Europe PMC Funders Group
Author Manuscript
Nat Clin Pract Endocrinol Metab. Author manuscript; available in PMC 2021 February 23.

Published in final edited form as:
Nat Clin Pract Endocrinol Metab. 2008 March 01; 4(3): 156–63. doi:10.1038/ncpendmet0723.

Mechanisms of Disease: genetic insights into the etiology of type 2 diabetes and obesity

Cecilia M Lindgren, **Mark I McCarthy***
CM Lindgren is a Scientific Leadership Fellow at the Wellcome Trust Centre for Human Genetics at the University of Oxford, and MI McCarthy is the Robert Turner Professor of Diabetes, at the Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK

Summary

Until recently, progress in identification of the genetic variants influencing predisposition to common forms of diabetes and obesity has been slow, a sharp contrast to the large number of genes implicated in rare monogenic forms of both conditions. Recent advances have transformed the situation, however, enabling researchers to undertake well-powered scans able to detect association signals across the entire genome. For type 2 diabetes, the six high-density genome-wide association studies so far performed have extended the number of loci harboring common variants implicated in diabetes susceptibility into double figures. One of these loci, mapping to the fat mass and obesity associated gene (*FTO*), influences diabetes risk through a primary effect on fat mass, making this the first common variant known to influence weight and individual risk of obesity. These findings offer two main avenues for clinical translation. First, the identification of new pathways involved in disease predisposition—for example, those influencing zinc transport and pancreatic islet regeneration in the case of type 2 diabetes—offers opportunities for development of novel therapeutic and preventative approaches. Second, with continuing efforts to identify additional genetic variants, it may become possible to use patterns of predisposition to tailor individual management of these conditions.

Keywords

association; genetics; genome-wide; obesity; type 2 diabetes

Introduction

Type 2 diabetes mellitus (T2D) and obesity represent significant and growing burdens on society through their impact on morbidity and mortality. Despite this, the key mechanisms involved in the development of these related conditions remain imperfectly understood, and available strategies for both prevention and treatment inadequate.[1]

Correspondence: Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM), Oxford University, Churchill Hospital, Old Road Headington, Oxford OX3 7LJ, UK, mark.mccarthy@drl.ox.ac.uk.

Competing interests
The authors declared no competing interests.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

The strong familial clustering of both conditions emphasizes the central role of genetic factors in both disease initiation and progression. As a result, identification of the specific genes responsible offers a powerful tool for defining fundamental mechanisms involved in the development of disease (Box 1). Despite the attraction of the approach, however, success in susceptibility-gene identification has, for the most common multifactorial forms of diabetes and obesity, lingered far behind expectation. Only now, with the advent of far-more-powerful genetic tools applied to much larger study samples, is this approach delivering the anticipated insights into the pathogenesis of these conditions.

The principal obstacle to gene discovery efforts for common forms of obesity and diabetes lies in the recognition that these are multifactorial conditions, with individual susceptibility governed by the concerted effects of variation at many different genetic sites and subject to influence by a variety of environmental exposures. For such conditions, the impact of any single susceptibility variant is likely to be modest and variable, making both detection and replication extremely difficult.

The contrast with monogenic forms of diabetes and obesity is instructive. Whilst these constitute only a small proportion of the overall burden of disease, the strong genetic influence that characterizes these forms of disease has facilitated identification of the genes responsible. At least six different genes—including *GCK* (glucokinase gene), which encodes the β-cell's glucose sensor, and those encoding a number of pancreatic transcription factors —have been causally implicated in the development of maturity-onset diabetes of the young.[2] Mutations in several others—notably the genes *KCNJ11* (potassium inwardly rectifying channel, subfamily J, member 11 gene) and *ABCC8* (ATP-binding cassette, subfamily C, member 8 gene), which encode the components of the pancreatic β-cell K_{ATP} channel—are known to result in various types of neonatal diabetes mellitus.[3,4] Disruption of genes coding for key components of the satiety and weight-regulation pathway (such as those for leptin and its receptor) has been implicated in the genesis of rare syndromes of extreme childhood obesity.[5,6]

Not only have these discoveries shed new light on the mechanisms involved in the maintenance of metabolic and weight homeostasis, they have also rapidly translated into substantial improvements in clinical management, for particular patient groups at least.[4–7] Examples include the dramatic effects of leptin replacement in children with leptin-gene mutations,[7] and the benefit of transferring from insulin to sulfonylureas that is evident in most children with neonatal diabetes mellitus due to mutations in *KCNJ11*.[4]

In this article we discuss how systematic efforts to identify genes influencing susceptibility to multifactorial forms of obesity and diabetes are now leading to significant advances in our understanding of these conditions, and offering new opportunities for improvements in their clinical management.

Multifactorial Genetics: the Calm Before the Storm

Whilst traditional genetics tools (e.g. linkage mapping in multiply affected families) are well suited to finding the rare, penetrant mutations underlying monogenic forms of disease, they

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

are very poorly suited to identification of the common, low-effect-size variants currently thought to be particularly relevant in predisposition to multifactorial forms of disease.[8,9] For the latter, association mapping in large case–control sample sets represents a far more attractive strategy, albeit one compromised by the limited genomic extent over which association signals can be detected.[9]

Most early efforts to use association mapping to find genes influencing risk of T2D and obesity focused on particular ‘candidates’, selected on the basis that their known or presumed biological function was considered relevant to the disease in question. Too often, these studies were performed on underpowered samples, far too small to offer reasonable prospects for detection of variants with realistic effect sizes, and rarely did the findings replicate robustly. In retrospect, it is not difficult to understand why this approach yielded so few examples of genuine susceptibility variants.[10]

In the case of T2D, only two of the many hundreds of claimed associations arising from candidate-gene studies have stood the test of robust replication. The pro12Ala variant in *PPARG* (peroxisome proliferator-activated receptor γ gene; this encodes the target for thiazolidinediones)[11] and the Glu32Lys variant in *KCNJ11* (which encodes part of another diabetes therapeutic target, this time for sulfonylureas)[4] are both common single-nucleotide polymorphisms (SNPs) that have been shown to influence risk of diabetes in multiple studies. Their effect sizes are modest (each extra copy of a susceptibility allele increases the risk of disease by about 15-20%), however, and their contribution to the observed familial aggregation of diabetes limited. Since glucose homeostasis can be altered both by common variants within these genes and by drugs that act on the proteins they encode, it is no surprise to find that rare mutations in both genes are also associated with significant phenotypic effects.[2,4,12–14]

The harvest of equivalent efforts in obesity has been even more limited. The only locus contributing to a respectable proportion of cases of severe adult obesity is the one that includes *MC4R* (melanocortin 4 receptor gene).[6] The variants responsible are themselves rare, however, and have limited impact on variation in weight within the wider population.[5,6]

The Genome-Wide Association Era

The intrinsic limitation of the candidate-gene approach is that poor understanding of disease pathogenesis, itself the principal motivator behind efforts at susceptibility-gene discovery, inevitably frustrates efforts to define credible candidates. Indeed, the most novel and unexpected insights into disease pathogenesis are often likely to result from studies that avoid the obvious biological ‘suspects’ entirely.

The first clues that this ‘hypothesis-free’ approach to gene identification would succeed came from the discovery that variants within *TCF7L2* (transcription factor 7-like 2 gene) influenced susceptibility to T2D in Icelandic individuals.[15] This gene, which encodes a transcription factor active in the Wnt-signaling pathway (acomplex network of proteins known primarily for roles in embryogenesis and cancer) had no prior claims for candidacy with respect to T2D, and the susceptibility effect was detected following a trawl for

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

microsatellite associations within a large region of chromosome 10 that lay under a linkage 'peak'.[15] Fine-mapping efforts localized the causal variants to a cluster of intronic SNPs within *TCF7L2*. Replication studies confirmed that *TCF7L2* variants had a significantly stronger effect than those in *PPARG* and *KCJN11*, displaying a per-allele odds ratio of approximately 1.4 (Table 1).[15–17] Approximately one in ten Europeans is homozygous for the risk allele and has about twice as much risk of developing T2D as someone without the allele.[15–17]

Interestingly, despite this substantial effect size, the associated variants are not able to explain the strength of the linkage signal that prompted the mapping of this region: either this was a case of serendipity, or else, the gene also harbors rare, penetrant variants (as yet undetected) that underlie the linkage.[15] The strength of the evidence implicating variants within *TCF7L2* in susceptibility to T2D has naturally galvanized efforts to understand the mechanisms involved. Current evidence supports an effect mediated through disruption of pancreatic islet function,[18] possibly involving dysregulation of glucagon gene expression leading to reduced insulin secretion.[15,19]

The logical corollary of the success with *TCF7L2* involves extending the same kind of hypothesis-free analysis to the genome-wide scale, and a number of advances in genomics coalesced to make this possible. The first was the completion of a catalog of patterns of human genome sequence variation through the auspices of the International HapMap Consortium.[20] This catalog clearly demonstrated that, rather than having to type all of the 5–10 million SNPs that are commonly variant in non-African populations (the number is higher in samples of African origin, reflecting the greater genetic diversity on this ancestral continent), extensive correlations between neighboring SNPs ('linkage disequilibrium') mean that a far smaller number (around 500,000) will suffice.[21,22] In fact, around 80% of all common genome-wide variation will be well represented by such a subset.[21,22]

The second advance lay in new genotyping technologies that allowed very-high-throughput SNP-typing to be performed with high accuracy and much reduced per-genotype cost.[23] The third piece of the jigsaw was the realization by disease investigators that the likely spectrum of locus-effect sizes required the assembly and analysis of far larger sample sizes than previously deployed.[7] With these three items in place, it became possible to undertake truly genomewide surveys of association in appropriately large sample sets.

Novel Susceptibility Genes For Type 2 Diabetes and Obesity

In the past few months, the results of high-density genome-wide association data from over 19,000 individuals (7,142 patients with T2D and 11,996 controls, all of Northern European ances-tory) assembled by six different sets of investigators have been published (Table 2).[24–29] The consequence has been a bumper harvest of novel T2D-susceptibility regions—six to date. These regions contain *HHEX* (hematopoietically expressed homeobox gene); *IDE* (insulindegrading enzyme gene); *SLC30A8* (solute carrier family 30, member 8 gene); *FTO* (fat mass and obesity associated gene); *CDKAL1* (CDK5 regulatory subunit associated protein 1-like 1 gene); *CDKN2A* (cyclin-dependent kinase inhibitor 2A gene) and *CDKN2B*; and *IGF2BP2* (insulin-like growth factor 2 mRNA binding protein 2 gene).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

All of these susceptibility regions were independently replicated, and the same studies also provided further confirmation of the susceptibility effects of the three previously known genes (*PPARG*, *KCNJ11* and *TCF7L2*; Table 1).[24–29] In parallel, large-scale association efforts directed towards particular candidate pathways have contributed to the identification of a further two genes, *WFS1* (Wolfram syndrome 1 gene) and *HNF1B* (HNF1 homeobox B gene; this encodes hepatocyte nuclear factor 1-β), for which the evidence is commensurate with claims of genome-wide significance.[30–32]

The first two of the confirmed signals (those in the region around *HHEX* and *IDE*, and at *SLC30A8*) arose from the key study of Sladek and colleagues conducted in ~1,300 French individuals.[26] Though that original report advanced four potential loci, the fate of *EXT2* (exostoses 2 gene) and *LOC387761* as genuine T2D-susceptibility loci hangs in the balance, since evidence of disease association has not been corroborated.

The association with T2D of the locus with *HHEX* and *IDE* and the locus with *SLC30A8*, on the other hand, has now been extensively replicated.[24,27,29] The association signal involving *HHEX* and *IDE* resides within a 295-kilobase region of high linkage disequilibrium (flanking recombination hotspots provide a useful limit for the interval within which the etiological variants are likely to lie[33]) that contains not only *HHEX* (a strong biological candidate encoding a homeobox protein highly expressed in fetal and adult pancreas[34]) but also *KIF11* (kinesin family member 11 gene; this encodes a kinesin-interacting factor) and *IDE*. The homolog of *IDE* has been implicated in rodent models of diabetes, so it is clear that there is more than one good biological candidate in this region.[35,36]

This finding illustrates an important point of caution. The results from genome-wide association scans highlight particular variants, but any attempt to link those associations with dysfunction or dysregulation of a given gene using approximate mapping approaches alone must be considered very provisional. Not only may the association disclose a genetic interval that harbors several genes (as in the locus with *HHEX* and *IDE*), any of which might contain the causative variant(s), but also the growing evidence in support of remote regulatory effects may mean that functional effects are mediated through genes whose coding sequences lie well beyond the interval of immediate interest.[37]

Fortunately, neither of these caveats seems to apply to the second of the signals emerging from the study by Sladek and colleagues.[26] Not only is the *SLC30A8* signal limited to an interval containing a single, strong biological candidate (a zinc transporter expressed exclusively in insulin-producing β-cells and implicated in the maintenance of insulin granule function[38]), but the strongest association is with a SNP that changes the amino-acid structure of the encoded protein (Arg325Trp) and which may well be the variant that is responsible for the molecular effect (Table 1).

The next batch of novel candidate genes emerged from four studies published on the same date in April 2007.[24,25,27,29] Three of the groups responsible had worked cooperatively, sharing data that allowed them to achieve rapid confirmation of the promising findings emerging from any single study.[24,27,29] Three of the signals—involving regions containing

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

CDKAL1, *IGF2BP2* or *CDKN2A* plus *CDKN2B*—were widely replicated across the four studies. None was a strong biological candidate in advance of these studies (illustrating once again the value of genome-wide association studies to provide novel insights), and the pathogenetic mechanisms involved remain a matter of considerable conjecture (Table 1).

Having said that, at least two of the novel candidate regions (*CDKAL1*, and *CDKN2A* plus *CDKN2B*) include genes that encode proteins with putative or known inhibitor activity towards cyclin-dependent kinases, and for which independent evidence points towards effects on β-cell function or regeneration.[39–44] The notion that these genes act through disruption of β-cell functional mass should go some way to resolve controversies about the role played by reduced β-cell mass in T2D. Once again, however, confirming that associated SNPs genuinely act through a particular gene remains a substantial task, and the chromosome 9 (*CDKN2A*, *CDKN2B*) signal is emerging as one of the more challenging regions. Not only does the strongest T2D-association signal lie some distance from any of the genes in the region (approximately 125 kilobases away from *CDKN2B* for example), but the immediate region also contains a second, statistically independent signal for T2D (Table 1)[24,29] and a third, also completely independent, signal of association with coronary artery disease.[45]

Whilst some overlap of signals between T2D and coronary artery disease is not a surprise given their comorbidity and potential shared pathogenesis (e.g. through insulin resistance), the variants that impact on T2D have little effect on coronary disease risk, and vice versa. Understanding the molecular mechanisms whereby these different variants influence function and regulation of neighboring genes (such as *CDKN2A* and *CDKN2B*) to result in divergent phenotypic consequences is now a target for intensive research.

Alone of the T2D-susceptibility genes emerging from the high-density genome-wide association scans, *FTO* did not show compelling replication across the studies. The reason became apparent when it was realized that *FTO* variants were influencing T2D risk through a primary effect on weight and adiposity.[24,46] Since several of the other T2D scans had deliberately focused on lean T2D subjects, they were not well powered to detect susceptibility effects of this type.[25–29]

FTO represents the first compelling example of a common variant impacting on variation in weight and fat mass and on the individual risk of obesity (Table 1).[46–48] The effect of *FTO* is, in terms of overall variance, quite modest (less than 1% of the within-population variation in weight), but this amounts to a mean difference of ~2–3 kg between individuals carrying zero or two copies of the variant allele.[46] As with several of the other genes identified by genome-wide association studies, the involvement of *FTO* (or its genomic neighbors) in weight regulation came as a complete surprise, and, to date, almost nothing is known about the mechanisms whereby the variants influence fat mass. The hope is that work prompted by this discovery will add to understanding of the processes involved in weight regulation, and offer new therapeutic and preventative approaches to tackle the growing prevalence of obesity.

Europe PMC Funders Author Manuscripts

Biological Insights

One of the most striking features emerging from the genome-wide association data is the frequency with which variants in a given gene result in diverse disease phenotypes, often crossing conventional nosological boundaries. The signal on chromosome 9 provides one such example: not only are there variants impacting on T2D and coronary artery disease, but loss of *CDKN2A* function is also implicated in oncogenesis.[49]

A similar story is emerging with other diabetes-susceptibility genes. A large-scale association analysis of common variants within genes implicated in maturity-onset diabetes of the young had already provided suggestive evidence that a variant in intron 2 of *HNF1B* influenced risk of multifactorial T2D.[30] More recently, a genome-wide association scan for prostate cancer pinpointed exactly the same variant (the allele that was low risk for T2D increased the risk of prostate cancer).[32] Subsequent replication studies confirmed both the prostate cancer and T2D associations,[30,32] pointing towards a more general mechanism whereby variants that influence rates of cell division may have opposing effects on cancer predisposition and diabetes risk. The implications from the point of view of drug development are obvious.

Another emerging feature is that the predominant effect of most T2D-susceptibility alleles (excepting those in *FTO* and *PPARG*) is to disrupt islet function rather than insulin sensitivity (Table 1). This is consistent with evidence from monogenic forms of diabetes[2] (which disproportionately involve genes impacting β-cell function) and overall estimates of the relative heritability of the two processes.[50]

Conclusions

The ultimate measure of the success of these gene-discovery endeavors will be the extent to which the information gathered can be used to improve clinical management. There are two obvious ways in which this might come about. The first is that the identification of pathogenetic pathways offers new targets for drug discovery and the opportunity for novel therapeutic and preventative modalities: the findings at *SLC30A8* provide one example, which is being pursued by several companies. The second is the expectation that individual genetic profiling will provide a basis for personalized clinical care, for example through more-precise estimates of individual risk, or more efficient optimization of available therapeutic options.

Though recent discoveries have brought forward the day when these improvements become possible, the variants so far identified explain too small a proportion of overall risk (below 10%, even when combined) for viable clinical translation at present. With the current hectic pace of gene discovery, this may not remain the case for very long, especially as the search extends towards low frequency, intermediate penetrance variants, which may be particularly informative in this respect.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Acknowledgments

We would like to acknowledge the Throne-Holst Foundation and Diabetes UK. CM Lindgren is a Scientific Leadership Fellow at the Nuffield Department of Medicine at the University of Oxford.

References

1. Zimmet P, et al. Global and societal implications of the diabetes epidemic. Nature. 2001; 414:782–787. [PubMed: 11742409]
2. Owen KR, McCarthy MI. Genetics of type 2 diabetes. Curr Opin Genet Dev. 2007; 17:239–244. [PubMed: 17466512]
3. Vaxillaire M, et al. Kir6.2 mutations are a common cause of permanent neonatal diabetes in a large cohort of French patients. Diabetes. 2004; 53:2719–2722. [PubMed: 15448107]
4. Gloyn AL, et al. Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. N Engl J Med. 2004; 350:1838–1849. [PubMed: 15115830]
5. O'Rahilly S. Human obesity and insulin resistance: lessons from experiments of nature. Biochem Soc Trans. 2007; 35:33–36. [PubMed: 17212584]
6. Farooqi IS, O'Rahilly S. Genetic factors in human obesity. Obes Rev. 2007; 8(1):37–40.
7. Farooqi IS, et al. Effects of recombinant leptin therapy in a child with congenital leptin deficiency. N Engl J Med. 1999; 341:879–884. [PubMed: 10486419]
8. Wang WYS, et al. Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet. 2005; 6:109–118. [PubMed: 15716907]
9. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996; 273:1516–1517. [PubMed: 8801636]
10. Hirschhorn JN, Altshuler D. Once and again—issues surrounding replication in genetic association studies. J Clin Endocrinol Metab. 2002; 87:4438–4441. [PubMed: 12364414]
11. Altshuler D, et al. The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet. 2000; 26:76–80. [PubMed: 10973253]
12. Semple RK, et al. PPAR γ and human metabolic disease. J Clin Invest. 2006; 116:581–589. [PubMed: 16511590]
13. Sagen JV, et al. Permanent neonatal diabetes due to mutations in KCNJ11 encoding Kir6.2: patient characteristics and initial response to sulfonylurea therapy. Diabetes. 2004; 53:2713–2718. [PubMed: 15448106]
14. Barroso I, et al. Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension. Nature. 1999; 402:880–883. [PubMed: 10622252]
15. Grant SFA, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet. 2006; 38:320–323. [PubMed: 16415884]
16. Helgason A, et al. Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. Nat Genet. 2007; 39:218–225. [PubMed: 17206141]
17. Groves CJ, et al. Association analysis of 6,736 U.K. subjects provides replication and confirms TCF7L2 as a type 2 diabetes susceptibility gene with a substantial effect on individual risk. Diabetes. 2006; 55:2640–2644. [PubMed: 16936215]
18. Loos RJF, et al. TCF7L2 polymorphisms modulate proinsulin levels and β-cell function in a British Europid population. Diabetes. 2007; 56:1943–1947. [PubMed: 17416797]
19. Yi F, et al. TCF-4 mediates cell type-specific regulation of proglucagon gene expression by β-catenin and glycogen synthase kinase-3β. J Biol Chem. 2005; 280:1457–1464. [PubMed: 15525634]
20. International HapMap Consortium. A haplotype map of the human genome. Nature. 2005; 437:1299–1320. [PubMed: 16255080]
21. Pe'er I, et al. Evaluating and improving power in whole-genome association studies using fixed marker sets. Nat Genet. 2006; 38:663–667. [PubMed: 16715096]

Europe PMC Funders Author Manuscripts

22. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. Nat Genet. 2006; 38:659–662. [PubMed: 16715099]
23. Fan J-B, et al. Highly parallel genomic assays. Nat Rev Genet. 2006; 7:632–644. [PubMed: 16847463]
24. Zeggini E, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science. 2007; 316:1336–1341. [PubMed: 17463249]
25. Steinthorsdottir V, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. Nat Genet. 2007; 39:770–775. [PubMed: 17460697]
26. Sladek R, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature. 2007; 445:881–885. [PubMed: 17293876]
27. Scott LJ, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science. 2007; 316:1341–1345. [PubMed: 17463248]
28. Salonen JT, et al. Type 2 diabetes wholegenome association study in four populations: the DiaGen consortium. Am J Hum Genet. 2007; 81:338–345. [PubMed: 17668382]
29. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of Biomedical Research. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science. 2007; 316:1331–1336. [PubMed: 17463246]
30. Winckler W, et al. Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. Diabetes. 2007; 56:685–693. [PubMed: 17327436]
31. Sandhu MS, et al. Common variants in WFS1 confer risk of type 2 diabetes. Nat Genet. 2007; 39:951–953. [PubMed: 17603484]
32. Gudmundsson J, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet. 2007; 39:977–983. [PubMed: 17603485]
33. Myers S, et al. A fine-scale map of recombination rates and hotspots across the human genome. Science. 2005; 310:321–324. [PubMed: 16224025]
34. Bogue CW, et al. Hex expression suggests a role in the development and function of organs derived from foregut endoderm. Dev Dyn. 2000; 219:84–89. [PubMed: 10974674]
35. Hwang DY, et al. Significant change in insulin production, glucose tolerance and ER stress signaling in transgenic mice coexpressing insulin-siRNA and human IDE. Int J Mol Med. 2007; 19:65–73. [PubMed: 17143549]
36. Fakhrai-Rad H, et al. Insulin-degrading enzyme identified as a candidate diabetes susceptibility gene in GK rats. Hum Mol Genet. 2000; 9:2149–2158. [PubMed: 10958757]
37. ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007; 447:799–816. [PubMed: 17571346]
38. Chimienti F, et al. Identification and cloning of a β-cell-specific zinc transporter, ZnT-8localized into insulin secretory granules. Diabetes. 2004; 53:2330–2337. [PubMed: 15331542]
39. Wei F-Y, et al. Cdk5-dependent regulation of glucose-stimulated insulin secretion. Nat Med. 2005; 11:1104–1108. [PubMed: 16155576]
40. Ubeda M, et al. Inhibition of cyclin-dependent kinase 5 activity protects pancreatic β cells from glucotoxicity. J Biol Chem. 2006; 281:28858–28864. [PubMed: 16887799]
41. Moritani M, et al. Hypoplasia of endocrine and exocrine pancreas in homozygous transgenic TGF-β1. Mol Cell Endocrinol. 2005; 229:175–184. [PubMed: 15607541]
42. Mettus RV, Rane SG. Characterization of the abnormal pancreatic development, reduced growth and infertility in Cdk4 mutant mice. Oncogene. 2003; 22:8413–8421. [PubMed: 14627982]
43. Marzo N, et al. Pancreatic islets from cyclin-dependent kinase 4/R24C (Cdk4) knockin mice have significantly increased β cell mass and are physiologically functional, indicating that Cdk4 is a potential target for pancreatic β cell mass regeneration in Type 1 diabetes. Diabetologia. 2004; 47:686–694. [PubMed: 15298346]
44. Krishnamurthy J, et al. p16INK4a induces an age-dependent decline in islet regenerative potential. Nature. 2006; 443:453–457. [PubMed: 16957737]
45. Samani NJ, et al. Genomewide association analysis of coronary artery disease. N Engl J Med. 2007; 357:443–453. [PubMed: 17634449]

46. Frayling TM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science. 2007; 316:889–894. [PubMed: 17434869]
47. Scuteri A, et al. Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity related traits. PLoS Genet. 2007; 3:e115. [PubMed: 17658951]
48. Dina C, et al. Variation in FTO contributes to childhood obesity and severe adult obesity. Nat Genet. 2007; 39:724–726. [PubMed: 17496892]
49. Finkel T, et al. The common biology of cancer and ageing. Nature. 2007; 448:767–774. [PubMed: 17700693]
50. Stumvoll M, et al. Type 2 diabetes: principles of pathogenesis and therapy. Lancet. 2005; 365:1333–1346. [PubMed: 15823385]

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Box 1

Looking for clues: approaches used to identify genes contributing to disease.

Historically, linkage analysis was the main approach used to identify genes influencing strongly inherited (Mendelian, monogenic) disorders (such as maturity onset diabetes of the young). Linkage looks for evidence that particular markers co-segregate with disease status within families, and is only reasonably powered when disease status and genotype are strongly correlated. Linkage therefore performs far less well when the search is directed towards the common, low-penetrance variants thought to be involved in much of the susceptibility to common diseases.

Such variants are far less well correlated with disease status and, in such situations, association analyses generally perform better.[9,10] Association analyses typically look for evidence that a particular variant allele or genotype is overrepresented in disease cases compared with controls (though there are many variations on this theme).[9,10]

Until recently, such association analyses tended to focus on a limited number of variants chosen because they were considered likely to influence the function and/or regulation of a candidate gene of interest—that is, a gene which, on the basis of its known or presumed function, was thought to have a higher than average probability of involvement in disease pathogenesis. Though the candidate-gene approach has brought some success, the overall yield of confirmed disease-susceptibility genes gathered by this approach has been disappointing.

Recently, technological advances outlined in this article have allowed researchers to contemplate genome-wide association analyses,[8] which do not rely on prior assumptions regarding biological candidacy, and which are capable of identifying associations within genes that had little or no previous credibility as disease candidates. Because such genome-wide association studies involve simultaneous tests of association at hundreds of thousands of different markers, a high level of statistical significance has to be demanded if true signals are to be distinguished from chance.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Review Criteria

A search for original published articles focusing on research leading up to the advent of genome-wide association screens and the subsequent application of such studies in obesity and type 2 diabetes was performed in MEDLINE and PubMed. All papers identified were English-language, full-text papers. We also searched the reference lists of identified articles for further papers.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Key Points

- Genome-wide scans to detect evidence of association between common DNA variants and common disease are now technically possible
- Application of these methods to type 2 diabetes and obesity has led to identification of several novel genes that influence predisposition to these conditions and highlighted a number of entirely new etiological pathways
- The expectation is that these methods will contribute to the development of novel therapeutic strategies and, in the future, to efforts to use individual genetic profiling in the clinical management of these conditions

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Robustly replicated type 2 diabetes susceptibility genes identified to date.

Gene	Method of identification	Chromosome	SNP (amino acid change)	Risk-allele frequency in controls[a]	Per-allele odds ratio[a]	Suggested function
PPARG	Candidategene study	3p25	rs1801282 (Pro12Ala)	0.84	1.20	Nuclear hormone receptor transcription factor; adipocyte differentiation and function
KCNJ11	Candidategene study	11p15.1	rs5215 (Glu23Lys)	0.47	1.15	ATP-sensitive potassium channel; crucial for glucose-induced insulin secretion
TCF7L2	Fine-mapping of linkage peak	10q25.3	rs7901695	0.30	1.37	Wnt-signaling pathway in the islet; influences insulin and glucagon secretion
HHEX, IDE	GWAS	10q23-q25	rs5015480, rs1111875	0.54	1.14	Region has *HHEX* (pancreas, liver development) and *IDE* (may affect insulin action, secretion)
SLC30A8	GWAS	8q23.3	rs13266634 (Arg325Trp)	0.65	1.12	Zinc transporter; involved in islet insulin-granule function
FTO	GWAS	16q12.2	rs8050136	0.40	1.23[b]	Affects T2D via obesity, but the pathway is unknown
CDKAL1	GWAS	6p22.3	rs10946398	0.32	1.12	Possibly affects β-cell development, regeneration, function
CDKN2A, CDKN2B	GWAS	9p21	rs10811661, rs564398	0.86	1.20	Probable effect on β-cell development, regeneration, function
IGF2BP2	GWAS	3q28	rs4402960	0.32	1.17	Involved in translation of insulin-like growth factor II
WFS1	Candidatepathway study	4p16.1	rs10010131	0.60	1.19	Role in β-cell death and apoptosis
HNF1B	Candidategene study	17q21	rs757210	0.43	1.12	Transcription factor; involved in development of the pancreas, other tissues

[a] Approximate values of risk allele frequency and odds ratio are based on currently available data.

[b] The odds ratio for the effect of *FTO* on T2D is based on case–control samples not matched for BMI. Abbreviations: *CDKAL1*, CDK5 regulatory subunit associated protein 1-like 1 gene; *CDKN2A*, cyclin-dependent kinase inhibitor 2A gene; *FTO*, fat mass and obesity associated gene; GWAS, genome-wide association scan; *HHEX*, hematopoietically expressed homeobox gene; *HNF1B*, HNF1 homeobox B gene (this encodes hepatocyte nuclear factor 1-β); *IDE*, insulin-degrading enzyme gene; *IGF2BP2*, insulin-like growth factor 2 mRNA binding protein 2 gene; *KCNJ11*, potassium inwardly rectifying channel, subfamily J, member 11 gene; *PPARG*, peroxisome proliferator-activated receptor γ gene; *SLC30A8*, solute carrier family 30, member 8 gene; SNP, single-nucleotide polymorphism; T2D, type 2 diabetes mellitus; *TCF7L2*, transcription factor 7-like 2 gene; *WFS1*, Wolfram syndrome 1 gene.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 2

Overview of high-density genome-wide association scans for type 2 diabetes.

Study[a]	Number of cases[b]	Number of controls[c]	Sample source	Genotyping array	Reference
Wellcome Trust Case Control Consortium	1,924	2,938	UK	Affymetrix 500K GeneChip® (Affymetrix, Santa Clara, CA)	Zeggini *et al.*[24]
Diabetes Genetics Initiative	1,464	1,467	Finland and Sweden	Affymetrix 500K GeneChip®	Saxena *et al.*[29]
deCODE Genetics	1,399	5,275	Iceland	Illumina HumanHap300 BeadChip® (Illumina Inc, San Diego, CA)	Steinthorsdottir *et al.*[25]
Finland–US Investigation of NIDDM Genetics (FUSION)	1,161	1,174	Finland	Illumina HumanHap300 BeadChip®	Scott *et al.*[27]
Diabetes Gene Discovery Group	694	645	France	Illumina HumanHap300 BeadChip® plus Illumina Human I BeadChip® (100K)	Sladek *et al.*[26]
DiaGen	500	497	East Finland	Illumina HumanHap300 BeadChip®	Salonen *et al.*[28]

[a] Projects are listed in descending order of case sample size.

[b] Total number of cases 7,142.

[c] Total number of controls 11,996.