2 diabetes determine beta-cell function. PLoS One 2: e832

[106] Staiger H, Stancakova A, Zilinskaite J, et al. (2008) 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 57: 514-517

[107] Boesgaard TW, Zilinskaite J, Vänttinen M, et al. (2008) The common SLC30A8 Arg325Trp variant is associated with reduced first-phase insulin release in 846 non-diabetic offspring of type 2 diabetes patients--the EUGENE2 study. Diabetologia 51: 816-820

[108] Rung J, Cauchi S, Albrechtsen A, et al. (2009) Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. Nat Genet 41: 1110-1115

[109] Boesgaard TW, Gjesing AP, Grarup N, et al. (2009) 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 4: e7236

[110] Florez JC (2008) Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes? Diabetologia 51: 1100-1110

[111] Haring HU (2016) Novel phenotypes of prediabetes? Diabetologia 59: 1806-1818

[112] DeFronzo RA (2004) Pathogenesis of type 2 diabetes mellitus. Med Clin North Am 88: 787-835

[113] Hannon TS, Kahn SE, Utzschneider KM, et al. (2018) Review of methods for measuring beta-cell function: Design considerations from the Restoring Insulin Secretion (RISE) Consortium. Diabetes Obes Metab 20: 14-24

[114] Ferrannini E, Mari A (1998) How to measure insulin sensitivity. J Hypertens 16: 895-906

[115] Stumvoll M, Fritsche A, Haring H (2001) The OGTT as test for beta cell function? Eur J Clin Invest 31: 380-381

[116] Hansen T, Drivsholm T, Urhammer SA, et al. (2007) The BIGTT test: a novel test for simultaneous measurement of pancreatic β-cell function, insulin sensitivity, and glucose tolerance. Diabetes Care 30: 257-262

[117] Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC (1985) Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia 28: 412-419

[118] Stumvoll M, Van Haeften T, Fritsche A, Gerich J (2001) Oral glucose tolerance test indexes for insulin sensitivity and secretion based on various availabilities of sampling times. Diabetes Care 24: 796-797

[119] Stumvoll M, Mitrakou A, Pimenta W, et al. (2000) Use of the oral glucose tolerance test to assess insulin release and insulin sensitivity. Diabetes Care 23: 295-301

[120] Walford GA, Gustafsson S, Rybin D, et al. (2016) Genome-Wide Association Study of the Modified Stumvoll Insulin Sensitivity Index Identifies BCL2 and FAM19A2 as Novel Insulin Sensitivity Loci. Diabetes 65: 3200-3211

[121] Knowles JW, Xie W, Zhang Z, et al. (2015) Identification and validation of N-acetyltransferase 2 as an insulin sensitivity gene. J Clin Invest 125: 1739-1751

[122] Camporez JP, Wang Y, Faarkrog K, Chukijrungroat N, Petersen KF, Shulman GI (2017) Mechanism by which arylamine N-acetyltransferase 1 ablation causes insulin resistance in mice. Proc Natl Acad Sci U S A 114: E11285-e11292

[123] Chennamsetty I, Coronado M, Contrepois K, et al. (2016) Nat1 Deficiency Is Associated with Mitochondrial Dysfunction and Exercise Intolerance in Mice. Cell Rep 17: 527-540

[124] Prokopenko I, Poon W, Magi R, et al. (2014) A central role for GRB10 in regulation of islet function in man. PLoS Genet 10: e1004235

[125] Wood AR, Jonsson A, Jackson AU, et al. (2017) A Genome-Wide Association Study of IVGTT-Based Measures of First Phase Insulin Secretion Refines the Underlying Physiology of Type 2 Diabetes Variants. Diabetes 66: 2296-2309

[126] Lyssenko V, Lupi R, Marchetti P, et al. (2007) Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. J Clin Invest 117: 2155-2163

[127] Villareal DT, Robertson H, Bell GI, et al. (2010) TCF7L2 variant rs7903146 affects the risk of type 2 diabetes by modulating incretin action. Diabetes 59: 479-485

[128] Schafer SA, Tschritter O, Machicao F, et al. (2007) Impaired glucagon-like peptide-1-induced insulin secretion in carriers of transcription factor 7-like 2 (TCF7L2) gene polymorphisms. Diabetologia 50: 2443-2450

[129] Pilgaard K, Jensen CB, Schou JH, et al. (2009) The T allele of rs7903146 TCF7L2 is associated with impaired insulinotropic action of incretin hormones, reduced 24 h profiles of plasma insulin and glucagon, and increased hepatic glucose production in young healthy men. Diabetologia 52: 1298-1307

[130] Srinivasan S, Kaur V, Chamarthi B, et al. (2018) TCF7L2 Genetic Variation Augments Incretin Resistance and Influences Response to a Sulfonylurea and Metformin: The Study to Understand the Genetics of the Acute Response to Metformin and Glipizide in Humans (SUGAR-MGH). Diabetes Care 41: 554-561

[131] Gjesing AP, Ribel-Madsen R, Harder MN, et al. (2015) Genetic and phenotypic correlations between surrogate

measures of insulin release obtained from OGTT data. Diabetologia 58: 1006-1012

[132] Jonsson A, Ladenvall C, Ahluwalia TS, et al. (2013) Effects of common genetic variants associated with type 2 diabetes and glycemic traits on alpha- and beta-cell function and insulin action in humans. Diabetes 62: 2978-2983

[133] Bouatia-Naji N, Bonnefond A, Cavalcanti-Proença C, et al. (2009) A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk. Nat Genet 41: 89-94

[134] Lyssenko V, Nagorny CL, Erdos MR, et al. (2009) Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. Nat Genet 41: 82-88

[135] Prokopenko I, Langenberg C, Florez JC, et al. (2009) Variants in MTNR1B influence fasting glucose levels. Nat Genet 41: 77-81

[136] Sparsø T, Bonnefond A, Andersson E, et al. (2009) G-allele of intronic rs10830963 in MTNR1B confers increased risk of impaired fasting glycemia and type 2 diabetes through an impaired glucose-stimulated insulin release: studies involving 19,605 Europeans. Diabetes 58: 1450-1456

[137] Bonnefond A, Clement N, Fawcett K, et al. (2012) Rare MTNR1B variants impairing melatonin receptor 1B function contribute to type 2 diabetes. Nat Genet 44: 297-301

[138] Gaulton KJ, Ferreira T, Lee Y, et al. (2015) Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. Nat Genet 47: 1415-1425

[139] Bonnefond A, Froguel P (2017) The case for too little melatonin signalling in increased diabetes risk. Diabetologia 60: 823-825

[140] Mulder H (2017) Melatonin signalling and type 2 diabetes risk: too little, too much or just right? Diabetologia 60: 826-829

[141] Dimas AS, Lagou V, Barker A, et al. (2014) Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes 63: 2158-2171

[142] Mahajan A, Wessel J, Willems SM, et al. (2018) Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. Nat Genet 50: 559-571

[143] Udler MS, Kim J, von Grotthuss M, et al. (2018) Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. PLoS Med 15: e1002654

[144] Lotta LA, Gulati P, Day FR, et al. (2017) Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. Nat Genet 49: 17-26

[145] Yaghootkar H, Scott RA, White CC, et al. (2014) Genetic evidence for a normal-weight "metabolically obese" phenotype linking insulin resistance, hypertension, coronary artery disease, and type 2 diabetes. Diabetes 63: 4369-4377

[146] Simonis-Bik AM, Nijpels G, van Haeften TW, et al. (2010) 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 59: 293-301

[147] Polonsky KS, Rubenstein AH (1984) C-peptide as a measure of the secretion and hepatic extraction of insulin. Pitfalls and limitations. Diabetes 33: 486-494

[148] Goodarzi MO, Guo X, Cui J, et al. (2013) Systematic evaluation of validated type 2 diabetes and glycaemic trait loci for association with insulin clearance. Diabetologia 56: 1282-1290

[149] Duckworth WC, Bennett RG, Hamel FG (1998) Insulin degradation: progress and potential. Endocr Rev 19: 608-624

[150] Røder ME, Porte D, Jr., Schwartz RS, Kahn SE (1998) Disproportionately elevated proinsulin levels reflect the degree of impaired B cell secretory capacity in patients with noninsulin-dependent diabetes mellitus. J Clin Endocrinol Metab 83: 604-608

[151] Strawbridge RJ, Dupuis J, Prokopenko I, et al. (2011) Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. Diabetes 60: 2624-2634

[152] Ingelsson E, Langenberg C, Hivert MF, et al. (2010) Detailed physiologic characterization reveals diverse mechanisms for novel genetic loci regulating glucose and insulin metabolism in humans. Diabetes 59: 1266-1275

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

[154] Sparsø T, Andersen G, Nielsen T, et al. (2008) The GCKR rs780094 polymorphism is associated with elevated fasting serum triacylglycerol, reduced fasting and OGTT-related insulinaemia, and reduced risk of type 2 diabetes. Diabetologia 51: 70-75

[155] Kilpelainen TO, Zillikens MC, Stancakova A, et al. (2011) Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. Nat Genet 43: 753-760

[156] Yamauchi T, Hara K, Maeda S, et al. (2010) A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. Nat Genet 42: 864-868

[157] Schafer SA, Mussig K, Staiger H, et al. (2009) A common genetic variant in WFS1 determines impaired glucagon-like peptide-1-induced insulin secretion. Diabetologia 52: 1075-1082

[158] Sparsø T, Andersen G, Albrechtsen A, et al. (2008) Impact of polymorphisms in WFS1 on prediabetic phenotypes in a population-based sample of middle-aged people with normal and abnormal glucose regulation. Diabetologia 51: 1646-1652

[159] Harder MN, Appel EV, Grarup N, et al. (2015) The type 2 diabetes risk allele of TMEM154-rs6813195 associates with decreased beta cell function in a study of 6,486 Danes. PLoS One 10: e0120890

[160] Richards JB, Waterworth D, O'Rahilly S, et al. (2009) A genome-wide association study reveals variants in ARL15 that influence adiponectin levels. PLoS Genet 5: e1000768

[161] Ried JS, Jeff MJ, Chu AY, et al. (2016) A principal component meta-analysis on multiple anthropometric traits identifies novel loci for body shape. Nat Commun 7: 13357

[162] Thomsen SK, Raimondo A, Hastoy B, et al. (2018) Type 2 diabetes risk alleles in PAM impact insulin release from human pancreatic beta-cells. Nat Genet 50: 1122-1131

[163] Rose CS, Ek J, Urhammer SA, et al. (2005) A -30G>A polymorphism of the beta-cell-specific glucokinase promoter associates with hyperglycemia in the general population of whites. Diabetes 54: 3026-3031

[164] Harder MN, Ribel-Madsen R, Justesen JM, et al. (2013) Type 2 diabetes risk alleles near BCAR1 and in ANK1 associate with decreased beta-cell function whereas risk alleles near ANKRD55 and GRB14 associate with decreased insulin sensitivity in the Danish Inter99 cohort. J Clin Endocrinol Metab 98: E801-806

[165] Loos RJ, Franks PW, Francis RW, et al. (2007) TCF7L2 polymorphisms modulate proinsulin levels and beta-cell function in a British Europid population. Diabetes 56: 1943-1947

[166] Yasuda K, Miyake K, Horikawa Y, et al. (2008) Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat Genet 40: 1092-1097

[167] Mussig K, Staiger H, Machicao F, et al. (2009) Association of type 2 diabetes candidate polymorphisms in KCNQ1 with incretin and insulin secretion. Diabetes 58: 1715-1720

[168] Stancáková A, Kuulasmaa T, Paananen J, et al. (2009) 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 58: 2129-2136

[169] Nielsen T, Sparso T, Grarup N, et al. (2011) Type 2 diabetes risk allele near CENTD2 is associated with decreased glucose-stimulated insulin release. Diabetologia 54: 1052-1056

[170] Steinthorsdottir V, Thorleifsson G, Sulem P, et al. (2014) Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. Nat Genet 46: 294-298

[171] Yaghootkar H, Stancakova A, Freathy RM, et al. (2015) Association analysis of 29,956 individuals confirms that a low-frequency variant at CCND2 halves the risk of type 2 diabetes by enhancing insulin secretion. Diabetes 64: 2279-2285

[172] Romeo S, Kozlitina J, Xing C, et al. (2008) Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 40: 1461-1465

[173] Romeo S, Sentinelli F, Cambuli VM, et al. (2010) The 148M allele of the PNPLA3 gene is associated with indices of liver damage early in life. J Hepatol 53: 335-338

[174] Pingitore P, Romeo S (2018) The role of PNPLA3 in health and disease. Biochim Biophys Acta 18

[175] Trepo E, Romeo S, Zucman-Rossi J, Nahon P (2016) PNPLA3 gene in liver diseases. J Hepatol 65: 399-412

[176] Liu DJ, Peloso GM, Yu H, et al. (2017) Exome-wide association study of plasma lipids in >300,000 individuals. Nat Genet 49: 1758-1766

[177] Tundo GR, Sbardella D, Ciaccio C, et al. (2017) Multiple functions of insulin-degrading enzyme: a metabolic crosslight? Crit Rev Biochem Mol Biol 52: 554-582

[178] Maianti JP, McFedries A, Foda ZH, et al. (2014) Anti-diabetic activity of insulin-degrading enzyme inhibitors mediated by multiple hormones. Nature 511: 94-98

[179] Tang WJ (2016) Targeting Insulin-Degrading Enzyme to Treat Type 2 Diabetes Mellitus. Trends Endocrinol Metab 27: 24-34

[180] Rosengren AH, Braun M, Mahdi T, et al. (2012) Reduced insulin exocytosis in human pancreatic beta-cells with gene variants linked to type 2 diabetes. Diabetes 61: 1726-1733

[181] Grarup N, Rose CS, Andersson EA, et al. (2007) 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 56: 3105-3111

[182] Helgadottir A, Thorleifsson G, Manolescu A, et al. (2007) A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction. Science 316: 1491-1493

[183] McPherson R, Pertsemlidis A, Kavaslar N, et al. (2007) A Common Allele on Chromosome 9 Associated with Coronary Heart Disease. Science 316: 1488-1491

[184] Hannou SA, Wouters K, Paumelle R, Staels B (2015) Functional genomics of the CDKN2A/B locus in cardiovascular and metabolic disease: what have we learned from GWASs? Trends Endocrinol Metab 26: 176-184

[185] Kohler CU, Olewinski M, Tannapfel A, Schmidt WE, Fritsch H, Meier JJ (2011) Cell cycle control of beta-cell replication in the prenatal and postnatal human pancreas. Am J Physiol Endocrinol Metab 300: E221-230

[186] Krishnamurthy J, Ramsey MR, Ligon KL, et al. (2006) $p16^{INK4a}$ induces an age-dependent decline in islet regenerative potential. Nature 443: 453-457

[187] Moritani M, Yamasaki S, Kagami M, et al. (2005) Hypoplasia of endocrine and exocrine pancreas in homozygous transgenic TGF-beta1. Mol Cell Endocrinol 229: 175-184

[188] Pal A, Potjer TP, Thomsen SK, et al. (2016) Loss-of-Function Mutations in the Cell-Cycle Control Gene CDKN2A Impact on Glucose Homeostasis in Humans. Diabetes 65: 527-533

[189] Fogarty MP, Panhuis TM, Vadlamudi S, Buchkovich ML, Mohlke KL (2013) Allele-specific transcriptional activity at type 2 diabetes-associated single nucleotide polymorphisms in regions of pancreatic islet open chromatin at the JAZF1 locus. Diabetes 62: 1756-1762

[190] Fogarty MP, Cannon ME, Vadlamudi S, Gaulton KJ, Mohlke KL (2014) Identification of a regulatory variant that binds FOXA1 and FOXA2 at the CDC123/CAMK1D type 2 diabetes GWAS locus. PLoS Genet 10: e1004633

[191] Coutinho M, Gerstein HC, Wang Y, Yusuf S (1999) The relationship between glucose and incident cardiovascular events. A metaregression analysis of published data from 20 studies of 95,783 individuals followed for 12.4 years. Diabetes Care 22: 233-240

[192] Lawlor DA, Fraser A, Ebrahim S, Smith GD (2007) Independent associations of fasting insulin, glucose, and glycated haemoglobin with stroke and coronary heart disease in older women. PLoS Med 4: e263

[193] Lawes CM, Parag V, Bennett DA, et al. (2004) Blood glucose and risk of cardiovascular disease in the Asia Pacific region. Diabetes Care 27: 2836-2842

[194] Auton A, Brooks LD, Durbin RM, et al. (2015) A global reference for human genetic variation. Nature 526: 68-74

[195] Stitziel NO, Kiezun A, Sunyaev S (2011) Computational and statistical approaches to analyzing variants identified by exome sequencing. Genome Biol 12: 227

[196] Ladouceur M, Dastani Z, Aulchenko YS, Greenwood CM, Richards JB (2012) The empirical power of rare variant association methods: results from sanger sequencing in 1,998 individuals. PLoS Genet 8: e1002496

[197] Liu DJ, Peloso GM, Zhan X, et al. (2014) Meta-analysis of gene-level tests for rare variant association. Nat Genet 46: 200-204

[198] Zuk O, Schaffner SF, Samocha K, et al. (2014) Searching for missing heritability: Designing rare variant association studies. Proc Natl Acad Sci U S A 111: E455-E464

[199] Yang J, Benyamin B, McEvoy BP, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42: 565-569

[200] Gibson G (2012) Rare and common variants: twenty arguments. Nat Rev Genet 13: 135-145

[201] Pritchard JK, Cox NJ (2002) The allelic architecture of human disease genes: common disease-common variant...or not? Hum Mol Genet 11: 2417-2423

[202] Bodmer W, Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet 40: 695-701

[203] Pritchard JK (2001) Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 69: 124-137

[204] McClellan J, King MC (2010) Genetic heterogeneity in human disease. Cell 141: 210-217

[205] Tennessen JA, Bigham AW, O'Connor TD, et al. (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337: 64-69

[206] Eyre-Walker A (2010) Evolution in health and medicine Sackler colloquium: Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proc Natl Acad Sci U S A 107 Suppl 1: 1752-1756

[207] Cirulli ET, Goldstein DB (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11: 415-425

[208] Shendure J, Balasubramanian S, Church GM, et al. (2017) DNA sequencing at 40: past, present and future. Nature 550: 345-353

[209] Albert TJ, Molla MN, Muzny DM, et al. (2007) Direct selection of human genomic loci by microarray hybridization. Nat Methods 4: 903-905

[210] Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH (2005) Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. Nat Genet 37: 161-165

[211] Flannick J, Thorleifsson G, Beer NL, et al. (2014) Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. Nat Genet 46: 357-363

[212] Rivas MA, Pirinen M, Conrad DF, et al. (2015) Effect of predicted protein-truncating genetic variants on the human transcriptome. Science 348: 666-669

[213] Ng SB, Turner EH, Robertson PD, et al. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461: 272-276

[214] Ng SB, Buckingham KJ, Lee C, et al. (2010) Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42: 30-35

[215] Surakka I, Horikoshi M, Magi R, et al. (2015) The impact of low-frequency and rare variants on lipid levels. Nat Genet 47: 589-597

[216] Fuchsberger C, Flannick J, Teslovich TM, et al. (2016) The genetic architecture of type 2 diabetes. Nature 536: 41-47

[217] Helgadottir A, Gretarsdottir S, Thorleifsson G, et al. (2016) Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease. Nat Genet 48: 634-639

[218] Borrego F (2013) The CD300 molecules: an emerging family of regulators of the immune system. Blood 121: 1951-1960

[219] Cannon JP, O'Driscoll M, Litman GW (2012) Specific lipid recognition is a general feature of CD300 and TREM molecules. Immunogenetics 64: 39-47

[220] Umemoto E, Hayasaka H, Bai Z, et al. (2011) Novel regulators of lymphocyte trafficking across high endothelial venules. Crit Rev Immunol 31: 147-169

[221] Takatsu H, Hase K, Ohmae M, et al. (2006) CD300 antigen like family member G: A novel Ig receptor like protein exclusively expressed on capillary endothelium. Biochem Biophys Res Commun 348: 183-191

[222] Kooner JS, Saleheen D, Sim X, et al. (2011) Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet 43: 984-989

[223] Heid IM, Jackson AU, Randall JC, et al. (2010) Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet 42: 949-960

[224] Teslovich TM, Musunuru K, Smith AV, et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466: 707-713

[225] Cooney GJ, Lyons RJ, Crew AJ, et al. (2004) Improved glucose homeostasis and enhanced insulin signalling in Grb14-deficient mice. EMBO J 23: 582-593

[226] Hu L, Xiao Y, Xiong Z, et al. (2017) MACF1, versatility in tissue-specific function and in human disease. Semin Cell Dev Biol 69: 3-8

[227] Nicolson TJ, Bellomo EA, Wijesekara N, et al. (2009) Insulin storage and glucose homeostasis in mice null for the granule zinc transporter ZnT8 and studies of the type 2 diabetes-associated variants. Diabetes 58: 2070-2083

[228] Tamaki M, Fujitani Y, Hara A, et al. (2013) The diabetes-susceptible gene SLC30A8/ZnT8 regulates hepatic insulin clearance. J Clin Invest 123: 4513-4524

[229] Rutter GA, Chimienti F (2015) SLC30A8 mutations in type 2 diabetes. Diabetologia 58: 31-36

[230] Lohmueller KE, Sparsø T, Li Q, et al. (2013) Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. Am J Hum Genet 93: 1072-1086

[231] Timpson NJ, Walter K, Min JL, et al. (2014) A rare variant in APOC3 is associated with plasma triglyceride and VLDL levels in Europeans. Nat Commun 5: 4871

[232] Jørgensen AB, Frikke-Schmidt R, Nordestgaard BG, Tybjaerg-Hansen A (2014) Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. N Engl J Med 371: 32-41

[233] TG and HDL Working Group of the Exome Sequencing Project NH, Lung, and Blood Institute, Crosby J, Peloso GM, et al. (2014) Loss-of-function mutations in APOC3, triglycerides, and coronary disease. N Engl J Med 371: 22-31

[234] Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH (2004) Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 305: 869-872

[235] Cohen JC, Pertsemlidis A, Fahmi S, et al. (2006) Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. Proc Natl Acad Sci U S A 103: 1810-1815

[236] Dewey FE, Murray MF, Overton JD, et al. (2016) Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science 354: aaf6814

[237] Lange LA, Hu Y, Zhang H, et al. (2014) Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. Am J Hum Genet 94: 233-245

[238] Lee S, Abecasis GR, Boehnke M, Lin X (2014) Rare-variant association analysis: study designs and statistical tests. Am J Hum Genet 95: 5-23

[239] Huyghe JR, Jackson AU, Fogarty MP, et al. (2013) Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nat Genet 45: 197-201

[240] McCarthy MI, Abecasis GR, Cardon LR, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9: 356-369

[241] Justice AE, Karaderi T, Highland HM, et al. (2018) Protein-coding variants implicate novel genes related to lipid-homeostasis contributing to body fat distribution. BioRxiv 10.1101/352674

[242] Gudbjartsson DF, Helgason H, Gudjonsson SA, et al. (2015) Large-scale whole-genome sequencing of the Icelandic population. Nat Genet 47: 435-444

[243] Kong A, Masson G, Frigge ML, et al. (2008) Detection of sharing by descent, long-range phasing and haplotype imputation. Nat Genet 40: 1068-1075

[244] Rafnar T, Gudbjartsson DF, Sulem P, et al. (2011) Mutations in BRIP1 confer high risk of ovarian cancer. Nat Genet 43: 1104-1107

[245] Roth JR, Lawrence JG, Bobik TA (1996) Cobalamin (coenzyme B12): synthesis and biological significance. Annu Rev Microbiol 50: 137-181

[246] Scott JM (1999) Folate and vitamin B_{12}. Proc Nutr Soc 58: 441-448

[247] Knight BA, Shields BM, Brook A, et al. (2015) Lower Circulating B12 Is Associated with Higher Obesity and Insulin Resistance during Pregnancy in a Non-Diabetic White British Population. PLoS One 10: e0135268

[248] Rafnsson SB, Saravanan P, Bhopal RS, Yajnik CS (2011) Is a low blood level of vitamin B12 a cardiovascular and diabetes risk factor? A systematic review of cohort studies. Eur J Nutr 50: 97-106

[249] Krishnaveni GV, Hill JC, Veena SR, et al. (2009) Low plasma vitamin B12 in pregnancy is associated with gestational 'diabesity' and later diabetes. Diabetologia 52: 2350-2358

[250] Adaikalakoteswari A, Jayashri R, Sukumar N, et al. (2014) Vitamin B12 deficiency is associated with adverse lipid profile in Europeans and Indians with type 2 diabetes. Cardiovasc Diabetol 13: 129

[251] Weikert C, Dierkes J, Hoffmann K, et al. (2007) B vitamin plasma levels and the risk of ischemic stroke and transient ischemic attack in a German cohort. Stroke 38: 2912-2918

[252] Hazra A, Kraft P, Lazarus R, et al. (2009) Genome-wide significant predictors of metabolites in the one-carbon metabolism pathway. Hum Mol Genet 18: 4677-4687

[253] Hazra A, Kraft P, Selhub J, et al. (2008) Common variants of FUT2 are associated with plasma vitamin B12 levels. Nat Genet 40: 1160-1162

[254] Lin X, Lu D, Gao Y, et al. (2012) Genome-wide association study identifies novel loci associated with serum level of vitamin B12 in Chinese men. Hum Mol Genet 21: 2610-2617

[255] Hustad S, Midttun O, Schneede J, Vollset SE, Grotmol T, Ueland PM (2007) The methylenetetrahydrofolate reductase 677C-->T polymorphism as a modulator of a B vitamin network with major effects on homocysteine metabolism. Am J Hum Genet 80: 846-855

[256] Visscher PM, Wray NR, Zhang Q, et al. (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 101: 5-22

[257] Davey Smith G, Hemani G (2014) Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 23: R89-98

[258] Pingault JB, O'Reilly PF, Schoeler T, Ploubidis GB, Rijsdijk F, Dudbridge F (2018) Using genetic data to strengthen causal inference in observational research. Nat Rev Genet 19: 566-580

[259] Allin KH, Friedrich N, Pietzner M, et al. (2017) Genetic determinants of serum vitamin B12 and their relation to body mass index. Eur J Epidemiol 32: 125-134

[260] Husemoen LL, Skaaby T, Thuesen BH, et al. (2016) Mendelian randomisation study of the associations of vitamin B12 and folate genetic risk scores with blood pressure and fasting serum lipid levels in three Danish population-based studies. Eur J Clin Nutr 70: 613-619

[261] Moen GH, Qvigstad E, Birkeland KI, Evans DM, Sommer C (2018) Are serum concentrations of vitamin B-12 causally related to cardiometabolic risk factors and disease? A Mendelian randomization study. Am J Clin Nutr 108: 398-404

[262] Debreceni B, Debreceni L (2014) The role of homocysteine-lowering B-vitamins in the primary prevention of cardiovascular disease. Cardiovasc Ther 32: 130-138

[263] Rimm EB, Stampfer MJ (2011) Folate and cardiovascular disease: one size does not fit all. Lancet 378: 544-546

[264] Imamura M, Takahashi A, Yamauchi T, et al. (2016) Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. Nat Commun 7: 10531

[265] Cho YS, Chen CH, Hu C, et al. (2012) Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat Genet 44: 67-72

[266] Morris AP (2018) Progress in defining the genetic contribution to type 2 diabetes susceptibility. Curr Opin Genet Dev 50: 41-51

[267] Lohmueller KE (2014) The impact of population demography and selection on the genetic architecture of complex traits. PLoS Genet 10: e1004379

[268] Gulløv HC (2004) Grønlands forhistorie. Gyldendal, Copenhagen

[269] Raghavan M, DeGiorgio M, Albrechtsen A, et al. (2014) The genetic prehistory of the New World Arctic. Science 345: 1255832

[270] Moltke I, Fumagalli M, Korneliussen TS, et al. (2015) Uncovering the genetic history of the present-day Greenlandic population. Am J Hum Genet 96: 54-69

[271] Grarup N, Moltke I, Albrechtsen A, Hansen T (2015) Diabetes in Population Isolates: Lessons from Greenland. Rev Diabet Stud 12: 320-329

[272] Diamond J (2003) The double puzzle of diabetes. Nature 423: 599-602

[273] Schulz LO, Chaudhari LS (2015) High-Risk Populations: The Pimas of Arizona and Mexico. Curr Obes Rep 4: 92-98

[274] Baier L, Hanson R (2004) Genetic studies of the etiology of type 2 diabetes in Pima Indians: hunting for pieces to a complicated puzzle. Diabetes 53: 1181-1186

[275] Alberti KG, Zimmet PZ (1998) 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 15: 539-553

[276] Sagild U, Littauer J, Jespersen CS, Andersen S (1966) Epidemiological studies in Greenland 1962-1964. I. Diabetes mellitus in Eskimos. Acta Med Scand 179: 29-39

[277] Jørgensen ME, Borch-Johnsen K, Stolk R, Bjerregaard P (2013) Fat Distribution and Glucose Intolerance Among Greenland Inuit. Diabetes Care 36: 2988-2994

[278] Bjerregaard P, Curtis T, Borch-Johnsen K, et al. (2003) Inuit health in Greenland: a population survey of life style and disease in Greenland and among Inuit living in Denmark. Int J Circumpolar Health 62 Suppl 1: 3-79

[279] Young TK, Bjerregaard P, Dewailly E, Risica PM, Jorgensen ME, Ebbesson SE (2007) Prevalence of obesity and its metabolic correlates among the circumpolar inuit in 3 countries. Am J Public Health 97: 691-695

[280] Bjerregaard P, Jorgensen ME (2013) Prevalence of obesity among Inuit in Greenland and temporal trend by social position. Am J Hum Biol 25: 335-340

[281] Andersen S, Rex KF, Noahsen P, et al. (2014) Forty-five year trends in overweight and obesity in an indigenous arctic Inuit Society in transition and spatiotemporal trends. Am J Hum Biol 26: 511-517

[282] Jørgensen ME, Glumer C, Bjerregaard P, Gyntelberg F, Jørgensen T, Borch-Johnsen K (2003) Obesity and central fat pattern among Greenland Inuit and a general population of Denmark (Inter99): relationship to metabolic risk factors. Int J Obes Relat Metab Disord 27: 1507-1515

[283] Jørgensen ME, Borch-Johnsen K, Bjerregaard P (2006) Lifestyle modifies obesity-associated risk of cardiovascular disease in a genetically homogeneous population. Am J Clin Nutr 84: 29-36

[284] Jorgensen ME, Borch-Johnsen K, Stolk R, Bjerregaard P (2013) Fat distribution and glucose intolerance among Greenland Inuit. Diabetes Care 36: 2988-2994

[285] Jørgensen ME, Borch-Johnsen K, Witte DR, Bjerregaard P (2012) Diabetes in Greenland and its relationship with urbanization. Diabet Med 29: 755-760

[286] Jeppesen C, Bjerregaard P, Jorgensen ME (2014) Dietary patterns in Greenland and their relationship with type 2 diabetes mellitus and glucose intolerance. Public Health Nutr 17: 462-470

[287] Dahl-Petersen IK, Bjerregaard P, Brage S, Jorgensen ME (2013) Physical activity energy expenditure is associated with 2-h insulin independently of obesity among Inuit in Greenland. Diabetes Res Clin Pract 102: 242-249

[288] Voight BF, Kang HM, Ding J, et al. (2012) The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genet 8: e1002793

[289] Manousaki D, Kent JW, Jr., Haack K, et al. (2016) Toward Precision Medicine: TBC1D4 Disruption Is Common Among the Inuit and Leads to Underdiagnosis of Type 2 Diabetes. Diabetes Care 39: 1889-1895

[290] Sano H, Kane S, Sano E, et al. (2003) Insulin-stimulated Phosphorylation of a Rab GTPase-activating Protein Regulates GLUT4 Translocation. J Biol Chem 278: 14599-14602

[291] Wang HY, Ducommun S, Quan C, et al. (2013) AS160 deficiency causes whole-body insulin resistance via composite effects in multiple tissues. Biochem J 449: 479-489

[292] Lansey MN, Walker NN, Hargett SR, Stevens JR, Keller SR (2012) Deletion of Rab GAP AS160 modifies glucose uptake and GLUT4 translocation in primary skeletal muscles and adipocytes and impairs glucose homeostasis. Am J Physiol Endocrinol Metab 303: E1273-E1286

[293] Treebak JT, Frøsig C, Pehmøller C, et al. (2009) Potential role of TBC1D4 in enhanced post-exercise insulin action in human skeletal muscle. Diabetologia 52: 891-900

[294] Vind BF, Pehmøller C, Treebak JT, et al. (2011) Impaired insulin-induced site-specific phosphorylation of TBC1 domain family, member 4 (TBC1D4) in skeletal muscle of type 2 diabetes patients is restored by endurance exercise-training. Diabetologia 54: 157-167

[295] Jørgensen ME, Schnurr TM, Jørsboe E, et al. (2016) Physical activity energy expenditure attenuates the effect of the TBC1D4 p.Arg684Ter loss-of-function variant. Diabetologia 59(S1): 209

[296] Langenberg C, Lotta LA (2018) Genomic insights into the causes of type 2 diabetes. Lancet 391: 2463-2474

[297] Sarwar N, Aspelund T, Eiriksdottir G, et al. (2010) Markers of Dysglycaemia and Risk of Coronary Heart Disease in People without Diabetes: Reykjavik Prospective Study and Systematic Review. PLoS Med 7: e1000278

[298] Faerch K, Vistisen D, Johansen NB, Jorgensen ME (2014) Cardiovascular risk stratification and management in pre-diabetes. Curr Diab Rep 14: 493

[299] Kodama S, Saito K, Tanaka S, et al. (2012) Fasting and Post-Challenge Glucose as Quantitative Cardiovascular Risk Factors: A Meta-Analysis. J Atheroscler Thromb 19: 385-396

[300] Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL (2005) Support from the relationship of genetic and geographic

distance in human populations for a serial founder effect originating in Africa. Proc Natl Acad Sci U S A 102: 15942-15947

[301] Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147: 915-925

[302] Krogh A, Krogh M (1915) A study of the diet and metabolism of Eskimos undertaken in 1908 on an expedition to Greenland. Medd Grønl 51: 1-52

[303] Kleinert M, Clemmensen C, Hofmann SM, et al. (2018) Animal models of obesity and diabetes mellitus. Nat Rev Endocrinol 14: 140-162

[304] Thomsen SK, Gloyn AL (2017) Human genetics as a model for target validation: finding new therapies for diabetes. Diabetologia 60: 960-970

[305] MacArthur DG, Balasubramanian S, Frankish A, et al. (2012) A systematic survey of loss-of-function variants in human protein-coding genes. Science 335: 823-828

[306] Dewey FE, Grove ME, Pan C, et al. (2014) Clinical interpretation and implications of whole-genome sequencing. JAMA 311: 1035-1045

[307] Dewey FE, Gusarova V, Dunbar RL, et al. (2017) Genetic and Pharmacologic Inactivation of ANGPTL3 and Cardiovascular Disease. N Engl J Med 377: 211-221

[308] DeBoever C, Tanigawa Y, Lindholm ME, et al. (2018) Medical relevance of protein-truncating variants across 337,205 individuals in the UK Biobank study. Nat Commun 9: 1612

[309] Emdin CA, Khera AV, Chaffin M, et al. (2018) Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease. Nat Commun 9: 1613

[310] Andersen MK, Pedersen CE, Moltke I, Hansen T, Albrechtsen A, Grarup N (2016) Genetics of Type 2 Diabetes: the Power of Isolated Populations. Curr Diab Rep 16: 65

[311] Sulem P, Helgason H, Oddson A, et al. (2015) Identification of a large set of rare complete human knockouts. Nat Genet 47: 448-452

[312] Lim ET, Würtz P, Havulinna AS, et al. (2014) Distribution and Medical Impact of Loss-of-Function Variants in the Finnish Founder Population. PLoS Genet 10: e1004494

[313] Fumagalli M, Moltke I, Grarup N, et al. (2015) Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science 349: 1343-1347

[314] Pedersen CT, Lohmueller KE, Grarup N, et al. (2017) The Effect of an Extreme and Prolonged Population Bottleneck on Patterns of Deleterious Variation: Insights from the Greenlandic Inuit. Genetics 205: 787-801

[315] Saleheen D, Natarajan P, Armean IM, et al. (2017) Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. Nature 544: 235-239

[316] Speliotes EK, Willer CJ, Berndt SI, et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 42: 937-948

[317] Warrington NM, Howe LD, Paternoster L, et al. (2015) A genome-wide association study of body mass index across early life and childhood. Int J Epidemiol 44: 700-712

[318] Saeed S, Bonnefond A, Tamanini F, et al. (2018) Loss-of-function mutations in ADCY3 cause monogenic severe obesity. Nat Genet 50: 175-179

[319] Tian Y, Peng B, Fu X (2018) New ADCY3 Variants Dance in Obesity Etiology. Trends Endocrinol Metab 29: 361-363

[320] Tong T, Shen Y, Lee HW, Yu R, Park T (2016) Adenylyl cyclase 3 haploinsufficiency confers susceptibility to diet-induced obesity and insulin resistance in mice. Sci Rep 6: 34179

[321] Pitman JL, Wheeler MC, Lloyd DJ, Walker JR, Glynne RJ, Gekakis N (2014) A gain-of-function mutation in adenylate cyclase 3 protects mice from diet-induced obesity. PLoS One 9: e110226

[322] Wang Z, Li V, Chan GC, et al. (2009) Adult type 3 adenylyl cyclase-deficient mice are obese. PLoS One 4: e6979

[323] Chen X, Luo J, Leng Y, et al. (2016) Ablation of Type III Adenylyl Cyclase in Mice Causes Reduced Neuronal Activity, Altered Sleep Pattern, and Depression-like Phenotypes. Biol Psychiatry 80: 836-848

[324] Vaisse C, Reiter JF, Berbari NF (2017) Cilia and Obesity. Cold Spring Harb Perspect Biol 9: a028217

[325] Siljee JE, Wang Y, Bernard AA, et al. (2018) Subcellular localization of MC4R with ADCY3 at neuronal primary cilia underlies a common pathway for genetic predisposition to obesity. Nat Genet 50: 180-185

[326] Mackay TF (2001) The genetic architecture of quantitative traits. Annu Rev Genet 35: 303-339

[327] Boyle EA, Li YI, Pritchard JK (2017) An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169: 1177-1186

[328] Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Proc Roy Soc Edinb 52: 399-433

[329] Barton NH, Etheridge AM, Veber A (2017) The infinitesimal model: Definition, derivation, and implications. Theor Popul Biol 118: 50-73

[330] Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era - concepts and misconceptions. Nat Rev Genet 9: 255-266

[331] Lande R (1976) Natural selection and random genetic drift in phenotypic evoluation. Evolution 30: 314-334

[332] Sigma Type Diabetes Consortium, Estrada K, Aukrust I, et al. (2014) Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. JAMA 311: 2305-2314

[333] The Sigma Type Diabetes Consortium, Williams AL, Jacobs SBR, et al. (2014) Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature 506: 97-101

[334] Mercader JM, Liao RG, Bell AD, et al. (2017) A Loss-of-Function Splice Acceptor Variant in IGF2 Is Protective for Type 2 Diabetes. Diabetes 66: 2903-2914

[335] Baier LJ, Muller YL, Remedi MS, et al. (2015) ABCC8 R1420H Loss-of-Function Variant in a Southwest American Indian Community: Association With Increased Birth Weight and Doubled Risk of Type 2 Diabetes. Diabetes 64: 4322-4332

[336] Hegele RA, Cao H, Harris SB, Hanley AJ, Zinman B (1999) The hepatic nuclear factor-1alpha G319S variant is associated with early-onset type 2 diabetes in Canadian Oji-Cree. J Clin Endocrinol Metab 84: 1077-1082

[337] Albert JS, Yerges-Armstrong LM, Horenstein RB, et al. (2014) Null mutation in hormone-sensitive lipase gene and risk of type 2 diabetes. N Engl J Med 370: 2307-2315

[338] Minster RL, Hawley NL, Su CT, et al. (2016) A thrifty variant in CREBRF strongly influences body mass index in Samoans. Nat Genet 48: 1049-1054

[339] Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB (2018) Genetic architecture: the shape of the genetic contribution to human traits and disease. Nat Rev Genet 19: 110-124

[340] Southam L, Gilly A, Suveges D, et al. (2017) Whole genome sequencing and imputation in isolated populations identify genetic associations with medically-relevant complex traits. Nat Commun 8: 15606

[341] Tachmazidou I, Dedoussis G, Southam L, et al. (2013) A rare functional cardioprotective APOC3 variant has risen in frequency in distinct population isolates. Nat Commun 4: 2872

[342] Manning A, Highland HM, Gasser J, et al. (2017) A Low-Frequency Inactivating AKT2 Variant Enriched in the Finnish Population Is Associated With Fasting Insulin Levels and Type 2 Diabetes Risk. Diabetes 66: 2019-2032

[343] Dupuis J, Langenberg C, Prokopenko I, et al. (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 42: 105-116

[344] Froguel P, Vaxillaire M, Sun F, et al. (1992) Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. Nature 356: 162-164

[345] Horikawa Y, Iwasaki N, Hara M, et al. (1997) Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. Nat Genet 17: 384-385

[346] Yamagata K, Furuta H, Oda N, et al. (1996) Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). Nature 384: 458-460

[347] Yamagata K, Oda N, Kaisaki PJ, et al. (1996) Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). Nature 384: 455-458

[348] Barroso I, Gurnell M, Crowley VE, et al. (1999) Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. Nature 402: 880-883

[349] Gloyn AL (2004) Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. N Engl J Med 350: 1838-1849

[350] Senee V (2006) Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. Nat Genet 38: 682-687

[351] Inoue H, Tanizawa Y, Wasson J, et al. (1998) A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). Nat Genet 20: 143-148

[352] Krude H, Biebermann H, Luck W, Horn R, Brabant G, Grüters A (1998) Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. Nat Genet 19: 155-157

[353] Yeo GS, Farooqi IS, Aminian S, Halsall DJ, Stanhope RG, O'Rahilly S (1998) A frameshift mutation in MC4R associated with dominantly inherited human obesity. Nat Genet 20: 111-112

[354] Clement K, Vaisse C, Lahlou N, et al. (1998) A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. Nature 392: 398-401

[355] Han JC, Liu QR, Jones M, et al. (2008) Brain-derived neurotrophic factor and obesity in the WAGR syndrome. N Engl J Med 359: 918-927

[356] Bochukova EG, Huang N, Keogh J, et al. (2010) Large, rare chromosomal deletions associated with severe early-onset obesity. Nature 463: 666-670

[357] Jackson RS, Creemers JW, Ohagi S, et al. (1997) Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. Nat Genet 16: 303-306

[358] Odom DT, Zizlsperger N, Gordon DB, et al. (2004) Control of pancreas and liver gene expression by HNF transcription factors. Science 303: 1378-1381

[359] Boj SF, Parrizas M, Maestro MA, Ferrer J (2001) A transcription factor regulatory circuit in differentiated pancreatic cells. Proc Natl Acad Sci U S A 98: 14481-14486

[360] McCarthy MI, Hattersley AT (2008) Learning from molecular genetics: novel insights arising from the definition of genes for monogenic and type 2 diabetes. Diabetes 57: 2889-2898

[361] Pearson ER, Starkey BJ, Powell RJ, Gribble FM, Clark PM, Hattersley AT (2003) Genetic cause of hyperglycaemia and response to treatment in diabetes. Lancet 362: 1275-1281

[362] Martagon AJ, Bello-Chavolla OY, Arellano-Campos O, et al. (2018) Mexican Carriers of the HNF1A p.E508K Variant Do Not Experience an Enhanced Response to Sulfonylureas. Diabetes Care 41: 1726-1731

[363] Majithia AR, Tsuda B, Agostini M, et al. (2016) Prospective functional classification of all possible missense variants in PPARG. Nat Genet 48: 1570-1575

[364] MacArthur DG, Manolio TA, Dimmock DP, et al. (2014) Guidelines for investigating causality of sequence variants in human disease. Nature 508: 469-476

[365] Larsen LH, Echwald SM, Sorensen TI, Andersen T, Wulff BS, Pedersen O (2005) Prevalence of mutations and functional analyses of melanocortin 4 receptor variants identified among 750 men with juvenile-onset obesity. J Clin Endocrinol Metab 90: 219-224

[366] Thanabalasingham G, Owen KR (2011) Diagnosis and management of maturity onset diabetes of the young (MODY). BMJ 343: d6044

[367] Pearson ER (2006) Switching from insulin to oral sulfonylureas in patients with diabetes due to Kir6.2 mutations. N Engl J Med 355: 467-477

[368] Babenko AP, Polak M, Cavé H, et al. (2006) Activating mutations in the ABCC8 gene in neonatal diabetes mellitus. N Engl J Med 355: 456-466

[369] Maruthur NM, Gribble MO, Bennett WL, et al. (2014) The pharmacogenetics of type 2 diabetes: a systematic review. Diabetes Care 37: 876-886

[370] Zhou K, Yee SW, Seiser EL, et al. (2016) Variation in the glucose transporter gene SLC2A2 is associated with glycemic response to metformin. Nat Genet 48: 1055-1059

[371] Owen KR (2018) Monogenic diabetes in adults: what are the new developments? Curr Opin Genet Dev 50: 103-110

[372] McCarthy MI (2017) Painting a new picture of personalised medicine for diabetes. Diabetologia 60: 793-799

[373] Inzucchi SE, Bergenstal RM, Buse JB, et al. (2015) Management of hyperglycaemia in type 2 diabetes, 2015: a patient-centred approach. Update to a position statement of the American Diabetes Association and the European Association for the Study of Diabetes. Diabetologia 58: 429-442

[374] Lyssenko V, Jonsson A, Almgren P, et al. (2008) Clinical risk factors, DNA variants, and the development of type 2 diabetes. N Engl J Med 359: 2220-2232

[375] Ahlqvist E, Storm P, Karajamaki A, et al. (2018) Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol 6: 361-369

[376] Merino J, Leong A, Liu CT, et al. (2018) Metabolomics insights into early type 2 diabetes pathogenesis and detection in individuals with normal fasting glucose. Diabetologia 61: 1315-1324

[377] Maurano MT, Humbert R, Rynes E, et al. (2012) Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science 337: 1190-1195

[378] Pai AA, Pritchard JK, Gilad Y (2015) The Genetic and Mechanistic Basis for Variation in Gene Regulation. PLOS Genetics 11: e1004857

[379] The Encode Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57-74

[380] Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. (2015) Integrative analysis of 111 reference human epigenomes. Nature 518: 317-330

[381] GTEx Consortium (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348: 648-660

[382] Gallagher MD, Chen-Plotkin AS (2018) The Post-GWAS Era: From Association to Function. Am J Hum Genet 102: 717-730

[383] Morris AP (2014) Fine mapping of type 2 diabetes susceptibility loci. Curr Diab Rep 14: 549

[384] Roy SS, Mukherjee AK, Chowdhury S (2018) Insights about genome function from spatial organization of the genome. Hum Genomics 12: 8

[385] Ford K, McDonald D, Mali P (2018) Functional Genomics via CRISPR-Cas. J Mol Biol 10.1016/j.jmb.2018.06.034

[386] Potter SS (2018) Single-cell RNA sequencing for the study of development, physiology and disease. Nat Rev Nephrol 14: 479-492

[387] Scannell JW, Blanckley A, Boldon H, Warrington B (2012) Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov 11: 191-200

[388] Nelson MR, Tipney H, Painter JL, et al. (2015) The support of human genetic evidence for approved drug indications. Nat Genet 47: 856-860

[389] Hurle MR, Nelson MR, Agarwal P, Cardon LR (2016) Trial watch: Impact of genetically supported target selection on R&D productivity. Nat Rev Drug Discov 15: 596-597

[390] Dewey FE, Gromada J, Shuldiner AR (2016) Variants in ANGPTL4 and the Risk of Coronary Artery Disease. N Engl J Med 375: 2305-2306

[391] Abifadel M, Varret M, Rabes JP, et al. (2003) Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. Nat Genet 34: 154-156

[392] Cohen JC, Boerwinkle E, Mosley TH, Hobbs HH (2006) Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med 354: 1264-1272

[393] Blom DJ, Hala T, Bolognese M, et al. (2014) A 52-week placebo-controlled trial of evolocumab in hyperlipidemia. N Engl J Med 370: 1809-1819

[394] El Khoury P, Elbitar S, Ghaleb Y, et al. (2017) PCSK9 Mutations in Familial Hypercholesterolemia: from a Groundbreaking Discovery to Anti-PCSK9 Therapies. Curr Atheroscler Rep 19: 49

[395] Scott RA, Freitag DF, Li L, et al. (2016) A genomic approach to therapeutic target validation identifies a glucose-lowering GLP1R variant protective for coronary heart disease. Sci Transl Med 8: 341ra376

[396] Drucker DJ, Nauck MA (2006) The incretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes. Lancet 368: 1696-1705

[397] Plenge RM, Scolnick EM, Altshuler D (2013) Validating therapeutic targets through human genetics. Nat Rev Drug Discov 12: 581-594

[398] Turnbull C, Scott RH, Thomas E, et al. (2018) The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. BMJ 361: k1952

[399] Gaziano JM, Concato J, Brophy M, et al. (2016) Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 70: 214-223

[400] Chen Z, Chen J, Collins R, et al. (2011) China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol 40: 1652-1666

[401] Wijmenga C, Zhernakova A (2018) The importance of cohort studies in the post-GWAS era. Nat Genet 50: 322-328

[402] Walters RG, Coin LJ, Ruokonen A, et al. (2013) Rare genomic structural variants in complex disease: lessons from the replication of associations with obesity. PLoS One 8: e58048

[403] Walters RG, Jacquemont S, Valsesia A, et al. (2010) A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. Nature 463: 671-675

[404] Franks PW, Pearson E, Florez JC (2013) Gene-Environment and Gene-Treatment Interactions in Type 2 Diabetes: Progress, pitfalls, and prospects. Diabetes Care 36: 1413-1421

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

[406] Kilpelainen TO, Qi L, Brage S, et al. (2011) Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. PLoS Med 8: e1001116

[407] Graff M, Scott RA, Justice AE, et al. (2017) Genome-wide physical activity interactions in adiposity - A meta-analysis of 200,452 adults. PLoS Genet 13: e1006528

[408] Justice AE, Winkler TW, Feitosa MF, et al. (2017) Genome-wide meta-analysis of 241,258 adults accounting for

smoking behaviour identifies novel loci for obesity traits. Nat Commun 8: 14977

[409] Karczewski KJ, Snyder MP (2018) Integrative omics for health and disease. Nat Rev Genet 19: 299-310

ABBREVIATIONS

Abbreviations

AIR	acute insulin response
BMI	body mass index
CIR	corrected insulin response
CVD	cardiovascular disease
GLP1	glucagon-like peptide 1
GSIS	glucose-stimulated insulin response
GWAS	genome-wide association study
HDL	high-density lipoprotein
HOMA	homeostasis-model assessment
hr	hour
HRC	haplotype reference consortium
IVGTT	intravenous glucose tolerance test
LD	linkage disequilibrium
LDL	low-density lipoprotein
MAF	minor allele frequency
MODY	maturity-onset diabetes of the young
OGTT	oral glucose tolerance test
RAF	risk allele frequency
SNP	single-nucleotide polymorphism
T2D	type 2 diabetes
WHR	waist-hip ratio

Gene names

ABCC8	ATP binding cassette subfamily C member 8
ABO	ABO, alpha 1-3-N-acetylgalactosaminyltransferase and alpha 1-3-galactosyltransferase
ACSL1	acyl-CoA synthetase long chain family member 1
ADAMTS9	ADAM metallopeptidase with thrombospondin type 1 motif 9
ADCY3	adenylate cyclase 3
ADCY5	adenylate cyclase 5
ANGPTL3	angiopoietin like 3
ANK1	ankyrin 1
ANKH	ANKH inorganic pyrophosphate transport regulator
ANKRD55	ankyrin repeat domain 55
AP3S2	adaptor related protein complex 3 subunit sigma 2
APOB	apolipoprotein B
APOC3	apolipoprotein C3
APOE	apolipoprotein E
ARAP1	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1
ARL15	ADP ribosylation factor like GTPase 15
BCAR1	BCAR1, Cas family scaffold protein
BCL11A	BCL11A, BAF complex component
BCL2	BCL2, apoptosis regulator
BDNF	brain derived neurotrophic factor
BPTF	bromodomain PHD finger transcription factor
C2CD4A	C2 calcium dependent domain containing 4A
C2CD4B	C2 calcium dependent domain containing 4B
CCND2	cyclin D2
CD300LG	CD300 molecule like family member g
CDC123	cell division cycle 123
CDKAL1	CDK5 regulatory subunit associated protein 1 like 1
CDKN2A	cyclin dependent kinase inhibitor 2A
CDKN2B	cyclin dependent kinase inhibitor 2B
CENPW	centromere protein W
CEP68	centrosomal protein 68
CMIP	c-Maf inducing protein
COBLL1	cordon-bleu WH2 repeat protein like 1
CREBRF	CREB3 regulatory factor
DGKB	diacylglycerol kinase beta
FADS1	fatty acid desaturase 1
FAF1	Fas associated factor 1
FAM13A	family with sequence similarity 13 member A
FAM19A2	family with sequence similarity 19 member A2, C-C motif chemokine like
FAM234A	family with sequence similarity 234 member A
FOLR3	folate receptor 3
FTO	FTO, alpha-ketoglutarate dependent dioxygenase
FUT2	fucosyltransferase 2
FUT6	fucosyltransferase 6
GCK	glucokinase
GCKR	glucokinase regulator
GIPR	gastric inhibitory polypeptide receptor
GLIS3	GLIS family zinc finger 3
GLP2R	glucagon like peptide 2 receptor
GPR151	G protein-coupled receptor 151
GPSM1	G protein signaling modulator 1
GRB10	growth factor receptor bound protein 10
GRB14	growth factor receptor bound protein 14
HHEX	hematopoietically expressed homeobox
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
HSD17B12	hydroxysteroid 17-beta dehydrogenase 12
IDE	insulin degrading enzyme
IGF1	insulin like growth factor 1
IGF2	insulin like growth factor 2
IGF2BP2	insulin like growth factor 2 mRNA binding protein 2
INS	insulin
IRS1	insulin receptor substrate 1
ITGA1	integrin subunit alpha 1
JADE2	jade family PHD finger 2
JAZF1	JAZF zinc finger 1

KCNJ11	potassium voltage-gated channel subfamily J member 11
KCNQ1	potassium voltage-gated channel subfamily Q member 1
KIF9	kinesin family member 9
KL	klotho
KLF14	Kruppel like factor 14
KLHL42	kelch like family member 42
LAMA1	laminin subunit alpha 1
LARGE1	LARGE xylosyl- and glucuronyltransferase 1
LDLR	low density lipoprotein receptor
LEPR	leptin receptor
LIPE	lipase E, hormone sensitive type
LPL	lipoprotein lipase
LPP	LIM domain containing preferred translocation partner in lipoma
MACF1	microtubule-actin crosslinking factor 1
MADD	MAP kinase activating death domain
MAEA	macrophage erythroblast attacher
MAP3K11	mitogen-activated protein kinase kinase kinase 11
MC4R	melanocortin 4 receptor
MLX	MLX, MAX dimerization protein
MNX1	motor neuron and pancreas homeobox 1
MPHOSPH9	M-phase phosphoprotein 9
MTHFR	methylenetetrahydrofolate reductase
MTMR3	myotubularin related protein 3
MTNR1B	melatonin receptor 1B
NAT2	N-acetyltransferase 2
NFAT5	nuclear factor of activated T cells 5
NOTCH2	notch 2
NPC1L1	NPC1 like intracellular cholesterol transporter 1
NRG4	neuregulin 4
NRXN3	neurexin 3
NTRK2	neurotrophic receptor tyrosine kinase 2
PAM	peptidylglycine alpha-amidating monooxygenase
PAX4	paired box 4
PCSK1	proprotein convertase subtilisin/kexin type 1
PCSK9	proprotein convertase subtilisin/kexin type 9
PDE3B	phosphodiesterase 3B
PDX1	pancreatic and duodenal homeobox 1
PEPD	peptidase D
PIM3	Pim-3 proto-oncogene, serine/threonine kinase
PLEKHA1	pleckstrin homology domain containing A1
PNPLA3	patatin like phospholipase domain containing 3
POC5	POC5 centriolar protein
POMC	proopiomelanocortin
PPARG	peroxisome proliferator activated receptor gamma
PRC1	protein regulator of cytokinesis 1
PROX1	prospero homeobox 1
PSMD6	proteasome 26S subunit, non-ATPase 6
PTPN9	protein tyrosine phosphatase, non-receptor type 9
RASGRP1	RAS guanyl releasing protein 1
RBMS1	RNA binding motif single stranded interacting protein 1
RNF6	ring finger protein 6
RREB1	ras responsive element binding protein 1
SH2B1	SH2B adaptor protein 1
SLC16A11	solute carrier family 16 member 11
SLC22A3	solute carrier family 22 member 3
SLC2A2	solute carrier family 2 member 2
SLC30A8	solute carrier family 30 member 8
SLC35D3	solute carrier family 35 member D3
SPRY2	sprouty RTK signaling antagonist 2
ST6GAL1	ST6 beta-galactoside alpha-2,6-sialyltransferase 1
TBC1D4	TBC1 domain family member 4
TCF7L2	transcription factor 7 like 2
TFAP2B	transcription factor AP-2 beta
THADA	THADA, armadillo repeat containing
TLE1	TLE family member 1, transcriptional corepressor
TLE4	TLE family member 4, transcriptional corepressor
TM6SF2	transmembrane 6 superfamily member 2
TMEM154	transmembrane protein 154
TMEM18	transmembrane protein 18
TP53INP1	tumor protein p53 inducible nuclear protein 1
TSPAN8	tetraspanin 8
TTLL6	tubulin tyrosine ligase like 6
UBE2E2	ubiquitin conjugating enzyme E2 E2
VEGFA	vascular endothelial growth factor A
VPS13C	vacuolar protein sorting 13 homolog C
WFS1	wolframin ER transmembrane glycoprotein
WSCD2	WSC domain containing 2
ZBED3	zinc finger BED-type containing 3
ZMIZ1	zinc finger MIZ-type containing 1
ZZEF1	zinc finger ZZ-type and EF-hand domain containing 1