Revealing the genetic roots of obesity and type 2 diabetes

Vliet-Ostaptchouk, Jana Vladimirovna van

Chapter 7

General discussion

General discussion

The aim of this thesis was to gain insight into genetic susceptibility to obesity and type 2 diabetes. The research was performed in a period in which major breakthroughs in the genetic studies of complex diseases took place. Recent advances in obesity and type 2 diabetes genetics have led to the identification of the first disease susceptibility genes and started to provide valuable information for understanding the disease pathogenesis. In this chapter, I will discuss our current model of the genetic architecture of obesity and type 2 diabetes, the potential implications and limitations of these recent discoveries, and provide future perspectives.

Genetic architecture of obesity and type 2 diabetes

Our insight into the genetics of both obesity and type 2 diabetes (T2D) has been revolutionized by genome-wide association studies (GWAS) in a number of ways (1). First, the recent GWAS findings have made clear that both obesity and T2D are much more heterogeneous and polygenic than previously believed (2-5). For years it was suggested that there might be two or three major genetic forms of T2D along with some minor forms. The GWAS results indicate however that T2D may represent genetically a large number of different genetical disorders (possibly as many as 50 or more), each resulting from a distinct combination of genetic variants and perhaps having different pathophysiology (2).

Second, the genetic effects of the currently identified disease risk variants are relatively small, with allelic odds ratios (OR) usually in the range of 1.1 – 1.3 (see **Table 3**, Introduction). Given the relatively large sample sizes of collaborative GWAS published to date, it is unlikely to find new common variants with effect sizes equal to or larger than those of *TCF7L2* or *FTO* (at least among populations of European ancestry). However, as study sample sizes continue to increase, we may find many more polymorphisms of modest effect sizes (OR 1.1 – 1.2) (4). Next, while the estimated heritability for both obesity and T2D is relatively high (about 30-70% for both disorders), only a small proportion of heritability can be explained by the combined effect of all currently identified susceptibility loci (i.e. less then 2% for obesity and as little as 5-10% for T2D (**Figure 1**)). So, the major question is: where is the remainder of the disease-associated genetic variation?

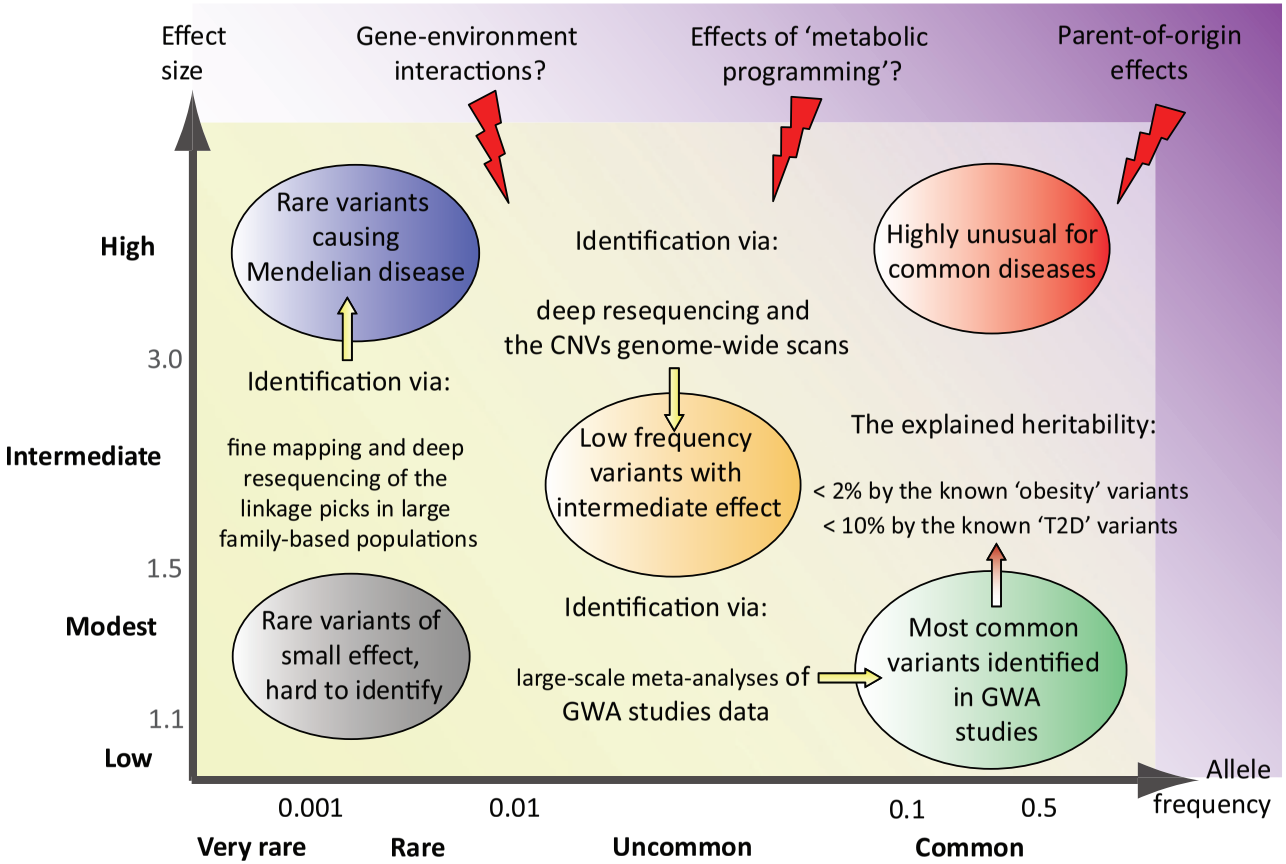

Figure 1: Genetic variants and disease susceptibility. See text for further details. Abbreviations: CNV, copy number variants; GWA, genome-wide association. Adapted from McCarthy MI et al., with permission, from Nat Gen Rev, REF. (1). © (2009) Macmillan Publishers Ltd.

Third, current findings suggest evidence for a possible overlap in the genetic determinants of monogenic and common forms of obesity and T2D. For example, the *MC4R* gene in which coding mutations are known to be the most frequent cause for severe obesity (6), has been also identified in GWAS for obesity (7, 8). In the same way, there is an overlap between the MODY genes (maturity onset diabetes of the young) such as *HNF4a* and *TCF2* genes, and the susceptibility to common T2D (9). These results suggest the hypothesis that these genes may have a "balancing effect" on the disorder: while the mutations that cause a substantial defect in gene function lead to highly penetrant forms of disease, more subtle genetic changes affecting the functions of these genes might play a role in the genetic predisposition to more common multifactorial forms of the disease.

Fourth, although the high co-occurrence rate of obesity and T2D may indicate the presence of shared predisposing genetic factors for these disorders, so far no real overlapping loci for both conditions have been identified (10). Despite of the fact that the obesity-associated gene *FTO* was found through a T2D GWAS, there is an ongoing discussion whether *FTO* can be considered as a risk variant for T2D. In the original study the gene association with diabetes was completely abolished after the adjustment for body mass index, indicating that the *FTO*-T2D relationship was mediated by an effect on adiposity. Recently an independent association between *FTO* and T2D was reported in the French MONICA study of 3400 diabetic patients (of whom only ~20% are obese) (11), suggesting that *FTO* variants could also be associated with diabetes-related phenotypes other than fat mass. Further investigation is necessary to draw any conclusions on the direct contribution of the *FTO* genetic variants to T2D risk. Also, in our work we extensively investigated whether some of the novel T2D loci – *TCF7L2* (Chapter 5) and *HHEX* (Chapter 6) – have an effect on body mass index (BMI), but we were also not able to find any evidence for such a relationship (12, 13).

Fifth, there is an indication of co-localization of the T2D loci with other common diseases. For example, several novel T2D susceptibility loci were also associated with rather unrelated diseases varying from prostate cancer to Crohn's disease (14). A region on chromosome 9p21.3 that harbours two independent signals for influencing the risk for T2D (15-17) was also found to be associated to cardiovascular disease (CVD) (18, 19). These risk variants are located near the protein-coding genes *CDKN2A* and *CDKN2B*, which are strongly implicated in the development of malignancy, and are associated with T2D and CVD independently from each other (20). While the association with CVD can be explained by a noncoding RNA *ANRIL* in this region which is expressed in tissues involved in atherosclerosis, the effect on the T2D risk might be explained via the

modulation of the *CDKN2A/B* function, thereby affecting β-cell regenerative capacity (21). Taking together, these examples of overlapping loci between different diseases may indicate the presence of variants within different regulatory domains that result in tissue- and disease-specific effects mediated through the same genes and pathways.

Missing heritability

The current findings that obesity and T2D genetics explain only a small proportion of the disease heritability, suggest that there are more genetic factors to be found. Where is the remainder of the heritability and can this in part be attributed to the limitations of the GWAS scans? Much of the discussion on missing heritability from the GWAS discoveries has focused on the bias of the current high-throughput genotyping platforms towards common variants – single nucleotide polymorphisms (SNPs) – with minor allele frequencies (MAF) over 5% (**Figure 1**) (22, 23). Thus, the choice in the GWAS design was partly driven by power constrains (i.e. there is no power to detect association to SNPs with low MAF). Moreover, the genetic model underlying common diseases, the so-called 'common disease – common variant' hypothesis, suggested that the joint action of several common variants influences disease susceptibility (24). Finally, SNPs that were not behaving well with respect to e.g. Hardy-Weinberg equilibrium, were also omitted from the commercial SNP arrays. Hence, current GWAS arrays are less effective to detect rare variants and copy number variations (CNV), such as insertions and deletions. Fortunately, a recent analysis of linkage disequilibrium between common CNVs and common SNPs predicts that common CNVs (or CNPs – copy number polymorphisms) can be successfully tagged by SNPs (25). In contrast, detection of rare CNVs requires sophisticated software.

Since structural variants occur commonly in the human genome (26), CNVs genome-wide scans may lead to new discoveries in the genetic architecture of obesity and type 2 diabetes. Indeed, the first screen for rare CNVs in patients with severe early-onset obesity has led to the identification of a deletion on 16p11.2 associated with highly penetrant familial severe early-onset obesity and severe insulin resistance (27). This deletion includes the *SH2B1* gene which was also identified in two obesity GWAS (28, 29) and an earlier candidate gene study (30) indicating the presence of both common variants influencing susceptibility to common obesity and more highly penetrant rare variants associated with severe forms of the disease in the same locus. Another promising new strategy for the detection of rare variants is deep sequencing

or next generation sequencing that is targeted at specific regions of interest or even the whole genome. The 1000 Genome Project (http://www.1000genomes.org/page.php), which will soon complete whole-genome sequencing of 1200 individuals will facilitate the building of a comprehensive catalogue of low frequency polymorphisms (1%<MAF<5%). It is expected that results from the 1000G project will lead to the development of new genome-wide chips that will more comprehensively capture the full repertoire of genetic variation in the genome. As the costs of sequencing have dropped dramatically over the past year it is anticipated that large scale whole-genome sequencing will come into reach within the next few years.

Another explanation for missing heritability might be the hidden effect of epigenetic modifications of DNA that can influence e.g. an individual's predisposition to metabolic syndrome. Evidence is emerging that molecular mechanisms of heritability may not be limited to DNA sequence differences. For example, a recent study of monozygotic and dizygotic twins has indicated that DNA methylation differences in the individual epigenomes can have a significant effect on phenotypic differences (31). These epigenetic modifications of DNA were shown to be stably inherited over a few generations in the absence of extensive genetic variation and with no selection (32). Recent data have also suggested that epigenetic processes play a key role in adaptive responses to nutritional and environmental factors during fetal and early life period that leads to a long term re-setting of cellular energy homeostasis, most probably via epigenetic modification of genes involved in a number of key regulatory pathways (33). This so called 'metabolic programming' is shown to play an important role in the individual's predisposition to develop obesity and type 2 diabetes, later in life. Interestingly, the hypothalamic genes, that are implicated to be strongly involved in common obesity, are reported to be affected by epigenetic modification during early-life events: for example, maternal food restriction or obesity have a programming effect on gene expression in the direction of reduced insulin sensitivity and increased production of glucose and fat (34). All these data highlight the potential impact of epigenetic variation on disease susceptibility and requires further investigation.

In addition, the role of parent-of-origin effects in the etiology of both obesity and T2D is currently of intense interest, but still largely unclear. Parent of origin effects refer to differential expression of a trait that is dependent on the sex of the parent from which transmission takes place. In mammals, these effects can be caused by imprinting, which is an epigenetic mechanism, as well as intrauterine effects or location of causative or susceptibility genes on the maternally inherited mitochondrial genome (35). The first genome-wide examination of parental-origin specific associations with

T2D has recently provided evidence of such mechanisms in disease aetiology (36). For instance, the effect of variant rs231362 located in the T2D susceptibility gene *KCNQ1* (37, 38) appeared to be limited to the maternally inherited allele. Also, a diabetes susceptibility variant on chromosome 11p15.5 (rs2334499) was shown to confer both T2D risk and protection depending on its parental origin: while the paternally inherited allele increased risk for T2D by 40%, the maternally inherited allele had a protective effect with odds ratio of 0.87 (36). These results indicate that the impact of parental origin on the genetically determined risk to disease is substantial. Therefore, further investigation of these effects in large family-based studies may lead to new discoveries in the genetic architecture of obesity and T2D.

Prediction of individual genetic risk of obesity and type 2 diabetes

Although some companies are already offering direct-to-consumer genetic testing for the risk of developing different diseases, the predictive value of many novel susceptibility variants for disease risk remains limited. In a recent study, 12 common genetic variants for BMI identified in recent GWA studies were combined to calculate predictive value for obesity risk (39). Together, all these polymorphisms had small but cumulative effects on obesity risk, with the *FTO* gene having the largest effect so far. Next, while each additional risk allele of any of these 12 disease susceptibility loci increased weight by 0.4 kg and increased the risk of obesity by 10%, the predictive value for obesity risk of the 12 SNPs combined was estimated to be only 2-3% (39). Similar results were obtained for T2D prediction. A genotype score for T2D, based on 18 established loci, was associated with a very modest increase of 12% in the relative risk of diabetes per risk allele. The inclusion of a combined genotype score to a model that was using common phenotypic risk factors (i.e. age, sex, family history, BMI, fasting glucose level, systolic blood pressure, high-density lipoprotein cholesterol level, and triglyceride level) resulted in the appropriate risk reclassification of 4% of the individuals only (40). These results suggest that the currently available genetic information for either obesity or T2D provides only little added value beyond classical clinical characteristics typically used to predict these conditions (39). Simulation and empirical studies indicate that with larger sample sizes or by large meta-analysis studies, accurate prediction of genetic risk will be possible, even if the causal variants or the underlying mechanisms by which they affect the disease risk are unknown (41). However, it is important to keep in mind

that although increasing the sample sizes of genome-wide association meta-analyses might lead to new obesity or T2D gene discoveries, these are very likely to have even smaller effect sizes than those already identified.

Lessons from the gene hunt for obesity and type 2 diabetes

The recent advances in the genetics of complex disease strongly indicate that a key factor of success lies in large-scale and meta-analysis data that usually requires international collaboration between different research groups. The expectations are that increasing sample sizes will lead to the identification of even more novel disease loci. The hypothesis is supported by the following calculations. If we assume that the undiscovered genetic variants have minor allele frequencies and effect sizes similar to those that have been identified (MAF >0.05 and OR 1.1 – 1.2, respectively), more than 800 genetic variants are required to explain the estimated heritability of 40%. In case the undiscovered genetic variants are mainly rare and with larger effect (MAF< 0.05 with ORs from 1.63 to 4.05), then ~85 variants are required (42). Indeed, four large meta-analysis for obesity and T2D and its clinical traits have identified a few new susceptibility loci to date (14, 15, 43, 44). Also, the results of the meta-analysis from the GIANT Consortium (Genetic Investigation of Anthropometric Traits) (29), an international collaborative initiative between eight research groups from across Europe and the USA that includes >100.000 individuals with GWAs data are expected soon.

The identification of the *INSIG2* gene in one of the first GWA scans for obesity initiated a discussion about the importance of study design that might be under-recognized in gene discovery and association replication. The initial finding of the association between *INSIG2*, a gene involved in cholesterol transport, and obesity was a long awaited discovery in the genetics of common disease (45). However, the results of the follow-up replications became a nightmare for the researchers of the initial study: while a few groups could replicated the original observation of the *INSIG2*-obesity association, many others failed to provide any evidence of the association (46-59). First, such inconsistency in the results was explained by a possible over-interpretation of the results in the original study, i.e. 'the winner's curse' (60). Later on, this debate has initiated a systematic meta-analysis of all available data to investigate the potential source of the between-study heterogeneity observed for the *INSIG2* association with obesity (61). The results revealed that the gene may be associated with extreme obesity (e.g., BMI≥37.5 kg/m^2) compared to normal controls meaning that the results

of the replication might be dependent on the proportion of extreme obesity included in the study. Thus, association with extreme degrees of obesity and consequently heterogeneous effects from different study designs may mask an underlying association when unaccounted for. The importance of study design might be underrecognized in gene discovery and association replication so far.

Next, it is important to keep in mind that both obesity and T2D are complex diseases with a broad range of disease phenotypes. Therefore, to improve our understanding of the genetic architecture of both disorders, specific obesity phenotypes or the quantitative traits for diabetes should be considered (62). A recent meta-analysis of 16 GWAS for genetic loci influencing central obesity and fat distribution has revealed two loci (*TFAP2B* and *MSRA*) associated with waist circumference and a further locus, near *LYPLAL1*, which shows gender-specific relationships with waist-hip ratio (43). Also, very recent large-scale meta-analyses of GWAS data have identified ten new loci associated with glycemic traits improving our understanding of the mechanisms involved in beta-cell function and glucose homeostasis (44, 63).

In addition, since the majority of GWA scans have so far been performed in populations of European descent, studies conducted in other ethnic populations may reveal additional candidate genes for obesity or T2D. As a matter of fact, the first GWA studies in East Asians have recently identified a previously unreported gene, *KCNQ1*, to be associated with T2D susceptibility (37, 38).

A systems biology approach in the studies of complex diseases genetics

Recently the concept of molecular networks as main drivers of common human diseases has been suggested (64). It is clear, that common diseases are the result of a complex interplay between DNA variation (both rare and common variations) and multiple additional factors such as age, gender, diet and exposure to environmental toxins. All these interacting factors affect the state of the entire molecular and physiological networks influencing the risk of disease (or phenotype in general) suggesting that disease states can be considered as different states of molecular networks. Hence, to be able to understand how any single gene may influence the risk to disease, it is important to understand which role this individual gene is playing in the context of molecular and physiological states associated with disease (64). This issue also was discussed in Chapter 2, in which we highlighted the importance of a complex approach

(i.e. the analysis of epistasis or interactions between genes) to understand the role of the whole hypothalamic system in obesity etiology. Other examples illustrating the importance of a systems approach are the studies of gene expression. As it has been recently shown, up to 80% of common genetic variants regulating gene expression are cell-type specific (65). Levels of gene expression too are dependent from age, metabolic conditions such as obesity and even fasting state (66-68). Thus, further investigation of gene expression and its regulation in different tissues in response to different environmental stimuli or diseases may improve our understanding of their connections with physiological states. To conclude, such a systems approach will require the integration of multiple levels of data, including genome-wide genotyping data on common and rare variants, whole-genome transcription data or RNA sequencing data (66), chromatin immunoprecipitation (ChIP) sequencing data and image data (69), data on mass spectrometry proteomics (70) (also, data on DNA-protein binding) and, finally, metabolomics data (i.e. metabolite levels in different states of physiological systems).

Future perspectives

Despite the progress in GWA findings over the last few years, we are still facing the 'dark matter' of genetic architecture of complex diseases such as obesity or type 2 diabetes, as: 1) most variants identified so far explain only a very small proportion of genetic predisposition to the diseases; 2) no causal variants that explain how the associated variants exert their effects on the disease pathogenesis were identified yet; 3) the predictive value of individual genetic risk for clinical practice remains poor, as 'linked to a telephone directory without addresses – some numbers are in it and maybe some names, but where those people live remains unknown'[1]. Admittedly, we have been too naïve by expecting that the GWAS will right away identify all the major genetic variants contributing to disease susceptibility. Nevertheless, the GWAS discoveries have provided new insights into genetic architecture of complex diseases that will lead to new research challenges in the next 5 years. As discussed above, deep resequencing of the genome, examination of the contribution of structural variants and epigenetic marks, large-scale meta-analysis GWAS, studies focused on a wide range of specific disease-associated phenotypes and on the investigation of the interactions with multiple environmental factors are the main directions for future research on obesity and T2D. The LifeLines project, a large population-based study of 165.000 participants

1 From the article by Lisa Nainggolan for TheHeart.org, June 2008

from the northern provinces of the Netherlands covering three generations (71), may represent a unique opportunity to investigate all above-mentioned aspects together. Importantly, the participants will be followed for a long period of time (~30 years) and a broad spectrum of biomedical data will be collected. Therefore, the LifeLines project will be essential in the providing in-depth insight into the origins and the course of complex diseases.