

UCLA

UCLA Previously Published Works

Title

Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks.

Permalink

<https://escholarship.org/uc/item/65z0w2c2>

Journal

Nature genetics, 41(4)

ISSN

1061-4036

Authors

Yang, Xia
Deignan, Joshua L
Qi, Hongxiu
et al.

Publication Date

2009-04-01

DOI

10.1038/ng.325

Peer reviewed



Published in final edited form as:

Nat Genet. 2009 April ; 41(4): 415–423. doi:10.1038/ng.325.

Validation of Candidate Causal Genes for Abdominal Obesity Which Affect Shared Metabolic Pathways and Networks

Xia Yang¹, Joshua L. Deignan¹, Hongxiu Qi¹, Jun Zhu², Su Qian³, Judy Zhong², Gevork Torosyan⁴, Sana Majid⁴, Brie Falkard⁴, Robert R. Kleinhanz², Jenny Karlsson⁶, Lawrence W. Castellani¹, Sheena Mumick³, Kai Wang², Tao Xie², Michael Coon², Chunsheng Zhang², Daria Estrada-Smith⁴, Charles R. Farber¹, Susanna S. Wang⁴, Atila Van Nas⁴, Anatole Ghazalpour⁴, Bin Zhang², Douglas J. MacNeil³, John R. Lamb², Katrina M. Dipple⁴, Marc L. Reitman⁵, Margarete Mehrabian¹, Pek Y. Lum², Eric E. Schadt², Aldons J. Lusis^{1,4}, and Thomas A. Drake⁶

¹Department of Medicine, David Geffen School of Medicine at UCLA, University of California, Los Angeles, California, 90095, USA

²Rosetta Inpharmatics, LLC, a Wholly Owned Subsidiary of Merck & Co. Inc., Seattle, Washington, 98109, USA

³Department of Metabolic Disorders, Merck & Co. Inc., Rahway, New Jersey, 07065, USA

⁴Department of Human Genetics, David Geffen School of Medicine at UCLA, University of, California, Los Angeles, California, 90095, USA

⁵Department of Clinical Pharmacology, Merck & Co. Inc., Rahway, New Jersey, 07065, USA

⁶Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at, UCLA, University of California, Los Angeles, California, 90095, USA

Abstract

A major task in dissecting the genetics of complex traits is to identify causal genes for disease phenotypes. We previously developed a method to infer causal relationships among genes through the integration of DNA variation, gene transcription, and phenotypic information. Here we validated our method through the characterization of transgenic and knockout mouse models of candidate genes that were predicted to be causal for abdominal obesity. Perturbation of eight out of the nine genes, with *Gas7*, *Me1* and *Gpx3* being novel, resulted in significant changes in obesity related traits. Liver expression signatures revealed alterations in common metabolic pathways and

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence to: Thomas A. Drake.

Author contributions

T.A. Drake, A.J. Lusis, E.E. Schadt, P.Y. Lum, X. Yang, K.M. Dipple, M.L. Reitman, and D.J. MacNeil designed the study. X. Yang, H. Qi, G. Torosyan, S. Majid, B. Falkard, S. Qian, L.W. Castellani, D. Estrada Smith, S. Mumick, S. Wang, A. van Nas, A. Ghazalpour, M. Mehrabian, and C. R. Farber, performed the experiments. X. Yang, J. Deignan, J. Zhu, G. Torosyan, J. Karlsson, K. Wang, J. Lamb, T. Xie, M. Coon, C. Zhang, and B. Zhang participated in data analysis. X. Yang, J. Deignan, J. Zhu, S. Qian, and J. Zhong wrote the manuscript, with advice and editing from T.A. Drake, A.J. Lusis, and E.E. Schadt. R. Kleinhanz coordinated the study. T.A. Drake, the corresponding author, certifies that all authors have agreed to all the content in the manuscript, including the data as presented.

networks contributing to abdominal obesity and overlapped with a macrophage-enriched metabolic network module that is highly associated with metabolic traits in mice and humans. Integration of gene expression in the design and analysis of traditional F2 intercross studies allows high confidence prediction of causal genes and identification of involved pathways and networks.

The discovery of novel genes which contribute to complex human disorders remains a challenge for geneticists. The conventional methodology of determining whether a particular locus is involved in a given disease involves testing for inheritance of specific genomic regions in successive generations of affected individuals. This typically leads to multiple loci (known as quantitative trait loci, or QTLs), each of which contributes modestly to the overall phenotype. Each locus may contain hundreds of genes, making the elucidation of the underlying gene or genes labor intensive and time consuming. Additionally, differentiating genes that are causal for the disease from those that are reactive to the biological alterations resulting from the disease has been difficult.

The advent of microarray technology has enabled scientists to simultaneously examine alterations in the mRNA levels of thousands of transcripts in a sample. Since microarrays yield quantitative estimates of gene expression changes, the loci that control their expression can be mapped. These loci are known as expression QTLs (or eQTLs). eQTLs that map near the gene and are likely to regulate gene expression in *cis* are termed *cis*-eQTLs 1. Genes with *cis*-eQTLs that are coincident with a clinical disease-related trait QTL (or cQTL) have an increased likelihood of contributing causally to the particular disorder, especially if expression of the gene is correlated with the severity of the disease trait 2,3.

However, correlation between an expression trait and a clinical phenotype does not imply a causal/reactive relationship due to linked causal mutations, and particular alleles may also influence RNA levels and phenotypes independently, further confounding the analysis. There is unambiguous biological directionality in that DNA changes influence alterations in transcript abundances and clinical phenotypes, so the number of possible relationships among correlated traits can be greatly reduced. For example, among two traits which are correlated and controlled by a unique DNA locus, only three likely relationship models exist, namely causal, reactive, and independent 4,5. Therefore, after constructing a network, one can simultaneously integrate all possible DNA variants and their underlying changes in transcript levels, and each relationship can be supported as being causal, reactive, or independent in relation to a particular phenotype such as obesity. This is referred to as the likelihood-based causality model selection (LCMS) procedure 4.

Using our LCMS procedure, we have predicted ~100 causal genes for abdominal obesity using an F2 intercross between the C57BL/6J and DBA/2J strains of mice (the BXD cross) 4. In order to validate the predictive power of LCMS, we carried out phenotypic characterization of transgenic or knockout mouse models for nine of the top candidate genes, and report here that in total eight out of the nine genes under characterization were found to influence obesity-related traits. We analyzed liver gene expression signatures of the transgenic and knockout mouse models and demonstrated that all nine genes affect common pathways and subnetworks that relate to metabolic pathways, suggesting that obesity is driven by a gene network instead of a single gene.

RESULTS

Knockout (ko) or transgenic (tg) mouse models for the nine candidate causal genes were constructed or obtained from vendors or institutional investigators 6–8 (Supplementary Table 1). Except for *Gas7* as noted below, transgene expression patterns were similar to those of the endogenous genes 4,7 (Supplementary Fig. 1). We previously demonstrated a preliminary validation of *Zfp90*, *C3ar1*, *Tgfb2*, *Lactb*, and *Lpl* in modifying adiposity using ko or tg mice 4,9. Here we present additional validation evidence from these mouse models, and the obesity-related phenotypic characteristics, gene expression signatures, and pathway/network analyses of four additional candidate causal genes: *Gas7* (tg), *Gpx3* (tg), *Gyk* (ko), and *Me1* (ko).

In vivo characterization of mouse models

We previously showed that *Zfp90* tg mice had a significantly increased fat mass to lean mass ratio 4. Additional phenotyping indicated that *Zfp90* tg mice also had significantly increased body weight, total fat pad mass, adiposity, and retroperitoneal, mesenteric, and subcutaneous fat pad masses (Supplementary Table 2) compared to wt mice. Plasma lipid profiles showed non-significant trends for LDL and other parameters (Supplementary Table 3), but our ability to assess these was limited since *Zfp90* tg mice were unable to breed and only three tg founders were studied.

As described previously, male *C3ar1* ko and male *Tgfb2* heterozygous ko mice (*Tgfb2* homozygous ko animals are not viable) had reduced fat/lean ratio as compared to their wild-type (wt) littermates 4. Analysis of additional mice from both sexes confirmed our previous results (Supplementary Table 2; Figure 1a–1d). Interestingly, female *C3ar1* ko and female *Tgfb2* heterozygous mice demonstrated opposing trends (Figure 1b and 1d compared to 1a and 1c). Individual fat pad masses were not significantly different between *Tgfb2* heterozygotes and wt, although males and females retained the opposite trends (Supplementary Table 2). Similarly, male and female *C3ar1* ko mice retained the opposite trends in individual fat pad masses, (Supplementary Table 2), suggesting the existence of a sex-by-genotype interaction for *C3ar1* and *Tgfb2*. Female *Tgfb2* heterozygous ko also exhibited a significant increase in endpoint free fatty acids, but no significant alterations were seen in endpoint lipid levels for *C3ar1* ko (Supplementary Table 3).

Heterozygous *Lpl* ko mice were previously shown to have increased adiposity 9. Male but not female heterozygotes also had significantly increased fat pad weights compared to controls (Supplementary Table 2). Both male and female heterozygous mice had significantly increased endpoint triglycerides. Females also exhibited increased unesterified and total cholesterol levels (Supplementary Table 3).

The increased adiposity of *Lactb* tg mice noted previously 9 was confirmed in an additional *Lactb* tg line in female mice, but not in males (Supplementary Figs. 2a and 2b). Neither male nor female tg mice showed alterations in individual fat pad weights or lipid levels (Supplementary Tables 2 and 3).

Gas7 tg mice were constructed as described in the Methods. Expression of the human transgene was found in all tissues analyzed except for the spleen, while the endogenous mouse *Gas7* was only expressed significantly in the brain (Supplementary Fig. 1b). Male but not female *Gas7* tg mice showed significantly decreased fat/lean ratio as compared to wt littermates (Figure 1e and 1f), which were confirmed in an additional independent *Gas7* tg line (Supplementary Figs. 2c and 2d). Male *Gas7* tg also had significantly decreased body weight, gonadal fat pad weight, and total fat pad weight, and females had significantly decreased mesenteric fat pad weight (Supplementary Table 2). Moreover, male *Gas7* tg showed significantly reduced triglycerides, unesterified cholesterol, and glucose levels and significantly increased endpoint total cholesterol and HDL levels, while female tg mice showed only significantly reduced unesterified cholesterol levels (Supplementary Table 3).

For *Gas7* tg males, the difference in adiposity appeared at the first measurement time point at 11 weeks of age and the increase in the difference with age was not as obvious as in the other models. To assess a possible embryonic effect of the tg, we measured body weights of male *Gas7* mice at weaning (3 weeks) and found that these were not significantly different than littermate controls for both transgenic lines. The body weights at weaning from the *Gas7* transgenic line analyzed in Figure 1e are presented as Supplementary Figure 2e.

Gpx3 male but not female tg mice showed a significant decrease in fat/lean ratio growth compared to controls (Figures 1g and 1h). No differences in individual fat pad weights were seen (Supplementary Table 2). Female *Gpx3* tg mice showed increased endpoint total cholesterol and HDL levels as compared to female controls (Supplementary Table 3).

Gyk is located on the X chromosome, and knockout males die by 4 days of life, while homozygous females are likely embryonic lethal 8. Therefore, only *Gyk* female heterozygous mice were characterized. There were no significant alterations in fat/lean ratio as compared to their wt littermates (Supplementary Fig. 2f), nor in individual or total fat pad weights (Supplementary Table 2). They did, however, exhibit decreases in free fatty acids and glucose levels (Supplementary Table 3).

The *Me1* ko mice were characterized at a separate facility where the growth curves of body weight instead of adiposity were recorded while mice were on several different diets including a medium high fat diet (44.9% Kcal from fat) and a high sucrose diet (76.5% kcal from carbohydrate). Both male and female *Me1* ko mice on a medium high fat diet demonstrated decreased body weight (Figure 1i and 1j). A similar trend was also observed in male though not female ko mice on a high sucrose diet (Supplementary Figs. 2g and 2h). We measured adiposity by NMR at the beginning and the end of the diet period in the *Me1* mice fed on high fat and high sucrose diets and did not find significant differences in initial or endpoint adiposity in mice on either diet (Supplementary Figs. 2i and 2j), though there was a trend of reduced endpoint fat mass as determined by NMR in male ko mice on high sucrose diet (7.61 ± 0.31 in ko and 6.75 ± 0.26 in wt; $p=0.06$), consistent with the decreased body weight. *Me1* ko mice on high fat but not a high sucrose diet showed a significant difference in food intake (males only) relative to littermate controls (Supplemental Figure 3). Thus, food intake alone cannot account for the significantly decreased body weight in *Me1* ko mice.

Gene expression profiling of mouse models

In order to explore the mechanisms underlying the observed phenotypic changes, we profiled male liver tissue from each mouse model (with the exception of the *Gyk* female ko) to obtain gene expression signatures for each of the individual candidate causal genes. As shown in Table 1, at $p < 0.05$ (t test), we observed hundreds to thousands of genes whose expression levels were altered in the livers of each of the tg or ko mouse strains compared to their wt littermate mice. The false discovery rate (FDR) ranged from 1.9%–86% at this p-value cutoff. The high FDR values in most of the profiling experiments are likely due to the relatively small number of mice ($n=3$ to 9) involved. The signature genes can be found in Supplementary Tables 4–12. We reasoned that although the relatively high levels of FDR would influence the confidence in individual signature genes identified, they should have less impact on the pathway analyses discussed below.

Two complementary methods, including Fisher's exact test-based enrichment analysis of the Gene Ontology (GO) functional categories 10, Panther pathways 11, or Ingenuity canonical pathways (Ingenuity® Systems, www.ingenuity.com), and Gene Set Enrichment Analysis (GSEA) of curated functional gene sets from public databases based on weighted Kolmogorov–Smirnov-like statistics 12 were used to analyze the functional relevance of the liver gene expression profiles to the phenotypic traits.

The liver gene signatures from these mouse models are enriched for many overlapping metabolic pathways (Table 1; Supplementary Table 13). Multiple mouse models showed enrichments for pathways related to steroid, fatty acid, amino acid, and glutathione metabolism pathways, as well as purine metabolism, the pentose phosphate pathway, and IL-10 signaling. Previously, we reported 13 TCA cycle-centered metabolic pathways as being differentially affected in fat and lean mice from the F2 DBA/J and C57BL/6J intercross 13. As depicted in Figure 2, two or more over-represented metabolic pathways identified for each strain of mice in the current study overlapped with these previously described pathways. Therefore, the causal genes identified by the LCMS procedure and tested in this study likely affect adiposity by modifying similar obesity-related pathways. Moreover, as shown in Table 2, the signature genes from each of the mouse models overlap significantly with the signature genes identified from one or more of the other mouse models.

Overlap between liver signature genes from mouse models and a macrophage-enriched metabolic network (MEMN)

Based on the co-expression networks constructed from liver and adipose tissues collected from a mouse cross between C57BL/6J (B6) and C3H/HeJ on an apolipoprotein E null background (BXH/apoE), we previously identified a macrophage-enriched metabolic network (MEMN) that is highly associated with metabolic traits and appears to be of macrophage-derived origin 9. Five of the nine genes under validation, namely, *Zfp90*, *Lactb*, *Lpl*, *C3ar1*, and *Tgfbr2*, are within this MEMN subnetwork. In addition, the liver gene signatures derived from five out of the nine validation mouse models, *Zfp90* tg, *Gpx3* tg, *Lpl* ko, *C3ar1* ko, and *Tgfbr2* ko, significantly overlapped with MEMN genes (Table 2). Recall that these candidate genes were identified as causal genes for obesity from a C57BL/6 ×

DBA/2J cross. Now, in a different mouse cross setting (BxH/apoE), many of these candidate genes and their downstream genes are confirmed to be within or highly overlap with a coexpression subnetwork that is relevant to obesity, diabetes, and atherosclerosis traits. Furthermore, we recently uncovered a human MEMN which is associated with obesity-related traits and contains extensive overlap with the mouse MEMN we describe here 14. Therefore, it is not surprising to find *C3ARI*, *LACTB*, and *LPL* as well as the previously characterized *HSD11B1* 15 among the human MEMN network genes, further highlighting the shared networks between mice and humans and suggesting a common mechanism leading to the development of obesity.

Bayesian network analysis of liver signatures in mouse models

As the sample sizes for defining perturbation signatures were small, the quality of the signatures derived could be noisy, thus limiting our ability to see additional significant overlaps between these signatures. To overcome the problem of noisy signatures, we projected them onto our liver transcriptional Bayesian network and then compared the subnetworks around the signatures instead of signatures themselves 16.

We previously described a method to reconstruct probabilistic, causal Bayesian networks by integrating genetic and gene expression data 4,17. A liver transcriptional network was constructed based on three F2 intercross populations derived from the C57BL/6J, C3H/HeJ, and CAST/Ei strains described previously 18 (see Methods). For each perturbation signature, we extracted the largest connected subnetwork in the whole liver transcriptional network as described 19 (see Methods). All of these subnetworks overlapped significantly with one another and with the MEMN module (Table 3), again suggesting that all causal genes affect a common pathway.

A core subnetwork consisting of 637 genes (Figure 3; Supplementary Table 14) was identified to be common in at least five of the nine perturbation signature subnetworks, while there was no gene that was common in at least five of the nine perturbation signatures themselves. Genes in this core subnetwork were significantly enriched in many GO biological processes (Table 4), and also significantly overlapped with the coexpression module MEMN described above. The core subnetwork consisted of *Zfp90*, *Lactb*, and well known key regulators for fatty acid and lipid metabolism *Insig1* and *Insig2* 20,21, and many classic cholesterol and fatty acid metabolic genes such as *Hmgcs1*, *Sqle*, *Dhcr7*, and *Fasn*. This suggests that the possible mechanism of the nine candidate causal genes giving rise to the obesity phenotype is through perturbations of this core subnetwork.

Relationship to human genome-wide association findings

Recent genome-wide association studies (GWAS) have identified more than ten loci that are associated with obesity-related traits 22–28. These loci include the highly replicated ones such as *FTO* and *MC4R* as well as the less replicated ones such as *INSIG2*, *GNPDA2*, *TMEM18*, *NEGR1*, and *SH2B1*. Although none of the nine causal genes that we have tested are within the obesity GWAS findings, the obesity GWAS gene *INSIG2* is within the core subnetwork that we identified, and *LPL* has been associated with triglyceride and HDL traits in GWAS 29,30.

One way to link our study to human GWAS is to test whether the causal genes we identified in mouse are enriched for harboring variations in human populations that show evidence of association (for example, low p values that may not reach the stringent threshold required for genome-wide significance) with obesity traits. Utilizing data from the Broad Institute GWAS control population 31, we tested the 667 core subnetwork genes for enrichment of neighboring SNPs (termed *cis*-SNPs) with low body mass index (BMI) association p values (see Methods for details). We did not find any enrichment for *cis*-SNPs with low association p values to BMI among the genes composing the full core sub-network set. However, when the 102 mouse MEMN genes within the core sub-network were considered, we did observe a significant enrichment for SNPs with low association p values to BMI.

Specifically, 12.5% (267 out of 2132) of *cis*-SNPs selected from the 102 mouse MEMN genes in the core subnetwork reached a BMI association cutoff of $p < 0.1$, as compared to an average of 10.4% [95% confidence interval (CI): 7.9% to 12.6%] in the *cis*-SNPs derived from 10^5 sets of randomly selected 102 genes by permutation. The enrichment p value was 0.031, defined as the probability of obtaining the observed result or even more extreme under random sampling. In addition, the average log of association p values (logP) for the *cis*-SNPs of the 102 mouse MEMN core subnetwork genes is -1.11 , as compared to an average of -1.00 [95% CI: -1.10 to -0.91] in the 10^5 random sets ($p = 0.010$). The comparison of the percentage of low association p values and the average logP between the mouse MEMN core subnetwork genes and the 10^5 random gene sets are shown in Supplementary Figures 4a and 4b, respectively.

DISCUSSION

As summarized in Table 5, genetic perturbation of eight out of the nine (~90%) candidate genes which were predicted to be causal for obesity in mice using our LCMS procedure caused significant alterations in fat/muscle ratios as well as relevant changes in body weight, adiposity (total fat/body weight), individual fat pad masses or plasma lipids. Furthermore, we identified corresponding changes in the liver expression of genes involved in metabolic pathways previously identified to be differentially regulated between fat and lean mice 13. Therefore, a large majority of the candidate causal genes for abdominal obesity predicted by LCMS were validated at both a phenotypic and a gene expression level. The liver gene expression signature genes from the validation mouse models highly overlapped with one another and with the metabolic trait-associated MEMN module genes. All of these causal candidate genes were found to impact a common liver transcriptional subnetwork that is enriched for GO metabolic pathways and the MEMN module. These multiple lines of evidence suggest that the perturbation of these predicted causal genes influences obesity via a common functional mechanism.

Interestingly, sex specificity in phenotypic effect was common, and we observed opposing effects on abdominal obesity between the sexes in *C3ar1* ko and *Tgfbr2* heterozygous ko mice. Sex hormones can affect *Tgfbr2* expression 32, and C3a stimulates the release of ACTH, which is involved in the production of androgens 33. Down-regulation of both genes in the ko mouse models may alter the impact of sex hormones and lead to sex-specific phenotypes.

Among the newly validated genes, *Gas7* was originally identified as a gene that was expressed in serum-starved NIH3T3 cells, and its protein structure resembles Oct2 and synapsins, which are involved in neuronal development and neurotransmitter release, respectively 34,35. It is selectively expressed in mature cerebral cortical, hippocampal, and cerebellar neurons 35. Our studies now indicate relevance of this gene to fat metabolism and other previously unknown pathways such as the insulin signaling pathway.

Gpx3 is involved in cellular protection against oxidative damage through the reduction of peroxides 36. The cytosolic isoform of *Gpx3*, *Gpx1*, has been associated with obesity 37, and recently the dysregulation of *Gpx3* in the plasma and adipose of obese subjects has been implicated in the increase in inflammatory signals and oxidative stress and hence obesity-related metabolic disorders 38. Our study provides primary evidence that *Gpx3* is a causal gene for obesity and supports that *Gpx3* overexpression modifies insulin resistance 38. Although the magnitude of the effect of *Gpx3* tg on phenotype was relatively weak (Figure 1g and 1h), the liver gene expression signature from the *Gpx3* animals highly overlaps that of *Gas7*, *Lactb*, *Gyk*, and *Lpl* as well as significantly overlaps the MEMN genes (Table 2), suggesting that *Gpx3* is causally affecting abdominal obesity along with the other genes validated in our study. The weak phenotypic validation might be a result of low copy number of the transgene 7 and susceptibility to compensatory mechanisms in gene networks 9.

Me1 encodes a cytosolic NADP(+)-dependent enzyme involved in the regeneration of pyruvate from malate back to the mitochondria, forming a link between the glycolytic pathway and the citric acid cycle 39. By assisting with the release of acetyl-CoA and NADPH from the mitochondria into the cytosol, they are made available for *de novo* fatty acid biosynthesis and other metabolic processes. *Me1* is considered lipogenic and altered levels of *Me1* enzyme activity has been associated with obesity mouse and rat models 40,41. Recently, *Me1* was identified as a primary candidate gene underlying a porcine QTL associated with backfat thickness 42.

Gyk encodes an enzyme responsible for the metabolism of endogenous and dietary glycerolipids 8. Deficiency in *Gyk* has been linked to altered fat and lipid metabolism 8,43, and deficiency in *Aqp7* which elevates *Gyk* expression has been associated with obesity development 44. In this study, we did not validate this gene at the phenotypic level. However, pathway analysis of the liver *Gyk* signature indicated that 5 out of 13 of the metabolic pathways previously linked to fat content were affected, and the *Gyk* heterozygous ko liver signature highly overlapped with the signatures derived from mouse models of other validated genes including *Gas7*, *Gpx3*, *Lactb*, and *Me1*, supporting a causal role. The lack of phenotypic validation might be a result of insufficient perturbation represented by the heterozygous ko.

When directly comparing our causal genes with the findings from the recent human GWAS studies, we only found limited overlaps. We reason that the GWAS genes/loci represent the *cis* variations in human population that confer disease risk, whereas by requiring multiple overlapping loci between the expression and fat mass traits in our LCMS method we implicitly required the causal genes to be affected in *trans* by a given genetic locus and then

cause variations in the obesity traits. Thus, it is not surprising to observe a limited overlap between the GWAS genes and the mouse causal genes we have identified. The causal genes that are affected by DNA variation in *trans* may not have been identified in the GWAS because the signals were too subtle to detect with the current scale, yet are still of interest because they are supported as causal. The fact that we found a weak enrichment for SNPs with low association p values to BMI from the Broad Institute GWAS population 31 when the mouse MEMN genes within the core subnetwork were considered supports this hypothesis.

In summary, we have validated the majority of the top genes predicted to be causal for abdominal obesity through phenotypic characterization and gene expression profiling, thus supporting the LCMS as a powerful tool in predicting causal genes for diseases. Although the genes are seemingly disparate, each appears to affect metabolic pathways that are linked to the TCA cycle. Future directions include the application of these network approaches to additional relevant tissues such as adipose, with incorporation of potential cross-tissue interactions, as well as environmental variations. Also, the investigation of negative predictive value of the LCMS procedure would be of value, though this is a more complicated problem than it appears on the surface, given that a list of LCMS predicted causal genes from one tissue is by no means comprehensive as many tissues are involved in the regulation of body fat. Considering that a large number of genes influence body weight, focusing on pathways and networks rather than pinpointing individual genes may be more efficient in elucidating the pathogenesis of obesity and the development of novel treatments.

METHODS

Construction of mouse models

Details regarding the *Gas7* transgenic and *Me1* knockout mouse models can be found in the Supplementary Methods. The *Zfp90*, *Lactb*, and *Gpx3* transgenic and *Gyk* and *Lpl* knockout mouse models were constructed as described previously 4,6–8. *Tgfbr2* and *C3ar1* knockout mouse models were obtained from Deltagen as described previously 4.

Breeding and genotyping of mice

All mice except *Me1* mice were bred at UCLA and were fed a 4% chow diet (Harlan Teklad 7017; 4% fat, 0% cholesterol) *ad libitum* and maintained on a 12 hour light/dark cycle. Genomic DNA was isolated from ear and tail samples using a DNeasy kit (Qiagen, CA) and genotyped using PCR. All reactions were carried out using initial enzyme activation at 95°C for 5 minutes, followed by 35 cycles at 95°C for 30 seconds, 56°C for 30 seconds, and 72°C for 1 minute, and finished with an extension at 72°C for 7 minutes. A detailed method for breeding and genotyping *Me1*^{-/-} mice has been described by Qian et al 41.

Phenotypic characterization of the mouse models

Starting at 11 weeks of age, mice (except *Me1* ko) were fed a 6% chow diet (Harlan Teklad 7013; 6.25% fat, 0% cholesterol) for 12 weeks. Each mouse was monitored for body weight and evaluated by NMR (Brucker Minispec) for body weight composition including lean mass, fat mass and water content over the course of the diet every two weeks. At the end of

the 12-week diet, mice were sacrificed using CO₂ asphyxiation. Gonadal (fat surrounding the gonads), retroperitoneal (fat beneath the kidneys), mesenteric (fat attached to the intestines), and subcutaneous (fat below the surface of the skin on the thighs) fat pads were collected and weighed. Liver was collected for RNA profiling. All procedures were done in accordance with the current National Research Council Guide for the Care and Use of Laboratory Animals and were approved by the UCLA Animal Research Committee. Additional details regarding the characterization of the *Gpx3* and *Me1* mice can be found in the Supplementary Methods.

Analysis of phenotypic data

The Student's t-test was used to analyze the differences in the phenotypic traits between tg or ko animals and their wt littermate controls. The significance level was set to $p < 0.05$. Significance of the difference in the growth curves of fat/muscle ratio between tg/ko and wt controls for all mouse models except *Me1* ko mice and in growth curves of body weight for *Me1* mice was determined using an autoregressive method described previously to enhance the power of difference detection by leveraging multiple repeated measures over a number of time points for each animal 4.

RNA sample preparation and microarray processing

For the liver tissues from the *Zfp90* transgenics, the *Tgfbr2* heterozygous ko, the *C3ar1* ko, the *Lpl* heterozygous ko, the *Me1* ko, and each of their respective littermate control mice, RNA preparation and array hybridizations were performed at Rosetta Informatics. For *C3ar1*, *Tgfbr2* and *Zfp90* mouse strains, the custom ink-jet microarrays used in this study were manufactured by Agilent Technologies (Palo Alto, CA) and consisted of 23,574 non-control oligonucleotides extracted from mouse Unigene clusters and combined with RefSeq sequences and RIKEN full-length cDNA clones. For *Lpl* and *Me1* mouse strains, the Agilent array consisted of 39,556 non-control probes representing 37,687 genes. For *Gas7* tg, *Gpx3* tg, *Lactb* tg, and *Gyk* female heterozygous ko, microarray profiling of the liver tissues was performed using Illumina MouseRef-8 beadchips. Each beadchip contained 24,886 oligonucleotide probes (849 control and 24,837 non-control) designed based on the Mouse Exonic Evidence Based Oligonucleotide (MEEBO) set, the RIKEN FANTOM 2 database, and the NCBI Reference Sequence (RefSeq) database. Additional details can be found in the Supplementary Methods. All expression data have been submitted to GEO database under the Super-series accession number GSE12000.

Selection of active or expressed gene sets based on microarray profiling

As a limited number of mice ($n=3-9$) per mouse model was used for gene expression profiling and the statistical power was low considering the large number of multiple tests for tens of thousands genes, we restricted attention to the subsets of genes that are more biologically relevant. The Agilent arrays do not provide a measure of "presence" or "absence," and therefore, we selected set of "most transcriptionally active genes" for mouse models profiled with Agilent arrays using the program Resolver 3,45,46. The active genes were defined as those with significance level $p < 0.05$ (as determined by error model 3,45,46) in at least 10% animals for each strain of mice including both tg/ko and controls. These

active genes represent those whose expression levels vary across samples and, thus, are more biologically relevant. For mouse strains that were profiled with Illumina arrays, the program BeadArray was used to normalize the expression intensity values across as well as within arrays using the “average” algorithm embedded in the software. Genes with detection scores of >0.99 (corresponding to detection $p < 0.01$) in at least 10% animals for each strain of mice were selected as “expressed” genes. This active/expressed gene selection procedure significantly reduced the size of starting gene sets for subsequent analysis from ~24,000 to ~1000–9000 genes per strain of mice, thus helping to alleviate multiple testing concerns.

Selection of Signature gene sets based on microarray profiling

A Student’s t-test was used to identify genes with significant differences between tg or ko animals and the corresponding wt control mice. These genes were defined as “signature” genes, representing the perturbed gene expression signature as a result of single gene modification. The significance level was set to $p < 0.05$. The false discovery rate at this significance level was calculated using Q-value as reported 47. All statistical analyses were carried out in the R statistical environment.

Pathway analysis

Each signature gene set identified above was classified using Gene Ontology (GO) 10 and Panther pathway 11 database assignments. The Ingenuity Pathway Analysis software (Ingenuity® Systems, www.ingenuity.com) was used to analyze the enrichment of canonical pathways in the signature genes identified above, and we also analyzed the enrichment of ~470 functional gene sets curated from public databases using Gene Set Enrichment Analysis (GSEA) 12. Additional details can be found in the Supplementary Methods.

Mouse Crosses and Tissue Collection

Three mouse crosses constructed from C57BL/6J (B6), C3H/HeJ (C3H), and CAST/Ei (CAST), namely, B6 \times C3H wildtype (BXH/wt), B6 \times C3H on an ApoE null background (BXH/apoE), and B6 \times CAST (BXC) were described previously 18,48. Additional details can be found in the Supplementary Methods. All procedures of housing and treatment of animals were performed in accordance with IACUC regulations.

Identification of the macrophage-enriched metabolic network (MEMN)

The construction of the coexpression network using liver and adipose tissues from BXH cross and the identification of the MEMN has been described previously 9. Briefly, both genotype and gene expression data were utilized to construct co-expression networks that consisted of highly connected genes from each tissue and sex. An iterative search algorithm was then used to detect highly interconnected subnetworks. One particular subnetwork that was highly enriched for causal genes for all metabolic traits tested, highly conserved between tissues and sexes, and highly enriched for macrophage genes was referred as MEMN.

Construction of Bayesian network and subnetwork for candidate causal gene perturbation signatures

Liver expression data generated from the above three mouse cross populations BXH/wt, BXH/apoE, and BXC was integrated with the genotypic data also generated in the same populations to reconstruct the Bayesian networks as previously described 21–24. Additional details can be found in the Supplementary Methods.

The construction of a subnetwork for a set of signature genes in the network is as follows: given a set of genes, we identified all of the nearest neighbors of these genes in the network (i.e., we identified all nodes in the network that were either in the input set or directly connected to a node in the input set). This first step produced a set of node pairs connected by an edge. In the next step, direct connections/edges among all the nodes in these pairs were identified and added. The resulting largest connected subnetwork (i.e., all smaller subnetworks that disconnect from the largest connected subnetwork were removed) was used as the subnetwork to represent the input set of signature genes.

The core subnetwork was identified by searching for genes that were present in more than half, in this case, five of the subnetworks representing the nine liver signature gene sets of the mouse models. The enrichment of the core subnetwork for 2283 gene sets from Gene Ontology Biological Processes category was analyzed using Fisher's exact test. A statistical cutoff of $2.2e-5$ was applied to the nominal p values to reflect the Bonferroni correction of multiple testing.

Comparison of signature genes across mouse models, between signature genes and the Bayesian subnetwork, and between signature genes and MEMN genes

The significance of overlap between different gene sets was estimated using Fisher's exact test statistics under the null hypothesis that the frequency of the genes in one signature set is the same between a reference set of 18,739 genes with Entrez Gene ID and the comparison gene set. The background for overlapping between gene sets and subnetworks in liver transcriptional network is 14,882 genes included in the whole network.

Enrichment of core subnetwork for genes/loci with low association p values to obesity traits in GWAS

The raw association p values between SNPs and body mass index (BMI) from the GWAS conducted by the BROAD Institute 31 were downloaded from the official website (<http://www.broad.mit.edu/diabetes/>). For each gene in the 667 core subnetwork and the 102 mouse MEMN module 9 genes within the core subnetwork, SNPs within 100kb distance (50kb upstream and 50kb downstream), termed *cis*-SNPs, were selected from dbSNP database and their association p values to BMI in the control population of the BROAD GWAS study were extracted. We compared each set of *cis*-SNPs of interest with 10^5 sets of *cis*-SNPs from randomly selected gene sets with matched size on the human 44k array, such that the number of *cis*-SNPs from the random gene sets roughly matched that of the gene sets of our interest. Two different tests were used to estimate the significance of enrichment for low association p values as detailed in Supplementary Methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank Dr. Richard Davis, Pingzi Wen, Melenie Rosales, Xiaohui Wu, Kathleen Ranola, and Xiaoyu Xia for helping with the tissue collection. We would also like to thank Leslie Ingram-Drake and Sharda Charugundla for technical support and Drs. Oleg Mirochnitchenko and Ira Goldberg for providing mouse models. The study was funded by NIH grants DK072206, HL28481, and HL30568.

References

1. Doss S, Schadt EE, Drake TA, Lusis AJ. Cis-acting expression quantitative trait loci in mice. *Genome Res.* 2005; 15:681–691. [PubMed: 15837804]
2. Monks SA, et al. Genetic inheritance of gene expression in human cell lines. *Am J Hum Genet.* 2004; 75:1094–1105. [PubMed: 15514893]
3. Schadt EE, et al. Genetics of gene expression surveyed in maize, mouse and man. *Nature.* 2003; 422:297–302. [PubMed: 12646919]
4. Schadt EE, et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet.* 2005; 37:710–717. [PubMed: 15965475]
5. Sieberts SK, Schadt EE. Moving toward a system genetics view of disease. *Mamm Genome.* 2007; 18:389–401. [PubMed: 17653589]
6. Merkel M, et al. Inactive lipoprotein lipase (LPL) alone increases selective cholesterol ester uptake in vivo, whereas in the presence of active LPL it also increases triglyceride hydrolysis and whole particle lipoprotein uptake. *J Biol Chem.* 2002; 277:7405–7411. [PubMed: 11751882]
7. Mirochnitchenko O, Palnitkar U, Philbert M, Inouye M. Thermosensitive phenotype of transgenic mice overproducing human glutathione peroxidases. *Proc Natl Acad Sci U S A.* 1995; 92:8120–8124. [PubMed: 7667255]
8. Huq AH, Lovell RS, Ou CN, Beaudet AL, Craigen WJ. X-linked glycerol kinase deficiency in the mouse leads to growth retardation, altered fat metabolism, autonomous glucocorticoid secretion and neonatal death. *Hum Mol Genet.* 1997; 6:1803–1809. [PubMed: 9302256]
9. Chen Y, et al. Variations in DNA elucidate molecular networks that cause disease. *Nature.* 2008; 452:429–435. [PubMed: 18344982]
10. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000; 25:25–29. [PubMed: 10802651]
11. Mi H, et al. The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* 2005; 33:D284–D288. [PubMed: 15608197]
12. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005; 102:15545–15550. [PubMed: 16199517]
13. Ghazalpour A, et al. Genomic analysis of metabolic pathway gene expression in mice. *Genome Biol.* 2005; 6:R59. [PubMed: 15998448]
14. Emilsson V, et al. Genetics of gene expression and its effect on disease. *Nature.* 2008; 452:423–428. [PubMed: 18344981]
15. Masuzaki H, et al. A transgenic model of visceral obesity and the metabolic syndrome. *Science.* 2001; 294:2166–2170. [PubMed: 11739957]
16. Liu M, et al. Network-based analysis of affected biological processes in type 2 diabetes models. *PLoS Genet.* 2007; 3:e96. [PubMed: 17571924]
17. Zhu J, et al. Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. *PLoS Comput Biol.* 2007; 3:e69. [PubMed: 17432931]
18. Schadt EE, et al. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* 2008; 6:e107. [PubMed: 18462017]

19. Zhu J, et al. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet.* 2008
20. Yabe D, Brown MS, Goldstein JL. Insig-2, a second endoplasmic reticulum protein that binds SCAP and blocks export of sterol regulatory element-binding proteins. *Proc Natl Acad Sci U S A.* 2002; 99:12753–12758. [PubMed: 12242332]
21. Janowski BA. The hypocholesterolemic agent LY295427 up-regulates INSIG-1, identifying the INSIG-1 protein as a mediator of cholesterol homeostasis through SREBP. *Proc Natl Acad Sci U S A.* 2002; 99:12675–12680. [PubMed: 12242342]
22. Chambers JC, et al. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat Genet.* 2008; 40:716–718. [PubMed: 18454146]
23. 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]
24. Loos RJ, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet.* 2008; 40:768–775. [PubMed: 18454148]
25. Mutch DM, Clement K. Unraveling the genetics of human obesity. *PLoS Genet.* 2006; 2:e188. [PubMed: 17196040]
26. Thorleifsson G, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet.* 2008 **advanced online publication.**
27. Willer CJ, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet.* 2008 **advanced online publication.**
28. Herbert A, et al. A common genetic variant is associated with adult and childhood obesity. *Science.* 2006; 312:279–283. [PubMed: 16614226]
29. Wallace C, et al. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *Am J Hum Genet.* 2008; 82:139–149. [PubMed: 18179892]
30. Willer CJ, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet.* 2008; 40:161–169. [PubMed: 18193043]
31. Saxena R, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science.* 2007; 316:1331–1336. [PubMed: 17463246]
32. Dhanasekaran SM, et al. Molecular profiling of human prostate tissues: insights into gene expression patterns of prostate development during puberty. *Faseb J.* 2005; 19:243–245. [PubMed: 15548588]
33. Francis K, et al. Complement C3a receptors in the pituitary gland: a novel pathway by which an innate immune molecule releases hormones involved in the control of inflammation. *Faseb J.* 2003; 17:2266–2268. [PubMed: 14563692]
34. Ju YT, et al. gas7: A gene expressed preferentially in growth-arrested fibroblasts and terminally differentiated Purkinje neurons affects neurite formation. *Proc Natl Acad Sci U S A.* 1998; 95:11423–11428. [PubMed: 9736752]
35. Moorthy PP, Kumar AA, Devaraj H. Expression of the Gas7 gene and Oct4 in embryonic stem cells of mice. *Stem Cells Dev.* 2005; 14:664–670. [PubMed: 16433621]
36. Ishibashi N, Mirochnitchenko O. Chemokine expression in transgenic mice overproducing human glutathione peroxidases. *Methods Enzymol.* 2002; 353:460–476. [PubMed: 12078519]
37. McClung JP, et al. Development of insulin resistance and obesity in mice overexpressing cellular glutathione peroxidase. *Proc Natl Acad Sci U S A.* 2004; 101:8852–8857. [PubMed: 15184668]
38. Lee YS, et al. Dysregulation of adipose GPx3 in obesity contributes to local and systemic oxidative stress. *Mol Endocrinol.* 2008
39. MacDonald MJ. Feasibility of a mitochondrial pyruvate malate shuttle in pancreatic islets. Further implication of cytosolic NADPH in insulin secretion. *J Biol Chem.* 1995; 270:20051–20058. [PubMed: 7650022]
40. van Schothorst EM, et al. Adipose gene expression response of lean and obese mice to short-term dietary restriction. *Obesity (Silver Spring).* 2006; 14:974–979. [PubMed: 16861601]
41. Qian S, et al. Deficiency in Cytosolic Malic Enzyme Does Not Increase Acetaminophen-Induced Hepato-Toxicity. *Basic Clin Pharmacol Toxicol.* 2008

42. Vidal O, et al. Malic enzyme 1 genotype is associated with backfat thickness and meat quality traits in pigs. *Anim Genet.* 2006; 37:28–32. [PubMed: 16441292]
43. Rahib L, MacLennan NK, Horvath S, Liao JC, Dipple KM. Glycerol kinase deficiency alters expression of genes involved in lipid metabolism, carbohydrate metabolism, and insulin signaling. *Eur J Hum Genet.* 2007; 15:646–657. [PubMed: 17406644]
44. Hibuse T, et al. Aquaporin 7 deficiency is associated with development of obesity through activation of adipose glycerol kinase. *Proc Natl Acad Sci U S A.* 2005; 102:10993–10998. [PubMed: 16009937]
45. Weng L, et al. Rosetta error model for gene expression analysis. *Bioinformatics.* 2006; 22:1111–1121. [PubMed: 16522673]
46. He YD, et al. Microarray standard data set and figures of merit for comparing data processing methods and experiment designs. *Bioinformatics.* 2003; 19:956–965. [PubMed: 12761058]
47. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A.* 2003; 100:9440–9445. [PubMed: 12883005]
48. Yang X, et al. Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* 2006; 16:995–1004. [PubMed: 16825664]

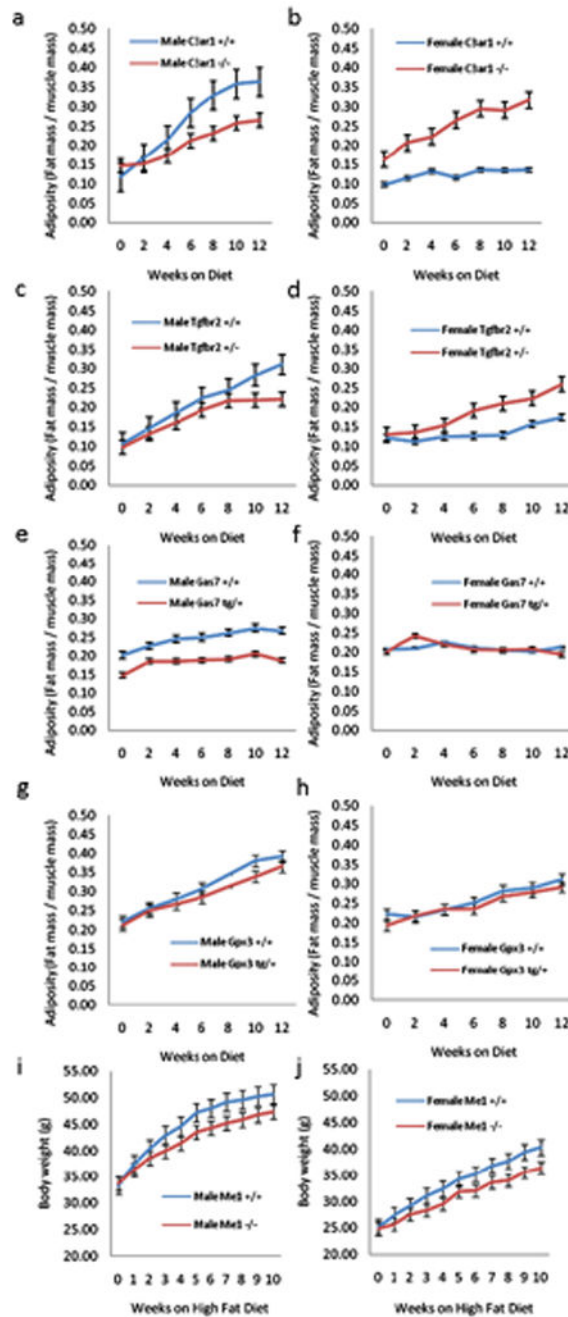


Figure 1. Adiposity (fat/muscle ratio) or body weight growth curves in the mouse models. The growth curves of males and females for each model are derived from the biweekly measurement of fat/muscle ratio every two weeks over the course of 14 weeks on a 6% fat diet. For *Me1* ko, the growth curves are from weekly measurement of body weight over the course of 10 weeks on a high fat diet. The p values are derived from the autoregressive model, which indicate the differences between the growth curves of tg/ko and wt, and are $< 10^{-10}$ for

panels a, b, c, i, and j; $< 10^{-5}$ for panels d and e; $< 10^{-2}$ for panel g, and > 0.05 for panels f and h.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

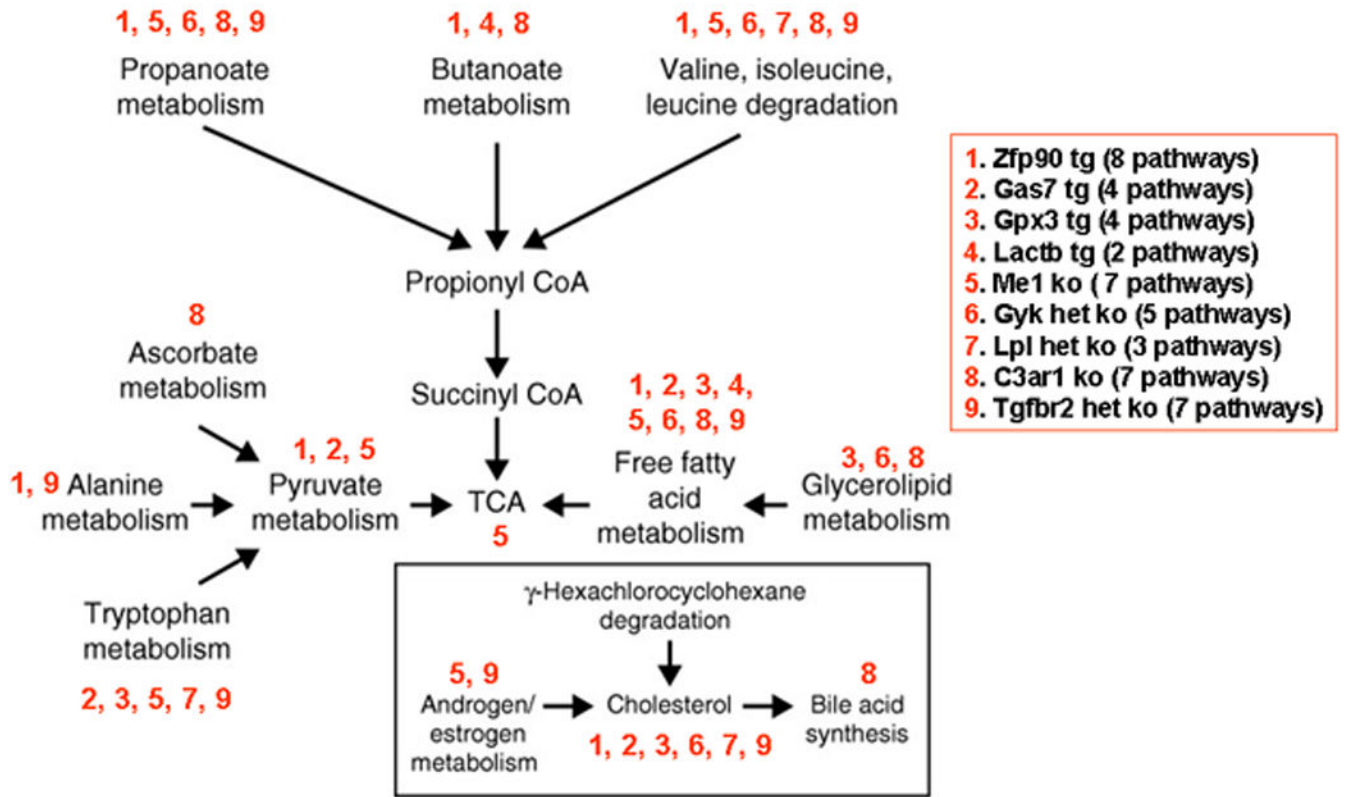


Figure 2. Disruption of metabolic pathways involved in fat pad mass trait in mouse models of the candidate genes. The nine mouse models are labeled as 1 to 9 in red. Each of the metabolic pathways previously identified to be different between fat and lean mouse is marked with the identifiers of the mouse models whose liver gene expression signatures are enriched for the specific pathway. The number of pathways that are over-represented in the liver gene expression signature of each mouse model is listed in parenthesis following the name of each mouse model.

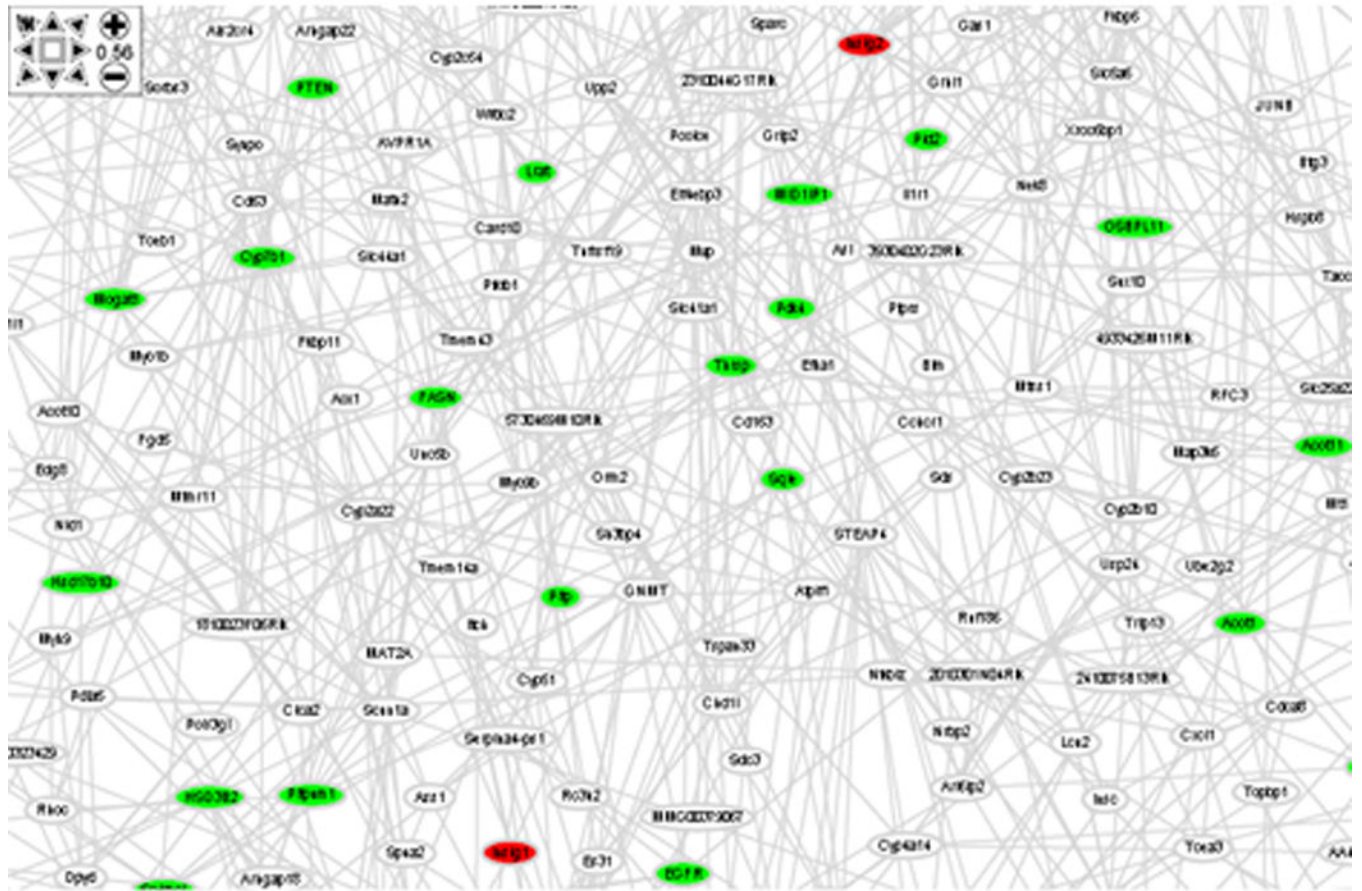


Figure 3.

A portion of the core subnetwork, derived from the liver transcriptional subnetworks representative of gene expression signatures of the mouse models of the candidate genes. The liver transcriptional network is the union of Bayesian networks constructed from three crosses derived from B6, C3H, and CAST. This core subnetwork consists of key regulators for fatty acid and lipid metabolism, including *Insig1* and *Insig2* (in red), and is enriched for genes involved in related GO biological processes. A scalable image of the full subnetwork is included as Supplementary Figure 5.

Table 1

Overlapping pathways in the liver gene expression signatures from the mouse models.

Gene	Signature size (p<0.05)	FDR at p<0.05	GO/Panther enrichment	Ingenuity pathway enrichment	GSEA
<i>Zfp90</i>	172	34.2%	Monocarboxylic acid,fatty acid, vitamin metab. Dehydrogenase Polyunsaturated fatty acid biosyn. Coenzyme/prosthetic group metab.	Propanoate, pyruvate, fatty acid metab. Lysine degradation Valine, leucine and isoleucine degrade. Fatty acid elongation in mitochondria	Up in tg: Biosynthesis of steroids, fatty acid Oxidative phosphorylation Down in tg: AKT signaling
<i>Gas7</i>	600	46.6%	Double-stranded DNA binding Proteasome complex Nucleosome Apoptosis	Acute phase response Purine, tryptophan, pyruvate metab. Pentose phosphate pathway Fatty acid elongation in mitochondria Insulin receptor signaling LXR/RXR FXR/RXR activation Hepatic cholestasis IL-10 signaling	Up in tg: Ribosome Complement and coagulation cascades Fatty acid metabolism Complement pathway
<i>Gpx3</i>	595	46.5%	Organelle/ribosome Mitochondrial ribosome Mitochondrial lumen Mitochondrial matrix Structural constituent of ribosome Steroid metabolism	Lysine biosynthesis Tryptophan, purine, arachidonic acid, fatty acid, linoleic acid metab. Pentose phosphate pathway FXR/RXR activation Acute phase response Metabolism of xenobiotics by CYP450	Down in tg: Insulin signaling CCR3 signaling in eosinophils
<i>Lactb</i>	207	85.7%		Glutathione metabolism Mitochondrial dysfunction	Up in tg: Glutathione metabolism Down in tg: Complement and coagulation cascades Fatty acid, butanoate metabolism
<i>Me1</i>	2904	1.86%	Protein biosynthesis Steroid, cofactor, acetyl CoA, coenzyme, glucose metab. TCA cycle Structural constituent of ribosome Electron transport Drug metabolic process		Up in ko: Androgen and estrogen, glycerine, serine and threonine, arachidonic acid metab. Down in ko: Valine, leucine, and isoleucine degrad. Propanoate, pyruvate, fatty acid, tryptophan starch and sucrose metab. Apoptosis Oxidative phosphorylation Citrate cycle, TCA cycle Cell communication
<i>Gyk</i>	957	33.5%	Structural constituent of ribosome Mitochondrial lumen, matrix Electron transport Reductase Small GTPase, G-protein Acetyltransferase	Mitochondrial dysfunction Valine, leucine and isoleucine degrad. Fatty acid, purine propanoate metab. Lysine biosynthesis, degradation Pentose phosphate pathway FXR/RXR activation	Up in ko: CCR3 signaling in eosinophils Insulin signaling Starch and sucrose metab Down in ko: Cell communication
<i>Lpl</i>	223	54.1%		LXR/RXR FXR/RXR activation IL-10 signaling	Up in ko: Valine, leucine and isoleucine degrad.

Gene	Signature size (p<0.05)	FDR at p<0.05	GO/Panther enrichment	Ingenuity pathway enrichment	GSEA
<i>C3ar1</i>	131	57.3%		Phenylalanine metabolism Hepatic cholestasis IL-10 signaling Lysine degradation	Up in ko: Ribosome Complement pathway Down in ko: Valine, leucine and isoleucine degrad. Butanoate, propanoate metabolism Fatty acid biosynthesis
<i>Tgfbir2</i>	125	28.4%	Steroid, cholesterol metabolism Oxygenase oxidoreductase activity monooxygenase activity immunoglobulin receptor activity Androgen and estrogen metabolism	Arachidonic acid, fatty acid metab. Metabolism of xenobiotics by CYP450 Tryptophan, phenylalanin metabolism Pentose phosphate pathway Linoleic acid metabolism	Up in ko: Biosynthesis of steroids Down in ko: Valine, leucine and isoleucine degrad. Propanoate, glutathione, fatty acid met. Oxidative phosphorylation

tg: transgenic; ko: knockout; het: heterozygote.

Overlap among liver signature gene sets derived from the mouse models of the candidate causal genes as well as between the signature genes and the previously identified macrophage-enriched metabolic network (MEMN).

Table 2

Signature Set	<i>Zfp90</i>	<i>Gas7</i>	<i>Gpx3</i>	<i>Lactb</i>	<i>Me1</i>	<i>Gyk</i>	<i>Lpl</i>	<i>C3arl</i>	<i>Tgfbir2</i>	MEMN module
<i>Zfp90</i>	0	1.14E-03	n.s	8.66E-03	1.83E-03	n.s	n.s	n.s	1.68E-02	1.45E-02
<i>Gas7</i>	1.14E-03	0	1.38E-17	9.55E-06	n.s	4.35E-12	4.29E-02	1.78E-03	n.s	n.s
<i>Gpx3</i>	n.s	1.38E-17	0	2.72E-06	n.s	1.12E-18	7.34E-04	n.s	n.s	1.86E-03
<i>Lactb</i>	8.66E-03	9.55E-06	2.72E-06	0	n.s	2.86E-07	4.61E-02	n.s	n.s	n.s
<i>Me1</i>	1.83E-03	n.s	n.s	n.s	0	n.s	5.87E-03	2.06E-02	n.s	n.s
<i>Gyk</i>	n.s	4.35E-12	1.12E-18	2.86E-07	n.s	0	n.s	n.s	n.s	n.s
<i>Lpl</i>	n.s	4.29E-02	7.34E-04	4.61E-02	5.87E-03	n.s	0	n.s	n.s	7.15E-04
<i>C3arl</i>	n.s	1.78E-03	n.s	n.s	2.06E-02	n.s	n.s	0	n.s	1.89E-02
<i>Tgfbir2</i>	1.68E-02	n.s	n.s	n.s	n.s	n.s	n.s	n.s	0	1.37E-15

Uncorrected p values < 0.05 from Fisher's exact test are listed and the ones that pass Bonferroni-corrected p < 0.05 are highlighted in red. For the overlap among the 9 liver signature gene sets, the Bonferroni corrected p value cutoff is 0.05/45 = 1.11e-3. For the overlap between the signature gene sets and MEMN module, the Bonferroni corrected p value cutoff is 0.05/9 = 5.56e-3.

Table 3

Overlap among the liver transcriptional subnetworks representing the liver signatures of the mouse models.

Subnetwork	<i>Zfp90</i>	<i>Gas7</i>	<i>Gpx3</i>	<i>Lactb</i>	<i>Me1</i>	<i>Gyk</i>	<i>Lpl</i>	<i>C3ar1</i>	<i>Tgfbp2</i>	MEMN
<i>Zfp90</i>	0	2.99E-81	2.93E-65	1.57E-24	4.36E-153	1.40E-66	9.35E-58	2.91E-40	1.55E-61	2.07E-28
<i>Gas7</i>	2.99E-81	0	2.55E-223	5.78E-67	1.88E-299	5.99E-185	7.97E-69	2.48E-93	1.39E-68	1.32E-17
<i>Gpx3</i>	2.93E-65	2.55E-223	0	5.11E-58	1.97E-249	9.85E-169	1.95E-72	2.33E-80	5.60E-60	3.36E-23
<i>Lactb</i>	1.57E-24	5.78E-67	5.11E-58	0	3.77E-82	1.88E-59	1.68E-25	3.47E-16	4.17E-18	5.51E-07
<i>Me1</i>	4.36E-153	1.88E-299	1.97E-249	3.77E-82	0	6.94E-240	1.82E-130	6.97E-138	6.72E-146	1.20E-44
<i>Gyk</i>	1.40E-66	5.99E-185	9.85E-169	1.88E-59	6.94E-240	0	1.21E-49	2.88E-54	4.44E-58	1.56E-18
<i>Lpl</i>	9.35E-58	7.97E-69	1.95E-72	1.68E-25	1.82E-130	1.21E-49	0	4.90E-37	1.28E-42	3.65E-22
<i>C3ar1</i>	2.91E-40	2.48E-93	2.33E-80	3.47E-16	6.97E-138	2.88E-54	4.90E-37	0	1.40E-39	1.74E-12
<i>Tgfbp2</i>	1.55E-61	1.39E-68	5.60E-60	4.17E-18	6.72E-146	4.44E-58	1.28E-42	1.40E-39	0	1.57E-99

All p values shown are derived from Fisher's exact test and pass Bonferroni corrected p<0.05.

GO Biological Process Categories enriched in the core subnetwork depicted in Figure 3.

Table 4

GO Biological Process	p_value	Overlap	GO Set Size	Core subnetwork size	background
lipid metabolic process	4.37E-10	71	762	637	14882
response to external stimulus	5.16E-10	89	1062	637	14882
cellular lipid metabolic process	4.44E-09	61	645	637	14882
alcohol metabolic process	1.09E-08	38	319	637	14882
steroid metabolic process	2.18E-08	29	209	637	14882
response to wounding	5.77E-08	65	757	637	14882
organic acid metabolic process	2.55E-07	47	495	637	14882
carboxylic acid metabolic process	4.75E-07	46	490	637	14882
steroid biosynthetic process	6.01E-07	18	107	637	14882
cholesterol biosynthetic process	7.92E-07	11	41	637	14882
sterol biosynthetic process	2.18E-06	11	45	637	14882
fat cell differentiation	3.02E-06	13	65	637	14882
sterol metabolic process	6.47E-06	16	102	637	14882
coenzyme metabolic process	1.52E-05	17	121	637	14882
cofactor metabolic process	2.01E-05	19	149	637	14882

All p values are derived from Fisher's exact test and pass Bonferroni corrected $p < 0.05$. The background for overlapping between gene sets and subnetworks in liver transcriptional network is 14,882 genes included in the whole network.

Table 5

Significant phenotypic traits observed in the mouse models.

Gene symbol	Mouse model	Sex	Fat-related traits	Lipid traits and glucose
<i>Zfp90</i>	tg	M/F	Increased fat/muscle growth, body weight, total fat pad mass, total fat/body weight, retroperitoneal fat pad, mesenteric fat pad, subcutaneous fat pad	
<i>Gas7</i>	tg	M	Decreased fat/muscle growth, body weight, gonadal fat pad, total fat pad mass	Decreased total cholesterol, HDL, unesterified cholesterol, triglyceride, and glucose
<i>Gas7</i>	tg	F	Decreased mesenteric fat pad mass	Decreased unesterified cholesterol
<i>Gpx3</i>	tg	M	Decreased fat/muscle growth	
<i>Gpx3</i>	tg	F		Decreased total cholesterol, HDL
<i>Lactb</i>	tg	M	Increased fat/muscle growth	
<i>Lactb</i>	tg	F	Increased fat/muscle growth	
<i>Me1</i>	ko	M	Decreased body weight	
<i>Me1</i>	ko	F	Decreased body weight	
<i>Gyk</i>	ko	F		Increased free fatty acids, decreased glucose
<i>Lpl</i>	ko (het)	M	Increased fat/muscle growth, total fat pad mass, total fat/body weight, mesenteric fat pad, subcutaneous fat pad	Increased triglycerides
<i>Lpl</i>	ko (het)	F	Increased fat/muscle growth	Increased triglycerides, decreased total cholesterol and unesterified cholesterol
<i>C3ar1</i>	ko	M	Decreased fat/muscle growth	
<i>C3ar1</i>	ko	F	Increased fat/muscle growth, gonadal fat pad, subcutaneous fat pad	
<i>Tgfb2</i>	ko (het)	M	Decreased fat/muscle growth	
<i>Tgfb2</i>	ko (het)	F	Increased fat/muscle growth	Increased free fatty acids

tg: transgenic; ko: knockout; het: heterozygote.