

# A strategy to search for common obesity and type 2 diabetes genes

Clara C. Elbers<sup>1,2</sup>, N. Charlotte Onland-Moret<sup>1</sup>, Lude Franke<sup>1</sup>, Anne G. Niehoff<sup>1,2</sup>, Yvonne T. van der Schouw<sup>2</sup> and Cisca Wijmenga<sup>1</sup>

<sup>1</sup> Complex Genetics Section, Department of Biomedical Genetics

<sup>2</sup> Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, PO Box 85060, 3508 AB Utrecht, the Netherlands

**Worldwide, the incidence of type 2 diabetes is rising rapidly, mainly because of the increase in the incidence of obesity, which is an important risk factor for this condition. Both obesity and type 2 diabetes are complex genetic traits but they also share some nongenetic risk factors. Hence, it is tempting to speculate that the susceptibility to type 2 diabetes and obesity might also partly be due to shared genes. By comparing all of the published genome scans for type 2 diabetes and obesity, five overlapping chromosomal regions for both diseases (encompassing 612 candidate genes) have been identified. By analysing these five susceptibility loci for type 2 diabetes and obesity, using six freely available bioinformatics tools for disease gene identification, 27 functional candidate genes have been pinpointed that are involved in eating behaviour, metabolism and inflammation. These genes might reveal a molecular link between the two disorders.**

## Obesity and type 2 diabetes

Worldwide, the incidence of type 2 diabetes (T2D) is rising rapidly, and there are already more than 170 million diabetic individuals. T2D results from the inability of the body to respond properly to the action of insulin produced by the pancreas; this results from impairment in both insulin sensitivity and insulin secretion [1]. T2D is a multifactorial disorder in which both genetic and nongenetic (environmental and life-style) factors have a role. The concordance rate of T2D among monozygotic twins is 76%, compared with 40% among dizygotic twins, providing convincing evidence that genetic factors contribute to the development of T2D. In addition, there is a 3.5-fold increased risk for a first-degree relative of a T2D patient to develop the disease [1]. Although both observations clearly imply that there is a genetic component to the disease, the model seems to be more complex, involving multiple genes and environmental factors. Several genes have been implicated that might contribute significantly to the risk of T2D, including the genes encoding peroxisome proliferator-activated receptor- $\gamma$  (*PPARG*), potassium inwardly-rectifying channel, subfamily J, member 11 (*KCNJ11*) [2] and, more recently, transcription factor 7-like 2 (*TCF7L2*) [3,4]. These genes are known to explain only part of the underlying genetic component, so there are

likely to be other, not yet identified, genes that are also important contributors to T2D susceptibility.

Besides a positive family history, T2D risk factors include ethnic background, age, hypertension, being overweight, increased abdominal fat and lack of physical exercise. Obesity is considered to be the most important risk factor for T2D, and the main one driving the current epidemic because 90% of T2D patients are obese. Worldwide obesity has reached epidemic proportions, with 300 million adults classified as clinically obese [based on data from the World Health Organization (WHO)], and 20% of these obese individuals suffer from T2D. Obesity is commonly assessed by the body mass index (BMI), defined as the weight in kilograms divided by the square of the height in metres ( $\text{kg}/\text{m}^2$ ). The definition of obesity is having a BMI value higher than 30, according to the WHO.

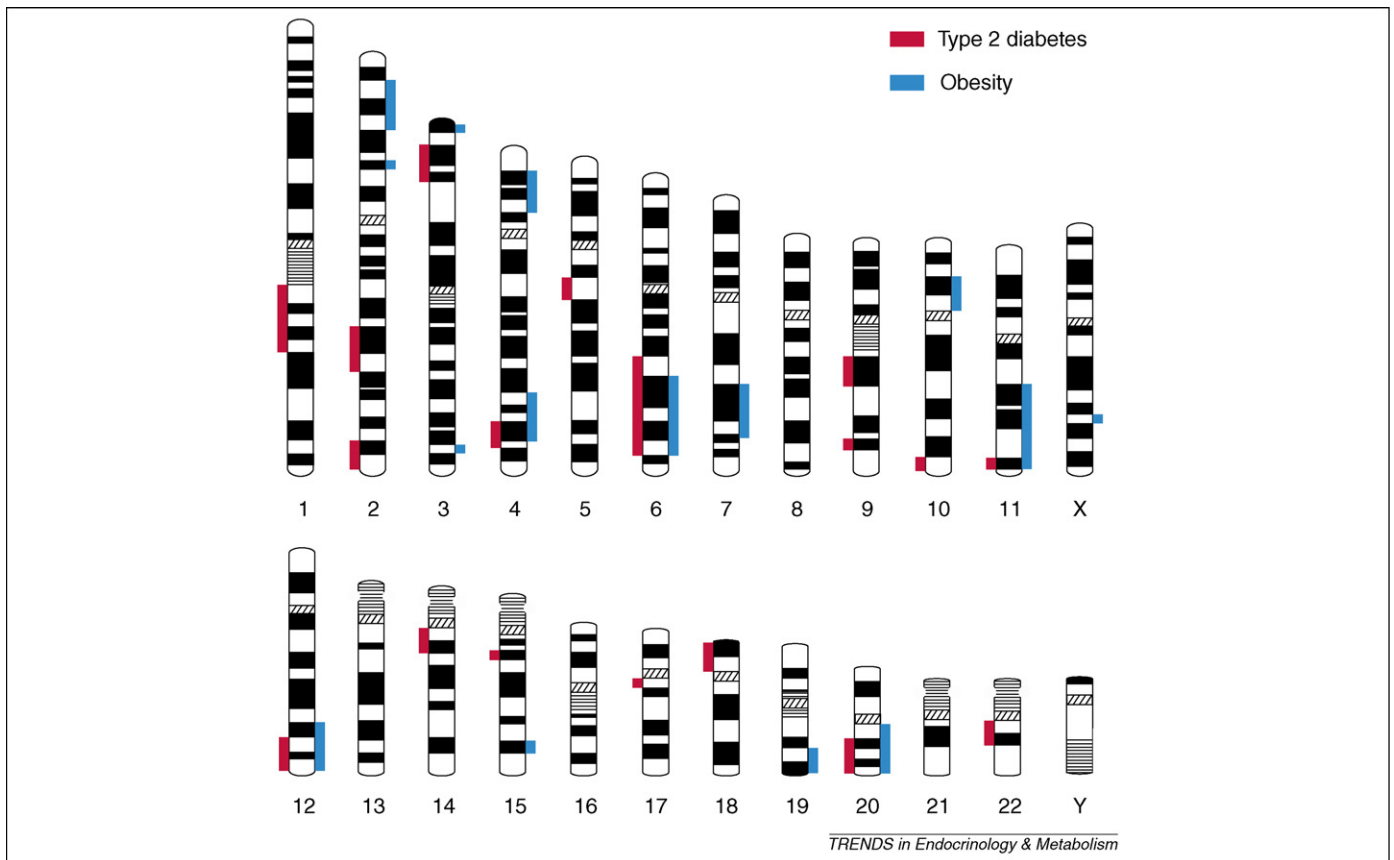
Clearly, obesity and T2D share some nongenetic factors because both are influenced by diet and physical inactivity. Also, both traits are characterized by insulin resistance, suggesting a shared pathology. It has been proposed that low-grade inflammation in visceral fat might be a potential mechanism whereby obesity results in insulin resistance [5].

It is tempting to speculate that the susceptibility to develop T2D and obesity is, in part, due to shared underlying genetic factors involved in common molecular mechanisms. Here, we explore whether there is any support for the hypothesis that T2D and obesity share some underlying susceptibility genes.

## Susceptibility loci for T2D and obesity

Genome scans are a useful approach to define susceptibility loci for disease candidate genes [6]. Genome-wide linkage scans involve the typing of families and sibling pairs using polymorphic markers that are positioned across the whole genome, followed by calculation of the degree of linkage of the marker to a disease trait. Positional candidate genes can then be identified by examining the regions around the peaks of linkage that are obtained. Linkage-based studies have implicated many susceptibility loci for both T2D and obesity. Bell *et al.* [7] collected and evaluated genome scans performed for obesity up until 2004, based on 31 papers. Since 2004, five additional genome scans for obesity have been published [8–12]. For T2D, 33 genome scans have been reported since 1996 [13–45] (Table S1 in the supplementary material online).

Corresponding author: Wijmenga, C. (t.n.wijmenga@med.uu.nl). Available online 28 November 2006.



**Figure 1.** Genetic linkage map for obesity and T2D. The red bars indicate susceptibility loci for T2D and the blue bars for obesity. Five chromosomal regions (4q32, 6q22–6q24, 11q24, 12q24 and 20q12–20q13) were found to be linked to both obesity and T2D. For detailed information on chromosomal locations, see the supplementary material online.

A total of 15 susceptibility loci for obesity (Figure 1; blue bars) and 18 susceptibility loci for T2D (Figure 1; red bars) fulfilled the inclusion criteria, according to the methods assessed in Box 1 (Table S2 in the supplementary material online gives a complete overview of the susceptibility loci reported for both T2D and obesity). Five of these chromosomal regions, encompassing a total of 612 genes, were found to be linked to both obesity and T2D.

#### Finding candidate genes using disease gene identification methods

Unfortunately, the data from linkage studies do not directly indicate the gene of interest, and identifying a potential gene is usually difficult [46,47] because linkage intervals can contain dozens to hundreds of candidate genes. To identify the gene of interest, a dense map of single nucleotide polymorphisms (SNPs) encompassing the candidate region needs to be tested for genetic association in large case-control studies. This strategy is based on the ‘common disease–common variant’ hypothesis [48] and assumes that common disease susceptibility alleles are involved in complex traits. If a risk polymorphism exists, it will either be genotyped directly or will be in strong linkage disequilibrium (LD) with one of the genotyped SNPs. The recent completion of phase I of the HapMap project has already resulted in a public database of more than 4 million common SNP variants across the genome [49]. This resource makes it feasible to carry out comprehensive genetic association studies needing a high

density of SNPs (usually more than one SNP in every ten kilobases). Consequently, large numbers of genotypes need to be generated, making these studies expensive, despite the recent drop in cost per genotype.

An attractive alternative strategy involves firstly prioritizing the positional candidate genes based on the function of the individual genes. Functional candidates are genes with products that can, in some way, be related to the pathogenesis of a disorder. Evidence for involvement of a gene in the disease process can, for example, comprise expression in the appropriate tissue or distribution of the gene product in a cell of interest. In the case of obesity, fat, adipose, hypothalamus, pituitary and gut are relevant tissues, whereas for T2D, pancreas, fat, adipose, liver, kidney, gut and muscle tissues are considered relevant. Evidence for involvement can also be drawn from a similarity to phenotypes associated with naturally occurring or engineered mutations in other species. For example, the *ob/ob* mouse, which has a defect in the leptin gene, is an excellent model for studying obesity. Strong mechanistic support can also come from a causal relationship of the phenotype with a variant nucleotide, with altered protein expression, or gene expression or function.

Until recently, investigating each gene separately for its likely involvement in the disease process had to be done manually with the aid of available public databases such as Online Mendelian Inheritance in Man (OMIM), Entrez and genome browsers, but now there are some promising bioinformatics tools (disease gene identification methods)

## Box 1. Methods

### Identification of susceptibility loci

The degree of evidence for all reported T2D loci was quantified as follows: a locus with a logarithm of odds ratio (LOD) score of 3 or more was considered significant, a LOD score between 2.2 and 3 was considered suggestive and a LOD score between 1 and 2.2 was considered nominal. For T2D, only those loci were included that were significant at least once, or were suggestive in at least one study and at least nominal in two or more studies. The inclusion of the second category of loci was based on a study by Wiltshire *et al.* [72], in which it was postulated that locus counting is a useful additional tool for the evaluation of genome scan data for complex trait loci. We used the same two criteria to determine the loci from the five papers published on obesity since 2004 and combined these loci with those from Bell *et al.* [7]. As obesity phenotypes, BMI, serum leptin levels, abdominal subcutaneous and visceral fat, and percentage body fat were included. All of these phenotypes were used as continuous quantitative traits, as well as with various cut-off levels.

### Gene identification methods

Prioritizer, Endeavour, DGP, Geneseeeker, G2D and PandS were combined to analyse the five overlapping T2D and obesity loci encompassing the 612 positional candidate genes. The 4q32, 6q22–6q24, 11q24, 12q24 and 20q12–20q13 loci were used as input in all systems. Additionally, Endeavour and PandS had to be trained with a set of genes. *ACDC*, *ADRA2A*, *ADRA2B*, *ADRB1*, *ADRB2*, *ADRB3*, *LEP*, *LEPR*, *NR3C1*, *UCP1*, *UCP2*, *UCP3* [7], *PPARG*, *KCNJ11* [2] and *TCF7L2* [3] are already known to be involved in T2D and/or obesity and were therefore used as training genes. Geneseeeker required disease-related tissue as input. Fat, adipose tissue, hypothalamus, pituitary, gut, liver, kidney and muscle were used. G2D needed OMIM number (#) as an additional input; for T2D, this was #125853. OMIM #601665 for obesity was not recognized by the program and was therefore omitted.

### Identification of candidate genes

Geneseeeker pinpoints genes that show expression in disease-related tissue. Therefore, we took all genes pinpointed by this method into consideration. All other tools produce rankings, and therefore the top 20 genes from each method were included for comparison.

A gene was considered to be interesting as a candidate gene if it was indicated by three or more of the tools. Because Endeavour, DGP and PandS partly use the same input information and show similar output, candidate genes were excluded if they were solely identified by these three methods.

Abbreviations: *ACDC*, adiponectin C1Q and collagen domain containing; *ADRA2A*, adrenoceptor  $\alpha$ 2A; *ADRA2B*, adrenoceptor  $\alpha$ 2B; *ADRB1*, adrenoceptor  $\beta$ 1; *ADRB2*, adrenoceptor  $\beta$ 2, surface; *ADRB3*, adrenoceptor  $\beta$ 3; *LEP*, leptin; *LEPR*, leptin receptor; *NR3C1*, nuclear receptor subfamily 3 group C member 1; *UCP1*, uncoupling protein 1 (mitochondrial proton carrier); *UCP2*, uncoupling protein 2 (mitochondrial proton carrier); *UCP3*, uncoupling protein 3 (mitochondrial proton carrier).

for disease gene identification. Six of these new tools are freely available online: Prioritizer [50], Geneseeeker [51], PROSPECTR and SUSPECTS (PandS) [52], Disease Gene Prediction (DGP) [53], Genes2Diseases (G2D) [54] and Endeavour [55] (Table 1). These tools use information extracted from public online databases, such as sequence data, medical literature, gene ontology and function annotation (GO), and information on biology, function and gene expression. Although the different tools have the same goal, they are based on different principles: Prioritizer ranks genes based on their functional interaction with genes on different susceptibility loci, assuming that

disease genes in a specific disorder are usually functionally related. Geneseeeker points to genes which are expressed in disease-related tissues. PROSPECTR differentiates between those genes that are likely to be involved in diseases and those which are not involved; it uses sequence-based features such as gene length, protein length and the percentage identity of homologues in other species. SUSPECTS scores candidate genes using PROSPECTR and also assesses the similarity between their annotation and that of already known disease genes. DGP assigns probabilities to genes that could indicate involvement in hereditary disease using parameters based on conservation, phylogenetic extent, protein length and paralogy. G2D scores all of the gene ontology GO terms according to their relevance to the disease. Endeavour is a software application for the computational prioritization of test genes based on a training set of genes already known to be involved in the disease of interest. The ranking of a test gene is based on its similarity with the training genes (Table 1).

These six tools were combined to analyse the five overlapping T2D and obesity loci encompassing the 612 positional candidate genes (Box 1). This strategy resulted in the identification of 27 candidate genes (Table 2), many of which qualify as ‘shared disease’ candidate genes when looking at their function.

When the 27 genes were grouped based on GO terms, it was found that five of the identified genes – that is, those encoding toll-like receptor 2 (*TLR2*), friend leukaemia virus integration 1 (*FLI1*), fibrinogen- $\beta$ - and - $\gamma$ -chain (*FGB*, *FGG*) and scavenger receptor class B member 1 (*SCARB1*) – were involved in immunity and defence. It is known that low-grade inflammation in the visceral fat of obese individuals causes insulin resistance and subsequently T2D. Although there is little evidence to date that causally links inflammation and obesity, there are recent data showing a role for inflammation in weight control [56]. Mice deficient in interleukin 18 show an increased food intake, resulting in accumulation of fat tissue. The insulin resistance seen in these interleukin 18-deficient mice is secondary to obesity and involves an enhanced expression of genes associated with gluconeogenesis in the liver, resulting from defective phosphorylation of signal transducer and activator of transcription 3 (STAT3). Hence, *TLR2*, *FLI1*, *FGG*, *FGB* and *SCARB1* might be interesting candidates with which to investigate further the link between satiety and inflammation, and are promising ‘shared disease’ candidate genes.

### The thrifty gene hypothesis

The group of 27 genes also contained ten genes involved in metabolism, sloth and gluttony (Table 3). This observation might point towards a role for thrifty genes as being important in the shared molecular basis of obesity and T2D.

Human evolution has shaped the genome of modern man, and one major driver of natural selection is famine [57]. During the periods of prolonged famine that plagued our early ancestors, a survival advantage would have been conferred by genes favouring the economical use and storage of energy; the so-called ‘thrifty’ genes. This theory was

**Table 1. Principles of the computational disease gene identification tools and detailed information on them**

Tool and website	Principle	Repositories used	Input
<b>Endeavour</b> <a href="http://www.esat.kuleuven.be/endeavour/">www.esat.kuleuven.be/endeavour/</a>	Ranks a test gene based on its similarity with the training genes	MEDLINE <sup>®</sup> abstracts, LocusLink, GO, InterPro and BIND protein–protein interactions, KEGG pathways, Microarray and EST-based expression data, TFBS, <i>cis</i> -regulatory modules, sequence similarity by BLAST	Susceptibility loci Known disease genes
<b>Prioritizer</b> <a href="http://www.prioritizer.nl">www.prioritizer.nl</a>	Ranks genes in different loci if they interact functionally	Protein–protein interactions from Reactome, HPRD and BIND. KEGG pathways, coexpression data from GEO, Y2H interactions in various species, GO	Susceptibility loci
<b>G2D</b> <a href="http://www.bork.embl-heidelberg.de/g2d/">www.bork.embl-heidelberg.de/g2d/</a>	Scores all GO terms according to their relevance to each disease	MEDLINE, Mesh-C and Mesh-D terms, GO, RefSeq collection	Susceptibility loci OMIM #
<b>Geneseeker</b> <a href="http://www.cmbi.kun.nl/GeneSeeker/">www.cmbi.kun.nl/GeneSeeker/</a>	Filters candidate genes based on expression and phenotypic data from humans and mice	OXFORD, MIMMAP, HGMD, MGD, Zuerich, PubMed, OMIM, UniProt, MLC, TBASE, GeneCards	Susceptibility loci Disease-related tissue
<b>DGP</b> <a href="http://cgg.ebi.ac.uk/services/dgp/">cgg.ebi.ac.uk/services/dgp/</a>	Assigns probabilities to genes, indicating their likelihood to mutate based on their sequence properties	OMIM, NCBI Locuslink, Ensembl database, CoGenT, SwissPROT, BLAST protein database	Susceptibility loci
<b>PROSPECTR and SUSPECTS</b> <a href="http://www.genetics.med.ed.ac.uk/suspects/">www.genetics.med.ed.ac.uk/suspects/</a>	Trained to differentiate between genes likely to be involved in a disease and those that are not, based on sequence features. (PROSPECTR) Scores similarity between the annotation and already known disease genes (SUSPECTS)	OMIM, HGMD, NCBI Homologene, InterPro protein domains, SwissPROT, Novartis, GO, Ensembl expression data	Susceptibility loci Known disease genes

<sup>a</sup>Abbreviations: BIND, biomolecular interaction network; BLAST, the basic local alignment search tool for finding regions of local similarity between sequences; *cis*-regulatory modules, combinations of TFBSs; CoGenT, complete genome-tracking database; Ensembl, project to produce and maintain automatic annotation on selected eukaryotic genomes; EST, expressed sequence tag; GeneCards, human gene-centric database; GEO, gene expression omnibus; HGMD, human gene mutation database; HPRD, human protein reference database; InterPro, database of protein families, domains and functional sites; KEGG, Kyoto encyclopaedia of genes and genomes; LocusLink, provides a single query interface to curated sequences and descriptive information about genetic loci; MEDLINE, medical literature analysis and retrieval system online; MGD, mouse genome database; MIMMAP, reformatted version of OMIM; MLC, mouse locus catalogue; NCBI, National Centre for Biotechnology Information; Novartis, Gene Expression Atlas of the Genomics Institute of the Novartis Research Foundation; OXFORD, human to mouse translation chromosomal locations; Reactome, curated database of biological processes in humans; RefSeq, reference sequence; SwissPROT, protein sequence database; TBASE, transgenic animals and targeted mutations; TFBS, transcription factor binding sites; UniProt; universal protein database; Y2H interactions, yeast two-hybrid interactions; Zuerich, chromosomal deletion and duplication map of malformation.

initially proposed by Neel [58], who focused on the efficient use of glucose as a biological fuel. He suggested that evolutionary pressure to preserve glucose for use by the brain during starvation led to a genetic propensity towards insulin resistance in peripheral tissue. In the Western world, food is, in general, easily available and plentiful, so these thrifty genes are maladaptive in modern society and might now contribute to susceptibility for obesity and T2D. However, although these evolutionary theories, focusing on the potential survival advantages of thrifty genes that are now maladaptive, are of great interest, they are speculative and difficult to prove [59]. Thriftiness can take many forms: (i) metabolic, an energy-sparing super-efficient metabolism; (ii) adipogenic, a propensity to rapid fat gain; (iii) physiological, an ability to switch off nonessential processes such as reproductive, thermogenic and immune capabilities; (iv) gluttony, a tendency to gorge when food is available and (v) sloth, a tendency to conserve energy through inactivity [60]. Physiological thriftiness is not very likely to cause obesity and/or T2D, because the ability to switch off nonessential processes during famine will not be clearly maladaptive during normal or excessive food intake. However, the other forms of thriftiness could be plausible characteristics of genes that are maladaptive in modern society. Although the ten genes presented in Table 3 would fit in with a ‘thrifty gene theory’ based on their function (because they might influence mechanisms such as energy reserve metabolism and eating behaviour), in-depth genetic studies are needed to prove

this theory. It would be interesting to compare the allele frequencies of these genes among different human populations with respect to food supply (past and present) and native climate. It would also be interesting to study the effect of long-term energy restriction on the expression of these genes [61].

### Candidate T2D and obesity genes

In addition to the inflammatory and thrifty genes mentioned above, the computational disease gene identification methods indicated some interesting genes already known to be associated with T2D or obesity. These include the genes encoding transcription factor 1 (*TCF1*), hepatocyte nuclear factor 4 $\alpha$  (*HNF4A*), opioid receptor  $\mu$ 1 (*OPRM1*), phosphoenolpyruvate carboxykinase 1 (*PCK1*), neuropeptide Y receptor 2 (*NPY2R*), ectonucleotide pyrophosphatase–phosphodiesterase 1 (*ENPP1*), guanine nucleotide binding protein,  $\alpha$  stimulating complex (*GNAS*), carboxypeptidase E (*CPE*) and nuclear receptor corepressor 2 (*NCOR2*). Many of these genes show genetic association to either T2D or obesity, or are in some other way functionally associated with either one of the disorders (Table 2).

Of particular interest is the *PCK1* gene, a main control point for the regulation of gluconeogenesis. A promoter SNP (–232C→G) in *PCK1* is associated with T2D [62]. The odds ratio (OR) for T2D among individuals with one or two copies of –232G compared with –232C homozygotes was 1.9 in a Canadian Oji-Cree Indian sample and 2.8 in a Caucasian sample. However, this association was not

**Table 2. List of 27 genes selected by the six disease gene identification tools, showing how the genes were prioritized and whether they are already genetically or functionally associated with obesity or T2D**

Gene	Tool						Genetic or functional association with type 2 diabetes or obesity	Refs
	Geneseeker	Prioritizer	DGB	PandS	G2D	Endeavour		
<b>Chr4</b>								
<i>NPY1R</i>			●	●	●	●	Involved in physiological regulation of energy balance	[7]
<i>NPY2R</i>		●	●	●		●	Lower BMI ( $P=0.017$ ) in 585T>C homozygous men. Lower allele and homozygosity frequency of 585T>C in obese and morbid obese men ( $P=0.007$ and $P=0.002$ , respectively). Obese T2D phenotype of ob/ob mice partially mediated by signalling through the NPY2R	[73] [74,75]
<i>NPY5R</i>		●	●			●	NPY5R mediates food intake in lean rats	[76]
<i>CPE</i>	●	●	●	●		●	CPE mRNA expression levels higher in visceral adipose tissue compared with subcutaneous adipose tissue in morbidly obese subjects ( $P<0.03$ ). BKS mice homozygous for the <i>CPE</i> fat mutation are severely obese, hyperinsulinaemic and hyperglycaemic	[77,78]
<i>FGB</i>		●	●	●		●		
<i>FGG</i>			●	●	●	●		
<i>CTSO</i>		●		●		●	Homologue of <i>CTSS</i> . Obesity is characterized by high circulating levels of CTSS ( $P<0.0001$ ); moreover, CTSS mRNA levels ( $P=0.006$ ) and protein levels ( $P<0.05$ ) in obese subcutaneous adipose tissue are increased compared with lean subjects	[79]
<i>GLRB</i>		●	●			●		
<i>TLR2</i>	●			●	●	●	Homologue of <i>TLR4</i> . TLR4 mRNA is induced during adipocyte differentiation and its level is enhanced in the fat tissues of obese db/db mice. TLR4 activation in 3T3-L1 adipocytes provokes insulin resistance	[80]
<b>Chr6</b>								
<i>GRM1</i>	●		●	●		●		
<i>OPRM1</i>			●	●	●	●	A single SNP (rs648007) in <i>OPRM1</i> associated with T2D ( $P=0.013$ ) in African-Americans	
<i>LATS1</i>				●	●	●		
<i>ESR1</i>	●	●	●	●		●		
<i>ENPP1</i>	●		●	●	●		Upregulation of <i>ENPP1</i> transcription in liver ( $P=0.025$ ) and brain ( $P=0.034$ ) of diabetic rabbits compared with controls. <i>ENPP1</i> 121Q allele predicts susceptibility to T2D in south Asians ( $P=0.01$ ) and Caucasians ( $P=0.003$ ). The Q allele of K121Q (OR = 1.6) and the T allele of rs997509 (OR = 4.7) were found to be associated with T2D in obese subjects from Poland. A risk haplotype was found to be associated with childhood obesity (OR = 1.69), adult morbid and moderate obesity (OR = 1.5 and OR = 1.37, respectively) and T2D (OR = 1.56)	[66–69]
<b>Chr11</b>								
<i>FLI1</i>				●	●	●		
<i>KCNJ5</i>		●	●	●	●	●	Homologue of <i>KCNJ11</i> . This gene contributes significantly to the risk of T2D	[2]
<i>ROBO4</i>		●		●		●		
<b>Chr12</b>								
<i>AACS</i>		●	●	●				
<i>SCARB1</i>			●	●		●	Key component in the reverse cholesterol transport pathway. Genetically associated with differences in insulin sensitivity in healthy subjects	[81]
<i>TCF1</i>	●	●	●	●			Responsible for MODY, an uncommon monogenetic form of early onset T2D	[2]
<i>NCOR2</i>		●	●	●		●	NCOR2 has an important role in the adipocyte by inhibiting adipocyte differentiation via repression of PPAR- $\gamma$ activity	[65]
<b>Chr20</b>								
<i>GNAS</i>	●		●	●		●	Targeted disruption of the <i>GNAS</i> gene in mice leads to distinct phenotypes in heterozygotes, depending on whether the maternal ( $m-/+$ ) or paternal ( $+/p-$ ) allele is mutated. $m-/+$ mice become obese, whereas $+/p-$ mice are thinner than normal. Both $m-/+$ and $+/p-$ mice have greater sensitivity to insulin, with low to normal fasting glucose levels, low fasting insulin levels, improved glucose tolerance and exaggerated hypoglycaemic response to administered insulin	[82]
<i>LAMA5</i>			●	●		●		

Table 2 (Continued)

Gene	Tool						Genetic or functional association with type 2 diabetes or obesity	Refs
	Geneseeker	Prioritizer	DGB	PandS	G2D	Endeavour		
<i>PCK1</i>		●	●	●		●	The OR for T2D among subjects with one or two copies of -232G compared with -232C homozygotes was found to be 1.9 in a native Canadian Oji-Cree sample and 2.8 in a Caucasian sample. This association was not replicated in a German Caucasian population. An <i>in vitro</i> experiment showed that the -232G construct was resistant to downregulation by insulin compared with a construct containing 232C	[62–64]
<i>PPGB</i>			●	●		●		
<i>HNF4A</i>	●		●	●	●		Responsible for MODY, an uncommon monogenetic form of early onset T2D	[2]
<i>PTGIS</i>			●	●	●			

Key and abbreviations: ●, gene is pointed out by Geneseeker in general or is ranked in the top 20 by all other bioinformatics tools; AACS, acetoacetyl-CoA synthetase; BKS mice, a diabetes-permissive inbred strain that arose through a genetic contamination of the C57BL/6J (B6) strain, probably by DBA/2J; CTSO, cathepsin O; CTSS, cathepsin S; *db/db* mice, mouse model for diabetic dyslipidaemia; ESR1, oestrogen receptor 1; GLRB, glycine receptor  $\beta$ ; GRM1, glutamate receptor metabotropic 1; LAMA5, laminin  $\alpha$ 5; LATS1, large tumour suppressor homologue; MODY, maturity onset diabetes of the young; NPY1R, neuropeptide Y receptor Y1; NPY2R, neuropeptide Y receptor Y2; NPY5R, neuropeptide Y receptor Y5; *ob/ob* mice, mice with a defect in the leptin gene, and an excellent model for studying obesity; OPRM1, opioid receptor  $\mu$ 1; OR, odds ratio; PPGB, protective protein for  $\beta$ -galactosidase; PTGIS, prostaglandin I2 (prostacyclin) synthase; ROBO4, roundabout homolog 4, magic roundabout (*Drosophila*).

replicated in a German Caucasian population [63]. An *in vitro* experiment in three different cell lines showed that the -232G construct was resistant to downregulation by insulin compared with a construct that did contain 232C [62]. The common assumption is that mutations in *PCK1* lead to excessive glucose production through hepatic gluconeogenesis. However, there is an alternative explanation, in which mutations at the *PCK1* locus could selectively affect *PCK1* expression in adipose tissue. This would result in changes in glyceroneogenesis that would affect the storage and release of fatty acids. Beale *et al.* [64] therefore proposed *PCK1* as a candidate gene for both T2D and obesity.

Another interesting gene is *NCOR2*. The protein encoded by this gene (*NCOR2*) interacts with PPAR- $\gamma$ . PPAR- $\gamma$  is an inflammatory factor that is also involved in the development of adipose tissue. Genetic studies have implicated *PPARG* in obesity as well as T2D. *NCOR2* also has an important role in the adipocyte, by inhibiting adipocyte differentiation through repression of PPAR $\gamma$  activity [65]. Hence, *NCOR2* is another interesting candidate gene to investigate for its susceptibility to both obesity and T2D.

*ENPP1* was also indicated by multiple gene identification systems. It encodes an inhibitor of the insulin receptor. Quantitative PCR analysis revealed a significant upregulation of *ENPP1* transcription in the

liver ( $P = 0.025$ ) and brain ( $P = 0.034$ ) of diabetic rabbits compared with controls [66]. The polymorphic *ENPP1* 121Q allele predicted genetic susceptibility to T2D in a south Asian sample ( $P = 0.01$ ) and a Caucasian sample ( $P = 0.003$ ) [67]. In obese Polish subjects, two SNPs were significantly associated with T2D: the Q allele of K121Q (OR = 1.6) and the T allele of rs997509 (OR = 4.7) [68]. A three-allele risk haplotype also showed an association with childhood obesity (OR = 1.69), adult morbid and moderate obesity (OR = 1.5 and OR = 1.37, respectively) and T2D (OR = 1.56;  $P = 0.00002$ ) [69]. This makes *ENPP1* the first example of a common genetic link between childhood obesity, adult obesity and T2D.

## Discussion

Complex traits such as obesity and T2D pose special challenges for genetic analyses because of gene–gene and gene–environment interactions, genetic heterogeneity and low penetrance of the individual genes. The heterogeneity means that it is difficult to generalize genome scan results over different populations and ethnicities. In addition, the multifactorial nature of complex traits assumes that the contribution of each of the susceptibility genes is likely to be small, and that only the joint effect of several susceptibility genes in combination with environmental factors will lead to disease [47]. It is therefore not surprising that large numbers of chromosomal regions have been implicated in disease susceptibility in T2D and/or obesity; analysing all of the individual positional candidate genes and loci will therefore be a daunting task. Applying computational disease gene identification methods can be hugely helpful in the hunt for complex disease genes.

Recently, using multiple bioinformatics tools, Tiffin *et al.* [70] analysed 9556 positional candidate genes that might be implicated in T2D and/or obesity. Their approach was different from the approach taken in the present study and resulted in a different list of genes, for the following reasons: (i) they included all susceptibility loci for each of the traits, which resulted in inclusion of nearly half of the genome. Hence, they will have indicated genes that could be responsible for either one of the disorders, whereas the

Table 3. Genes from the set of 27 genes pinpointed by multiple gene identification systems that are associated with the following thrifty GO annotation terms: metabolic, sloth and gluttony

Thrifty GO term	Genes
<b>Metabolic</b>	
Fatty acid metabolism	AACS, PTGIS
Gluconeogenesis	PCK1
Lipid, fatty acid and steroid metabolism	SCARB1, PTGIS, AACS
Glucose metabolism	NPY1R
Energy reserve metabolism	GNAS
Insulin processing	CPE
<b>Sloth</b>	
Locomotor behaviour	NPY1R, NPY2R
<b>Gluttony</b>	
Eating behaviour	NPY1R, NPY2R, NPY5R

approach in the present study focused only on the overlapping T2D and obesity loci. (ii) In addition, the present study used two extra computational methods (Prioritizer and Endeavour) that incorporate a wider range of biological data sources than do the other tools.

This study yielded an interesting list of candidate genes by investigating the overlapping chromosomal linkage regions for T2D and obesity, using a combination of six computational disease gene identification methods. Many of these identified genes are excellent candidates to study further for their role in the shared disease aetiology between obesity and T2D, and a few have already been genetically or functionally associated with both disorders (*ENPP1*, *NPY2R*). Although this cannot be taken as evidence that these computational methods work, it is tempting to assume that at least some of these genes might be true candidates, especially because this list includes genes belonging to the inflammatory pathway recently suggested to form an important molecular link between obesity and T2D [5]. Based on the candidate gene list presented here, we speculate that the molecular link between obesity and T2D extends beyond low-grade inflammation and might also involve thrifty genes. It will be interesting to see whether high-resolution SNP typing of these candidate genes in obese and/or T2D cohorts can be used to establish genetic association. In addition, these genes might be interesting candidates for identifying quantitative trait loci affecting obesity and T2D phenotypes in mice [71]. The list of 27 candidate genes and the associated pathways identified might also help in the further interpretation of genome-wide genetic association data for T2D and obesity.

#### Acknowledgements

The work described in this review was made possible through financial support from SenterNovem (IOP genomics grant IGE05012). We thank Jackie Senior for critical reading of the manuscript.

#### Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.tem.2006.11.003](https://doi.org/10.1016/j.tem.2006.11.003).

#### References

- Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (2003) Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 26 (Suppl. 1), S5-S20
- McCarthy, M.I. and Zeggini, E. (2006) Genetics of type 2 diabetes. *Curr. Diab. Rep.* 6, 147-154
- Grant, S.F. *et al.* (2006) Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat. Genet.* 38, 320-323
- Florez, J.C. *et al.* (2006) *TCF7L2* polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N. Engl. J. Med.* 355, 241-250
- Dandona, P. *et al.* (2004) Inflammation: the link between insulin resistance, obesity and diabetes. *Trends Immunol.* 25, 4-7
- Botstein, D. and Risch, N. (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* 33 (Suppl.), 228-237
- Bell, C.G. *et al.* (2005) The genetics of human obesity. *Nat. Rev. Genet.* 6, 221-234
- Chen, G. *et al.* (2005) A genome-wide scan for quantitative trait loci linked to obesity phenotypes among West Africans. *Int. J. Obes. (Lond)* 29, 255-259
- Cornes, B.K. *et al.* (2005) Sex-limited genome-wide linkage scan for body mass index in an unselected sample of 933 Australian twin families. *Twin Res. Hum. Genet.* 8, 616-632
- Lewis, C.E. *et al.* (2005) Sex-specific findings from a genome-wide linkage analysis of human fatness in non-Hispanic whites and African Americans: the HyperGEN study. *Int. J. Obes. (Lond)* 29, 639-649
- Norris, J.M. *et al.* (2005) Quantitative trait loci for abdominal fat and BMI in Hispanic-Americans and African-Americans: the IRAS Family study. *Int. J. Obes. (Lond)* 29, 67-77
- Wilson, S.G. *et al.* (2006) Linkage and potential association of obesity-related phenotypes with two genes on chromosome 12q24 in a female dizygous twin cohort. *Eur. J. Hum. Genet.* 14, 340-348
- Lindgren, C.M. *et al.* (2002) Genomewide search for type 2 diabetes mellitus susceptibility loci in Finnish families: the Botnia study. *Am. J. Hum. Genet.* 70, 509-516
- Aulchenko, Y.S. *et al.* (2003) A genome-wide search for genes involved in type 2 diabetes in a recently genetically isolated population from the Netherlands. *Diabetes* 52, 3001-3004
- Avery, C.L. *et al.* (2004) Linkage analysis of diabetes status among hypertensive families: the Hypertension Genetic Epidemiology Network study. *Diabetes* 53, 3307-3312
- Busfield, F. *et al.* (2002) A genomewide search for type 2 diabetes-susceptibility genes in indigenous Australians. *Am. J. Hum. Genet.* 70, 349-357
- Duggirala, R. *et al.* (1999) Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. *Am. J. Hum. Genet.* 64, 1127-1140
- Ehm, M.G. *et al.* (2000) Genomewide search for type 2 diabetes susceptibility genes in four American populations. *Am. J. Hum. Genet.* 66, 1871-1881
- Elbein, S.C. *et al.* (1999) A genome-wide search for type 2 diabetes susceptibility genes in Utah Caucasians. *Diabetes* 48, 1175-1182
- Ghosh, S. *et al.* (2000) The Finland-United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. I. An autosomal genome scan for genes that predispose to type 2 diabetes. *Am. J. Hum. Genet.* 67, 1174-1185
- Hanis, C.L. *et al.* (1996) A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nat. Genet.* 13, 161-166
- Hanson, R.L. *et al.* (1998) An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am. J. Hum. Genet.* 63, 1130-1138
- Hsueh, W.C. *et al.* (2003) Genome-wide and fine-mapping linkage studies of type 2 diabetes and glucose traits in the Old Order Amish: evidence for a new diabetes locus on chromosome 14q11 and confirmation of a locus on chromosome 1q21-q24. *Diabetes* 52, 550-557
- Hunt, K.J. *et al.* (2005) Genome-wide linkage analyses of type 2 diabetes in Mexican Americans: the San Antonio Family Diabetes/Gallbladder Study. *Diabetes* 54, 2655-2662
- Iwasaki, N. *et al.* (2003) Mapping genes influencing type 2 diabetes risk and BMI in Japanese subjects. *Diabetes* 52, 209-213
- Luo, T.H. *et al.* (2001) A genome-wide search for type II diabetes susceptibility genes in Chinese Hans. *Diabetologia* 44, 501-506
- Mahtani, M.M. *et al.* (1996) Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. *Nat. Genet.* 14, 90-94
- Martin, L.J. *et al.* (2002) A quantitative trait locus influencing type 2 diabetes susceptibility maps to a region on 5q in an extended French family. *Diabetes* 51, 3568-3572
- Mori, Y. *et al.* (2002) Genome-wide search for type 2 diabetes in Japanese affected sib-pairs confirms susceptibility genes on 3q, 15q, and 20q and identifies two new candidate loci on 7p and 11p. *Diabetes* 51, 1247-1255
- Nawata, H. *et al.* (2004) Genome-wide linkage analysis of type 2 diabetes mellitus reconfirms the susceptibility locus on 11p13-p12 in Japanese. *J. Hum. Genet.* 49, 629-634
- Ng, M.C. *et al.* (2004) Genome-wide scan for type 2 diabetes loci in Hong Kong Chinese and confirmation of a susceptibility locus on chromosome 1q21-q25. *Diabetes* 53, 1609-1613
- Parker, A. *et al.* (2001) A gene conferring susceptibility to type 2 diabetes in conjunction with obesity is located on chromosome 18p11. *Diabetes* 50, 675-680

- 33 Permutt, M.A. *et al.* (2001) A genome scan for type 2 diabetes susceptibility loci in a genetically isolated population. *Diabetes* 50, 681–685
- 34 Pratley, R.E. *et al.* (1998) An autosomal genomic scan for loci linked to prediabetic phenotypes in Pima Indians. *J. Clin. Invest.* 101, 1757–1764
- 35 Rotimi, C.N. *et al.* (2004) A genome-wide search for type 2 diabetes susceptibility genes in West Africans: the Africa America Diabetes Mellitus (AADM) Study. *Diabetes* 53, 838–841
- 36 Sale, M.M. *et al.* (2004) A genome-wide scan for type 2 diabetes in African-American families reveals evidence for a locus on chromosome 6q. *Diabetes* 53, 830–837
- 37 Silander, K. *et al.* (2004) A large set of Finnish affected sibling pair families with type 2 diabetes suggests susceptibility loci on chromosomes 6, 11, and 14. *Diabetes* 53, 821–829
- 38 van Tilburg, J.H. *et al.* (2003) A genome-wide scan in type 2 diabetes mellitus provides independent replication of a susceptibility locus on 18p11 and suggests the existence of novel loci on 2q12 and 19q13. *J. Clin. Endocrinol. Metab.* 88, 2223–2230
- 39 Vionnet, N. *et al.* (2000) Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27-qter and independent replication of a type 2-diabetes locus on chromosome 1q21-q24. *Am. J. Hum. Genet.* 67, 1470–1480
- 40 Watanabe, R.M. *et al.* (2000) The Finland-United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. II. An autosomal genome scan for diabetes-related quantitative-trait loci. *Am. J. Hum. Genet.* 67, 1186–1200
- 41 Wiltshire, S. *et al.* (2001) A genomewide scan for loci predisposing to type 2 diabetes in a UK population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am. J. Hum. Genet.* 69, 553–569
- 42 Xiang, K. *et al.* (2004) Genome-wide search for type 2 diabetes/impaired glucose homeostasis susceptibility genes in the Chinese: significant linkage to chromosome 6q21-q23 and chromosome 1q21-q24. *Diabetes* 53, 228–234
- 43 Zhao, J.Y. *et al.* (2005) An autosomal genomic scan for loci linked to type 2 diabetes in northern Han Chinese. *J. Mol. Med.* 83, 209–215
- 44 Reynisdottir, I. *et al.* (2003) Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34-q35.2. *Am. J. Hum. Genet.* 73, 323–335
- 45 Frayling, T.M. *et al.* (2003) Young-onset type 2 diabetes families are the major contributors to genetic loci in the Diabetes UK Warren 2 genome scan and identify putative novel loci on chromosomes 8q21, 21q22, and 22q11. *Diabetes* 52, 1857–1863
- 46 Carlson, C.S. *et al.* (2004) Mapping complex disease loci in whole-genome association studies. *Nature* 429, 446–452
- 47 Glazier, A.M. *et al.* (2002) Finding genes that underlie complex traits. *Science* 298, 2345–2349
- 48 Becker, K.G. (2004) The common variants/multiple disease hypothesis of common complex genetic disorders. *Med. Hypotheses* 62, 309–317
- 49 International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437, 1299–1320
- 50 Franke, L. *et al.* (2006) Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. *Am. J. Hum. Genet.* 78, 1011–1025
- 51 van Driel, M.A. *et al.* (2005) GeneSeeker: extraction and integration of human disease-related information from web-based genetic databases. *Nucleic Acids Res.* 33, W758–W761
- 52 Adie, E.A. *et al.* (2006) SUSPECTS: enabling fast and effective prioritization of positional candidates. *Bioinformatics* 22, 773–774
- 53 Lopez-Bigas, N. and Ouzounis, C.A. (2004) Genome-wide identification of genes likely to be involved in human genetic disease. *Nucleic Acids Res.* 32, 3108–3114
- 54 Perez-Iratxeta, C. *et al.* (2002) Association of genes to genetically inherited diseases using data mining. *Nat. Genet.* 31, 316–319
- 55 Aerts, S. *et al.* (2006) Gene prioritization through genomic data fusion. *Nat. Biotechnol.* 24, 537–544
- 56 Netea, M.G. *et al.* (2006) Deficiency of interleukin-18 in mice leads to hyperphagia, obesity and insulin resistance. *Nat. Med.* 12, 650–656
- 57 Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life*, John Murray
- 58 Neel, J.V. (1962) Diabetes mellitus: a 'thrifty' genotype rendered detrimental by 'progress'? *Am. J. Hum. Genet.* 14, 353–362
- 59 Lazar, M.A. (2005) How obesity causes diabetes: not a tall tale. *Science* 307, 373–375
- 60 Prentice, A.M. *et al.* (2005) Insights from the developing world: thrifty genotypes and thrifty phenotypes. *Proc. Nutr. Soc.* 64, 153–161
- 61 Kagawa, Y. *et al.* (2002) Single nucleotide polymorphisms of thrifty genes for energy metabolism: evolutionary origins and prospects for intervention to prevent obesity-related diseases. *Biochem. Biophys. Res. Commun.* 295, 207–222
- 62 Cao, H. *et al.* (2004) Promoter polymorphism in PCK1 (phosphoenolpyruvate carboxykinase gene) associated with type 2 diabetes mellitus. *J. Clin. Endocrinol. Metab.* 89, 898–903
- 63 Gouni-Berthold, I. *et al.* (2006) Association of the promoter polymorphism -232C/G of the phosphoenolpyruvate carboxykinase gene (PCK1) with type 2 diabetes mellitus. *Diabet. Med.* 23, 419–425
- 64 Beale, E.G. *et al.* (2004) Disregulated glyceroneogenesis: PCK1 as a candidate diabetes and obesity gene. *Trends Endocrinol. Metab.* 15, 129–135
- 65 Cohen, R.N. (2006) Nuclear receptor corepressors and PPAR $\gamma$ . *Nucl. Recept. Signal.* 4, e003
- 66 Eller, P. *et al.* (2006) Hepatic ENPP1 expression is induced in diabetic rabbits. *Mamm. Genome* 17, 886–891
- 67 Abate, N. *et al.* (2005) ENPP1/PC-1 K121Q polymorphism and genetic susceptibility to type 2 diabetes. *Diabetes* 54, 1207–1213
- 68 Bochenski, J. *et al.* (2006) New polymorphism of ENPP1 (PC-1) is associated with increased risk of type 2 diabetes among obese individuals. *Diabetes* 55, 2626–2630
- 69 Meyre, D. *et al.* (2005) Variants of ENPP1 are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes. *Nat. Genet.* 37, 863–867
- 70 Tiffin, N. *et al.* (2006) Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes. *Nucleic Acids Res.* 34, 3067–3081
- 71 Singer, J.B. *et al.* (2004) Genetic dissection of complex traits with chromosome substitution strains of mice. *Science* 304, 445–448
- 72 Wiltshire, S. *et al.* (2002) Evaluating the results of genomewide linkage scans of complex traits by locus counting. *Am. J. Hum. Genet.* 71, 1175–1182
- 73 Hung, C.C. *et al.* (2004) Studies of the peptide YY and neuropeptide Y2 receptor genes in relation to human obesity and obesity-related traits. *Diabetes* 53, 2461–2466
- 74 Lavebratt, C. *et al.* (2006) Common neuropeptide Y2 receptor gene variant is protective against obesity among Swedish men. *Int. J. Obes. (Lond)* 30, 453–459
- 75 Sainsbury, A. *et al.* (2002) Y2 receptor deletion attenuates the type 2 diabetic syndrome of ob/ob mice. *Diabetes* 51, 3420–3427
- 76 Wyss, P. *et al.* (1998) Stimulation of feeding in lean but not in obese Zucker rats by a selective neuropeptide Y Y5 receptor agonist. *Neuroreport* 9, 2675–2677
- 77 Ramis, J.M. *et al.* (2002) Carboxypeptidase E and thrombospondin-1 are differently expressed in subcutaneous and visceral fat of obese subjects. *Cell. Mol. Life Sci.* 59, 1960–1971
- 78 Collin, G.B. *et al.* (2005) Genetic modifiers interact with Cpefat to affect body weight, adiposity, and hyperglycemia. *Physiol. Genomics* 22, 182–190
- 79 Taleb, S. *et al.* (2005) Cathepsin S, a novel biomarker of adiposity: relevance to atherogenesis. *FASEB J.* 19, 1540–1542
- 80 Song, M.J. *et al.* (2006) Activation of Toll-like receptor 4 is associated with insulin resistance in adipocytes. *Biochem. Biophys. Res. Commun.* 346, 739–745
- 81 Perez-Martinez, P. *et al.* (2005) A polymorphism exon 1 variant at the locus of the scavenger receptor class B type I (SCARB1) gene is associated with differences in insulin sensitivity in healthy people during the consumption of an olive oil-rich diet. *J. Clin. Endocrinol. Metab.* 90, 2297–2300
- 82 Yu, S. *et al.* (2001) Increased insulin sensitivity in G $\alpha$ s knockout mice. *J. Biol. Chem.* 276, 19994–19998