

# Variations in DNA elucidate molecular networks that cause disease

Yanqing Chen<sup>1\*</sup>, Jun Zhu<sup>1\*</sup>, Pek Yee Lum<sup>1</sup>, Xia Yang<sup>1</sup>, Shirly Pinto<sup>2</sup>, Douglas J. MacNeil<sup>2</sup>, Chunsheng Zhang<sup>1</sup>, John Lamb<sup>1</sup>, Stephen Edwards<sup>1</sup>, Solveig K. Sieberts<sup>1</sup>, Amy Leonardson<sup>1</sup>, Lawrence W. Castellini<sup>3</sup>, Susanna Wang<sup>3</sup>, Marie-France Champy<sup>6</sup>, Bin Zhang<sup>1</sup>, Valur Emilsson<sup>1</sup>, Sudheer Doss<sup>3</sup>, Anatole Ghazalpour<sup>3</sup>, Steve Horvath<sup>4</sup>, Thomas A. Drake<sup>5</sup>, Aldons J. Lusis<sup>3,4</sup> & Eric E. Schadt<sup>1</sup>

**Identifying variations in DNA that increase susceptibility to disease is one of the primary aims of genetic studies using a forward genetics approach. However, identification of disease-susceptibility genes by means of such studies provides limited functional information on how genes lead to disease. In fact, in most cases there is an absence of functional information altogether, preventing a definitive identification of the susceptibility gene or genes. Here we develop an alternative to the classic forward genetics approach for dissecting complex disease traits where, instead of identifying susceptibility genes directly affected by variations in DNA, we identify gene networks that are perturbed by susceptibility loci and that in turn lead to disease. Application of this method to liver and adipose gene expression data generated from a segregating mouse population results in the identification of a macrophage-enriched network supported as having a causal relationship with disease traits associated with metabolic syndrome. Three genes in this network, lipoprotein lipase (*Lpl*), lactamase  $\beta$  (*Lactb*) and protein phosphatase 1-like (*Ppm1l*), are validated as previously unknown obesity genes, strengthening the association between this network and metabolic disease traits. Our analysis provides direct experimental support that complex traits such as obesity are emergent properties of molecular networks that are modulated by complex genetic loci and environmental factors.**

A challenge in the post-genome era is deciphering the biological function of individual genes and gene networks that drive disease. Given the availability of low-cost, high-throughput technologies for genotyping hundreds of thousands of DNA markers, successes are being realized in identifying associations between DNA variants and diseases such as age-related macular degeneration<sup>1–3</sup>, diabetes<sup>4</sup> and obesity<sup>5</sup>. Although these and coming discoveries from a slew of genome-wide association studies currently under way provide a peek into pathways that underlie disease, they are usually devoid of context, so elucidating the functional role of such genes in disease can linger for years, as has been the case for ApoE, an Alzheimer's-susceptibility gene identified 15 years ago<sup>6</sup>. Even when an association to disease has been localized to a given region representing a single gene, in the absence of experimental support the gene cannot be definitively claimed to be the susceptibility gene. This problem is exacerbated in experimental crosses derived from inbred mouse strains, for which in addition to the problem of inferring the function of positionally cloned genes from the genetic data alone, the extent of linkage disequilibrium operating in such populations makes positional cloning a difficult and time-consuming process.

An alternative to the forward genetics approach is the construction of molecular networks that define the molecular states of a system that underlie disease, where such networks are constructed from molecular phenotype data scored in populations that manifest disease. The information that defines how variations in DNA lead to variations in complex traits flows through molecular networks. Characterizing molecular networks that underlie complex traits such

as disease can provide a more comprehensive view, which in turn can lead to the direct identification of genes underlying disease processes and the functional roles of these genes with respect to disease. Recent studies characterizing gene networks have demonstrated how genetic loci associated with expression traits can be combined with clinical data to infer causal associations between expression and disease traits<sup>7–12</sup>. By leveraging DNA variations as a systematic source of perturbations on molecular networks and clinical traits, biological processes can be studied at the systems level, in addition to studying gene function at the level of individual pathways<sup>13,14</sup>.

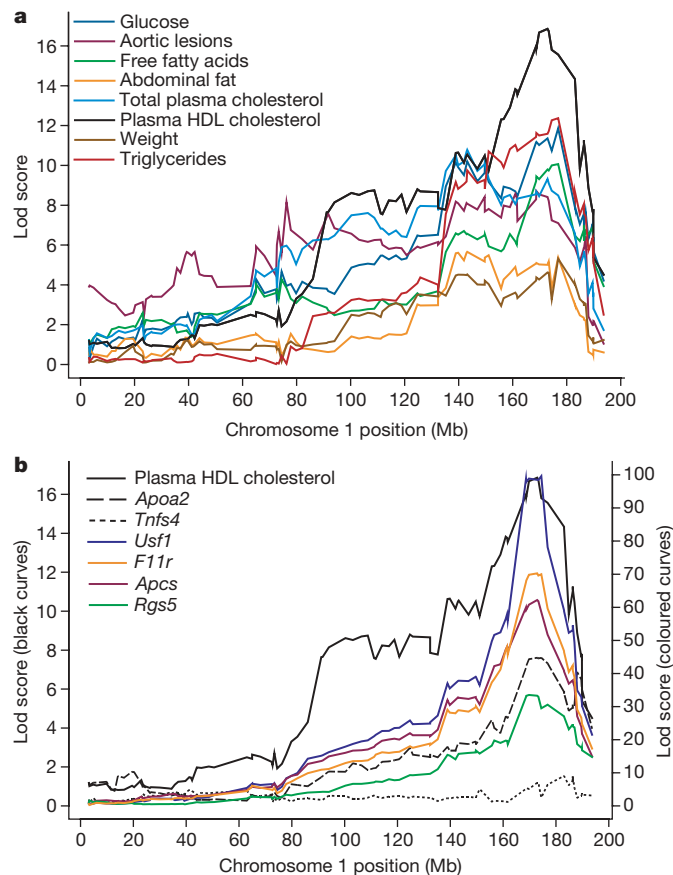
Here we report the development of an approach to uncover the components of co-expression networks that respond to variations in DNA associated with obesity-, diabetes- and atherosclerosis-related traits. In contrast to a forward genetics approach, we leverage quantitative trait loci (QTL) associated with disease to identify components of the co-expression network that are perturbed by the QTL and that in turn cause variations in disease traits. After constructing co-expression networks from liver and adipose tissues collected from a segregating mouse population, we identify sub-networks that are significantly associated with a complex of linked genetic loci related to obesity-, diabetes- and atherosclerosis-associated traits. A macrophage-enriched metabolic sub-network was found to be significantly enriched for expression traits supported as having a causal relationship with these metabolic traits. The connection to obesity and other metabolic syndrome traits is confirmed by validating three genes in this sub-network, *Lpl*, *Lactb* and *Ppm1l*, as previously unknown obesity genes.

<sup>1</sup>Rosetta Inpharmatics, LLC, Merck & Co., Inc., 401 Terry Avenue North, Seattle, Washington 98109, USA. <sup>2</sup>Department of Metabolic Disorders, Merck & Co., Inc., 126 East Lincoln Avenue, Rahway, New Jersey 07065, USA. <sup>3</sup>Department of Microbiology, Molecular Genetics, and Immunology, <sup>4</sup>Department of Human Genetics, and <sup>5</sup>Department of Pathology and Laboratory Medicine, UCLA, 650 Young Drive South, Los Angeles, California 90095, USA. <sup>6</sup>Institut de Genetique et de Biologie Moleculaire et Cellulaire, CNRS/INSERM/ULP, 67404 Illkirch, France.

\*These authors contributed equally to this work.

### A complex linkage to metabolic traits

A number of QTL mapping studies in experimental mouse cross populations have identified the distal half of chromosome 1 as a major contributor to metabolic traits such as weight, fat mass, and plasma glucose and cholesterol levels<sup>15–18</sup>. Much effort has been expended to map the quantitative trait genes (QTGs) underlying this locus, and these efforts have met with some success. For example, apolipoprotein A-II (*Apoa2*) and tumour necrosis factor superfamily, member 4 (*Tnfsf4*) have been mapped as QTGs for the cholesterol, fat mass, weight, insulin and atherosclerosis QTL mapped to the distal half of chromosome 1 (refs 19–23). However, it remains to be shown whether other genes in this chromosome 1 region contribute to these linkages beyond *Apoa2* and *Tnfsf4*. Furthermore, how the chromosome 1 QTL affect molecular networks in different tissues that in turn lead to pleiotropic effects on metabolic traits has not been characterized. An alternative to mapping QTGs for QTL is to incorporate molecular network data into these analyses to identify those network components that are perturbed by the QTL and that in turn lead to variations in disease traits. After characterizing the complexity of the chromosome 1 genomic region associated with metabolic traits, we implement a procedure to identify components of molecular networks that respond to genetic perturbations and in turn induce changes in metabolic traits. This procedure includes reconstructing co-expression networks and identifying highly interconnected functional sub-networks constituting these networks supported as having a causal relationship with disease traits.



**Figure 1 | The distal half of chromosome 1 strongly influences metabolic and gene expression traits. a**, Lod score curves for metabolic traits scored in the B × H cross demonstrate that they are all driven by one or more QTL on chromosome 1. **b**, Lod score curves for expression traits corresponding to genes mapped as QTGs for the metabolic traits in **a** (*Apoa2* and *Tnfsf4*) or to genes within ten-million base pairs of *Apoa2* that give rise to strong, putative *cis* eQTL and that are significantly correlated with at least one of the metabolic traits depicted in **a**.

In a previously described cross between C57BL6/J (B6) and C3H/HeJ (C3H) on an *Apoe*<sup>-/-</sup> background (referred to here as the B × H cross)<sup>17</sup>, the importance of distal chromosome 1 as a key driver of metabolic traits became apparent because every metabolic trait scored in the B × H cross links to this region of the chromosome (Fig. 1a). *Tnfsf4* and *Apoa2* are located within 10 megabases (Mb) of one another and are proximal to the peak log likelihood ratio (lod) score curves for the metabolic traits on chromosome 1. These two genes were positionally cloned from the B × H background and validated using transgenic and knockout animals as having a causal relationship with plasma cholesterol and high-density lipoprotein (HDL) levels, fat mass, weight, insulin levels and atherosclerotic lesion size<sup>19,21,22</sup>. *Apoa2* was specifically identified as having a mutation in C3H relative to B6 that affected *Apoa2* translational efficiency, leading to lower liver transcript and protein levels in C3H relative to B6 (refs 22 and 24). Liver gene expression traits scored in the B × H cross provide a unique opportunity to confirm *Apoa2* as a QTL and to assess its total contribution to the metabolic traits. Because the expression of *Apoa2* and its association to the chromosome 1 linkage region and metabolic traits can be considered simultaneously on the mixed genetic background in which the disease trait QTL were originally mapped, the gene can be validated in the exact context in which it was identified.

*Apoa2* liver gene expression in the B × H cross gave rise to a significant expression QTL (Fig. 2a) that was proximal to the *Apoa2* structural gene, confirming that *Apoa2* expression is significantly perturbed between B6 and C3H mice as previously reported<sup>22</sup>. However, of the eight metabolic traits tested (Fig. 1a), *Apoa2* liver expression levels were only modestly correlated with glucose levels (expected *P* value = 0.014), and not at all correlated with obesity traits (Supplementary Fig. 1a). Interestingly, *Apoa2* gene expression was strongly supported as being independent of each of the metabolic traits with respect to the chromosome 1 locus (see Fig. 2a, b for weight). Results for *Apoa2* liver protein expression in the B × H cross were consistent with these gene expression results (Supplementary Results). Although the lack of association between *Apoa2* expression and the metabolic traits cannot exclude *Apoa2* as at least one of many genes underlying the chromosome 1 metabolic trait QTL, it is consistent with genes other than *Apoa2* having a more dominant role in this linkage region. *Tnfsf4* was similarly examined in the B × H cross but was not found to be associated with any of the metabolic traits linked to chromosome 1 in the B × H cross (Supplementary Results). However, because heart and aorta were demonstrated as the relevant tissues for *Tnfsf4* activity associated with metabolic traits<sup>21</sup>, our failure to detect an association in this instance may be because we have not profiled the relevant tissue.

Whereas the expression data in this specific B × H cross did not support *Apoa2* and *Tnfsf4* as having a causal relationship with the metabolic traits, we identified 112 liver expression traits corresponding to genes located in the chromosome 1 linkage region (from 90 Mb to the end of the chromosome) that gave rise to expression QTL (eQTL) in this region supporting the metabolic trait QTL (Supplementary Table 1). Although none of these genes completely explains the linkage of the clinical traits to chromosome 1, the expression levels of 54 of these genes are statistically supported as at least partially explaining variation in the metabolic traits in a causal way<sup>11</sup> (Supplementary Table 1), suggesting that there may be many genes in this region that support the metabolic trait QTL. Figure 1b highlights strong liver *cis* eQTL for 4 of these 54 genes that are physically located within 10 Mb of *Apoa2* as well as the peak lod scores for each of the metabolic traits. Upstream transcription factor 1 (*Usf1*) was identified as a susceptibility gene for familial combined hyperlipidemia (FCH)<sup>25</sup>; F11 receptor (*F11r*) is supported as being a susceptibility gene for FCH and other inflammatory processes<sup>26,27</sup>; serum amyloid P component (*Apcs*) is implicated in atherosclerotic lesion formation<sup>28</sup>; and regulator of G-protein signalling 5 (*Rgs5*), a gene involved in vessel development and physiology, can distinguish the

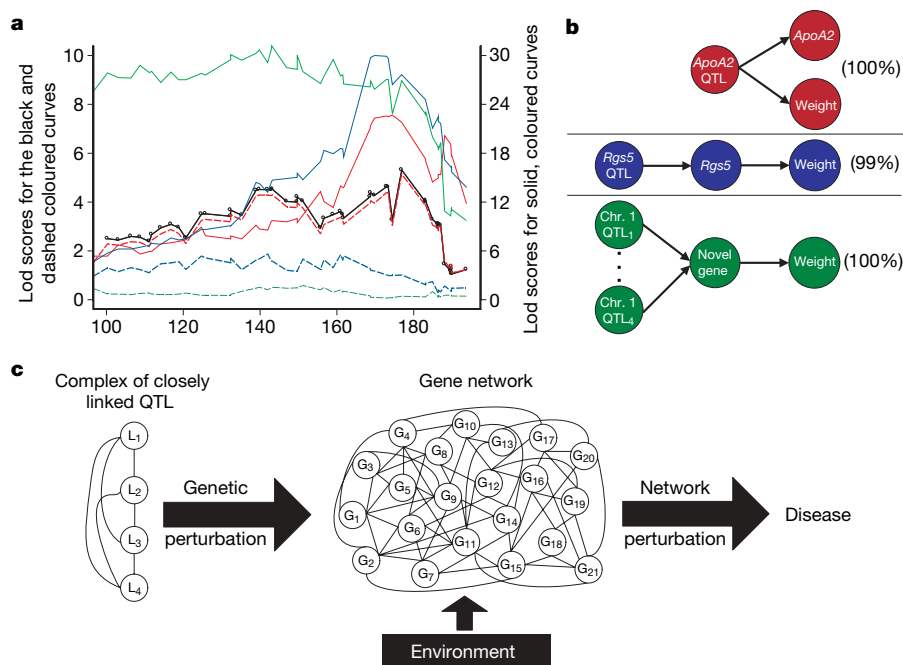
fibrous cap from other atherosclerotic plaque components<sup>29</sup> and has recently been associated with hypertension in humans<sup>30</sup>. Of these four expression traits, *Rgs5* is the most strongly associated with the metabolic traits linked to the chromosome 1 genomic region (see Fig. 2 and Supplementary Fig. 1c for weight). Therefore, unlike *Apoa2* and *Tnfrsf4*, these expression traits are significantly correlated with the metabolic traits, are strongly linked to the chromosome 1 locus, are physically located near the chromosome 1 linkage peaks, and are strongly supported as having a causal relationship with the metabolic traits.

The extensive linkage disequilibrium operating in the B × H cross, the number of possible QTGs in this region, the small-to-modest effects of each QTG and potential interactions among the QTGs make dissecting the individual contributions of the QTGs in the chromosome 1 region nearly impossible from the cross data alone. However, using gene expression data scored in the B × H cross, expression traits that capture the multiple genetic perturbations in this region and that in turn lead to variations in the metabolic traits<sup>11,31</sup> can be more readily identified. As an example, Fig. 2a highlights transcript abundances for an uncharacterized gene (GenBank accession number, BB433460) that is positioned in an intron of intraflagellar transport 88 homologue (*Ift88*). The liver expression of this gene is highly correlated with metabolic traits such as obesity (Supplementary Fig. 1d), is significantly linked across the entire distal half of chromosome 1 (lod score > 8 across most of the distal half of chromosome 1) and is supported as having a large contribution to the weight trait (Fig. 2a, b). Although BB433460 physically resides on chromosome 14, it captures more of the genetic variation driving the metabolic traits at the chromosome 1 locus than any of the genes physically located in this region, suggesting that networks of expression traits may be perturbed *in trans* by this complex of closely linked QTL and, as a result, lead to variation in the metabolic traits.

### Network changes induce phenotypic change

Liver and adipose co-expression networks were reconstructed from the B × H data to identify components of these networks that, like BB433460, mediate the transfer of information from QTL in the chromosome 1 region to the metabolic traits. Supplementary Fig. 3a depicts the most highly connected expression traits in this network as an ordered connectivity matrix. The pattern of distinct clusters or sub-networks that emerge among the highly connected nodes in liver and adipose (Supplementary Fig. 3) are notable and support a hierarchical structure in these networks (Supplementary Fig. 4). The different sub-networks highlighted are seen to be enriched for a number of biological processes (Supplementary Table 2), including insulin signalling (sub-network 1), inflammation (sub-network 5), muscle-related processes (sub-network 7) and cell cycle (sub-network 9). These sub-networks represent key functional units that make up the co-expression network and that underlie processes specific to the different cell types that constitute each tissue. For example, in the female liver co-expression network, sub-network 5 is enriched for genes involved in inflammatory processes, potentially reflecting activity in Kupffer cells. Sub-network 7 is enriched for muscle-related genes such as actin and myosin, potentially reflecting hepatic stellate cell activity, where these cells are known to control microvascular tone and, when activated, can turn into myofibroblasts and express smooth muscle actin filaments and desmin.

The sub-networks represent different sets of overlapping pathways and are readily seen to be enriched for genes that are perturbed by specific genetic loci. For example, 85% of the genes in liver sub-network 1 give rise to eQTL on chromosome 1 (Supplementary Fig. 5). To establish whether a given sub-network was supported as having a causal relationship with the metabolic traits linked to chromosome 1, we used a statistical procedure to test whether the gene expression traits in each sub-network supported a causal, reactive or independent relationship with each of the metabolic traits with



**Figure 2 | Genetic loci perturb molecular phenotypes that in turn lead to variations in disease-associated traits.** **a**, Lod score plots for weight (solid black line), *Apoa2* liver expression (solid red), *Rgs5* liver expression (solid blue) and BB433460 liver expression (solid green) traits in the B × H cross. The dashed curves represent the lod score curves for weight conditional on the *Apoa2* (dashed red), *Rgs5* (dashed blue) and BB433460 (dashed green) liver gene expression traits. Conditioning on *Apoa2* expression does not significantly reduce the weight lod score (independent relationship), whereas conditioning on *Rgs5* or BB433460 does (causal relationship).

**b**, Relationships supported between the expression and weight traits described in **a**: *Apoa2* (top), *Rgs5* (middle) and BB433460 (bottom) are predicted to be related to weight in an independent (*Apoa2*) and causal (*Rgs5* and BB433460) way. Percentages represent the number of times the model shown was inferred out of 1,000 random samples drawn from the B × H cross. **c**, Generalization of the relationship discovered between BB433460 and weight, in which genetic loci ( $L_i$ ) and environment perturb molecular networks of genes ( $G_j$ ) that in turn leads to disease.



respect to the genetic loci driving metabolic traits scored in the B × H cross: abdominal fat mass, weight, plasma insulin levels, free fatty acids, total plasma cholesterol levels and aortic lesion sizes. We identified a sub-network as having a causal relationship with a given metabolic trait if it was significantly enriched ( $P < 0.01$ ) for expression traits that have been supported as having a causal association with that trait. For liver, only five sub-networks were identified as being enriched for at least one of the metabolic traits (Supplementary Fig. 3c). Two of the sub-networks were weakly enriched for insulin, fat mass, weight or cholesterol candidate causal genes (sub-networks 6 and 14), whereas sub-networks 2 and 9 were strongly enriched for only cholesterol and weight candidate causal genes, respectively. However, one of the sub-networks (sub-network 5) was very significantly enriched for expression traits supported as having a causal relationship with every metabolic trait tested, directly implicating this sub-network as a key mediator of the genetic loci driving variation in the metabolic traits scored in the B × H cross (Supplementary Fig. 3c). This sub-network was also the most highly conserved between the sexes and tissues in the B × H cross. In fact, 90% of the genes in female liver sub-network 5 overlapped a corresponding male sub-network ( $P < 10^{-305}$  by the Fisher Exact Test), and 50% of these genes overlapped a corresponding adipose sub-network ( $P \sim 6.47 \times 10^{-147}$  by the Fisher Exact Test). Furthermore, the adipose sub-network corresponding to liver sub-network 5 was the only adipose sub-network found to be significantly enriched for expression traits supported as having a causal relationship with all of the metabolic traits tested (Supplementary Fig. 3d).

### A macrophage sub-network causes disease

To explore the strong pleiotropic effects of sub-network 5 on the metabolic traits in the B × H cross, we formed a supermodule by combining this sub-network with the corresponding sub-network identified in the adipose co-expression network (Supplementary Table 3). Compared to the individual sub-networks, this supermodule systematically increased the fold-change enrichments and corresponding significance scores for expression traits supported as having a causal relationship with the metabolic traits (Table 1). In fact, the percentage of expression traits in this supermodule supported as having a causal relationship with aortic lesions, weight or fat mass, plasma insulin or glucose levels, total cholesterol and HDL cholesterol were 75%, 50%, 45%, 50% and 47%, respectively (Supplementary Table 4). The probability that these overlaps occurred by chance are small. For example, the probability that 50% of the 762 expression traits supported as having a causal relationship with obesity fall in this single supermodule (out of the 23,574 transcripts represented on the array) is  $2.30 \times 10^{-262}$ . We also searched this

supermodule comprised of 1,406 transcribed sequences against a body atlas of gene expression representing 60 distinct mouse tissues. For each tissue in the atlas, gene sets were formed on the basis of tissue-specific expression (Supplementary Methods) and these sets were intersected with the supermodule. Bone-marrow-derived macrophages and spleen were the two most enriched tissues (Table 1 and Supplementary Table 4), not liver and adipose as one might expect given the module origins. These enrichments, combined with the significant enrichment of genes in inflammatory pathways, suggest that this module reflects the significant macrophage populations resident in liver and adipose tissues. This macrophage connection is further supported by a number of known macrophage markers represented in this supermodule, including *Cd14*, *Cd68* and *Emr1* (refs 32–34). Given the apparent macrophage-derived origins of this supermodule and its association with the metabolic traits in the B × H cross, we refer to it here as the macrophage-enriched metabolic network (MEMN) (Fig. 3a).

The MEMN is comprised of a number of expression traits corresponding to genes that we recently identified and validated as having a causal relationship with obesity traits, including *Zfp90* (ref. 11), *Tgfb2* (ref. 11), *C3ar1* (ref. 11) and *Alox5ap* (arachidonate 5-lipoxygenase-activating protein)<sup>31</sup>. Because this network comprises a highly interconnected set of expression traits supported as having a causal relationship with the different metabolic traits, we hypothesized that perturbing single genes in the MEMN that had been previously validated as having a causal relationship with these traits would significantly perturb the entire MEMN. To test this, we constructed single gene perturbation signatures for two of the genes, *Zfp90* and *Alox5*, recently validated as having a causal relationship with obesity-associated traits<sup>11,31</sup>. In addition, we constructed a single gene perturbation signature for *Pparg*, a gene that also resides in the MEMN and that has previously been validated as having a causal relationship with obesity and diabetes traits<sup>35</sup>. In all cases, the perturbation signatures (Supplementary Table 4) were significantly enriched for expression traits in the MEMN (Table 1). For example, the *Zfp90* transgenic signature comprised approximately 3,000 expression traits; 468 of these overlapped the MEMN, whereas only 179 would have been expected by chance—a greater than 2.5-fold enrichment (Fisher Exact  $P$  value =  $4.83 \times 10^{-94}$ ). Furthermore, genes validated as having a causal relationship with obesity were observed in these different perturbation signatures. For example, *Pparg* falls in the *Zfp90* signature, whereas *Tgfb2* and *C3ar1* fall in the *Pparg* and *Alox5* signatures, respectively. More generally, all signatures are enriched for expression traits supported as having a causal relationship with the metabolic traits. Therefore, expression traits supported as having a causal relationship with the metabolic traits

**Table 1 | Gene sets significantly over-represented in the MEMN**

Gene set type	Gene set description	Gene set count*	Overlap (fold enrichment)†	Enrichment nominal $P$ value (corrected $P$ value)‡
GO biological process categories	Immune response	1,503	246 (2.6)	$4.26 \times 10^{-43}$ ( $1.94 \times 10^{-39}$ )
	Defence response	1,565	251 (2.4)	$1.97 \times 10^{-42}$ ( $8.98 \times 10^{-39}$ )
	Inflammatory response	584	110 (2.8)	$4.66 \times 10^{-24}$ ( $2.12 \times 10^{-20}$ )
Tissue-specific expression	Bone-marrow-derived macrophage specific expression	289	65 (3.3)	$1.10 \times 10^{-18}$ ( $1.04 \times 10^{-16}$ )
	Spleen-specific expression	186	47 (3.8)	$7.56 \times 10^{-15}$ ( $5.81 \times 10^{-14}$ )
Environmental perturbations	Diet-induced obesity versus wild-type signature	1,108	415 (6.2)	$5.17 \times 10^{-232}$
Causal gene sets	Genes supported as causal for atherosclerotic lesions	159	119 (12.4)	$3.22 \times 10^{-111}$
	Genes supported as causal for obesity traits	762	375 (8.2)	$2.30 \times 10^{-262}$
	Genes supported as causal for diabetes	589	272 (7.7)	$4.76 \times 10^{-176}$
	Genes supported as causal for total cholesterol levels	245	131 (8.9)	$1.01 \times 10^{-93}$
	Genes supported as causal for HDL levels	77	36 (7.8)	$7.98 \times 10^{-24}$
Single gene perturbation experiments	<i>Zfp90</i> transgenic signature	3,006	468 (2.6)	$4.83 \times 10^{-94}$
	5-LO knockout signature	5,264	605 (1.9)	$5.95 \times 10^{-70}$
	Rosiglitazone signature	837	118 (2.3)	$3.03 \times 10^{-18}$

\* The number of sequences in the MEMN used to compare to these gene sets is 1,406.

† The overlap count is computed by counting the number of genes in the intersection between the indicated gene set and the MEMN. The fold enrichment is computed as the observed overlap count divided by the expected overlap count, estimated by multiplying the MEMN transcript count (1,406) by the fraction 'gene set count divided by total gene count (23,574)'.

‡ Nominal  $P$  values represent the significance of the Fisher Exact Test statistic under the null hypothesis that the frequency of the indicated gene set is the same between a reference set of all transcripts represented on the array and the set of genes comprising the MEMN. The corrected  $P$  values represent the Bonferroni-corrected  $P$  values (nominal  $P$  value multiplied by the number of gene sets searched).

falling in the MEMN and moving this network when perturbed provide direct support that the metabolic traits are an emergent property of this network, with hundreds of expression traits supported as having a causal relationship with the metabolic traits.

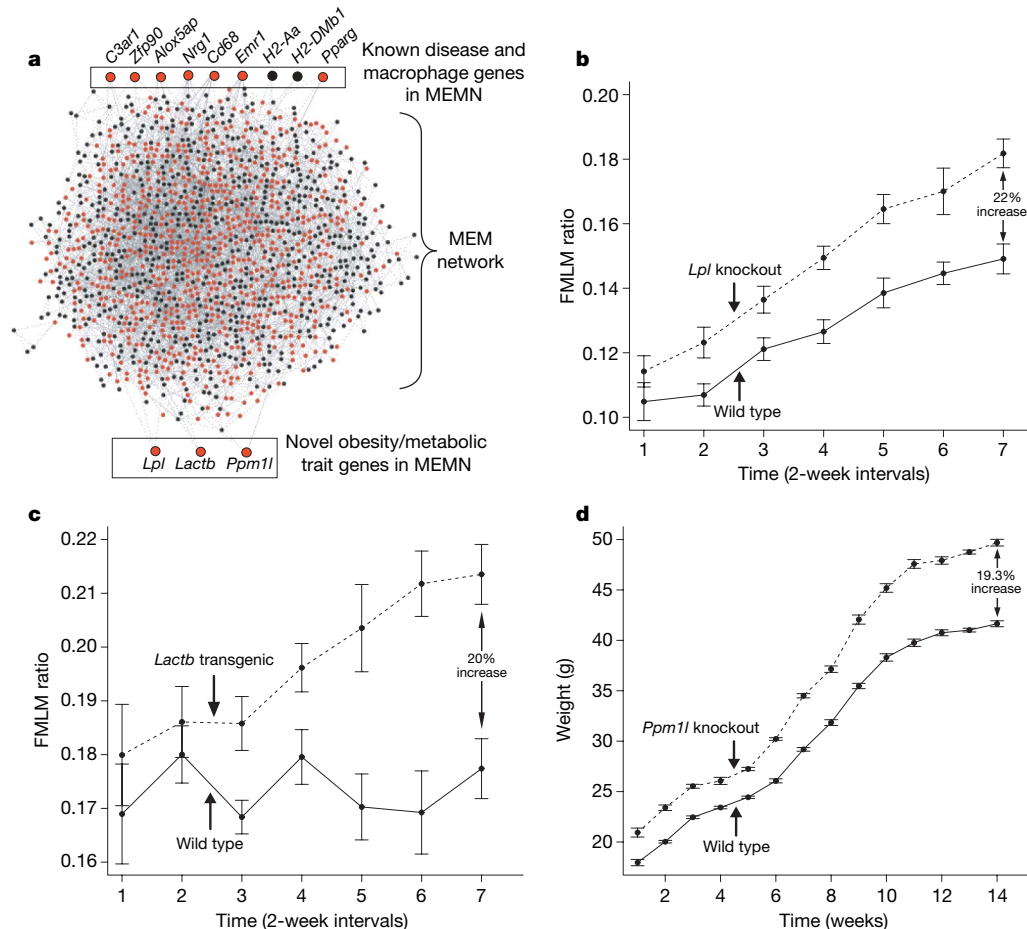
### *Lpl* and *Lactb* validated as obesity genes

In the MEMN, there were 375 expression traits supported as having a causal relationship with the obesity traits linked to the chromosome 1 locus. Although many of the genes corresponding to the expression traits in this network have been validated as having a causal relationship with metabolic traits (*Pparg*, *Alox5*, *Tgfr2*, *C3ar1* and *Zfp90*, to name just a few), many others have not. We used replication over multiple studies as a way to prioritize genes for validation. Genes supported in multiple independent experiments as having a causal relationship with disease are more likely to be truly causal. Therefore, we intersected the MEMN with a set of genes we previously predicted to have a causal relationship with obesity in a completely independent experiment<sup>11</sup>. Three of the ten genes predicted in an independent F<sub>2</sub> intercross population<sup>11</sup> were represented in the MEMN: *Zfp90*, *Lpl* and *Lactb*. *Zfp90* has already been validated as having a causal relationship with obesity, so we proceeded to validate the other two 'replicated' genes.

*Lpl* has previously been supported as a susceptibility gene for atherosclerosis- and diabetes-associated traits<sup>36</sup>. However, an association between *Lpl* and obesity has not been established. To our knowledge, *Lactb* has not ever been associated with any of the B × H metabolic traits. Given the prediction that *Lpl* and *Lactb* have

a causal relationship with obesity, we recorded weight, fat mass and lean mass for *Lpl*<sup>+/-</sup>, *Lactb* transgenic mice and wild-type littermate controls every 2 weeks starting at 11 weeks of age using quantitative NMR. As predicted, the growth curves for the *Lpl*<sup>+/-</sup> and *Lactb* transgenic animals were significantly different from those of controls (Fig. 3b, c), with the fat-mass-to-lean-mass (FMLM) ratio difference generally increasing over time. At the final quantitative NMR measurement, the FMLM ratios in the *Lpl*<sup>+/-</sup> and *Lactb* transgenic mice were increased by 22% and 20%, respectively, over the wild-type controls ( $P = 1.09 \times 10^{-5}$  and  $P = 4.48 \times 10^{-5}$ , respectively).

*Lpl* is the principal enzyme responsible for the hydrolysis of circulating triglycerides and is active in differentiated macrophages<sup>37</sup>, consistent with its presence in the MEMN. Although *Lpl* has not previously been functionally validated as a susceptibility gene for obesity, several studies have established an inverse relationship between *Lpl* activity and obesity-related traits, including a negative correlation observed between *Lpl* activity and percentage body fat in humans<sup>38</sup>. *Lactb* is a serine protease with high similarity to the bacterial lactamase gene, but very little is known about its function in eukaryotes<sup>39,40</sup>. Lactamase metabolizes peptidoglycan in the bacterial cell wall but neither the substrate nor the function of *Lactb* in eukaryotes is known<sup>41</sup>. *Lactb* has been detected in the mitochondria as part of the mitochondrial ribosomal complex<sup>42-44</sup>. Interestingly, a strain of rat that exhibits late-onset obesity was found to contain a mutation in the S26 subunit of the mitochondrial ribosome, at least partially explaining the obesity phenotype<sup>45</sup>.



**Figure 3** | Genes in the MEM network validated as having a causal relationship with obesity traits. **a**, The MEMN is enriched for genes supported as having a causal relationship with disease traits in the B × H cross (red nodes). The black nodes represent genes in the MEMN not supported as causal for disease traits in the B × H cross. **b**, FMLM ratio curves for *Lpl* knockout ( $n = 25$ ) and wild-type control ( $n = 23$ ) mice ( $P = 1.09 \times 10^{-5}$  that the difference at the last time point is significant).

**c**, FMLM ratio curves for the *Lactb* transgenic ( $n = 36$ ) and wild-type control ( $n = 27$ ) mice ( $P = 4.48 \times 10^{-5}$  that the difference at the last time point is significant). **d**, Weight curves for the *Ppm1l*<sup>-/-</sup> ( $n = 18$ ) and wild-type control ( $n = 18$ ) mice ( $P = 1.93 \times 10^{-11}$  that the difference at the last time point is significant). Error bars in **b-d** represent  $\pm 1$ s.d. of the indicated measures based on replicates and signal-to-noise ratios derived from the model applied to the weight and fat mass differences.

**Table 2 | Comparison of metabolic traits between *Ppm1l*<sup>-/-</sup> and *Ppm1l*<sup>+/+</sup> mice**

Trait	Age of mice (weeks)	<i>Ppm1l</i> <sup>-/-</sup>		<i>Ppm1l</i> <sup>+/+</sup>		Percentage change	Difference <i>P</i> value*
		Mean trait value	Sample size	Mean trait value	Sample Size		
Weight (g)	21	49.69	17	41.65	18	19.3	$1.93 \times 10^{-11}$
Total fat mass (g)	9	3.54	17	2.54	18	39.4	0.0037
Total fat mass (g)	20	22.10	17	15.06	18	46.7	0.00030
Baseline glucose (mg ml <sup>-1</sup> )	21	1.55	17	1.39	18	11.5	0.0075
OGTT area under curve (male mice only) (min (mg ml <sup>-1</sup> ))	21	186	8	279	9	-33.3	0.0069
OGTT insulin at 30 min (μg/l)	21	5.17	17	2.44	18	111.9	0.017
Free fatty acids (mequiv. l <sup>-1</sup> )	29	0.4116	14†	0.5457	17†	-24.6	0.00050
Non-invasive blood pressure (mm Hg)	25	90.13	17	86.07	18	4.7	0.027

\* All *P* values reported, except weight and OGTT AUC, represent the significance of the *t* statistic under the null hypothesis that the difference in mean, sex-adjusted trait values between the *Ppm1l* knockout and *Ppm1l* wild-type groups is equal to 0. For OGTT AUC, this same null hypothesis was tested but for males only. See Supplementary Methods for calculation of the *P* value for weight.  
 † By the 29 week time point, 3 male knockout mice and 1 male wild-type mouse had died.

### *Ppm1l* has a causal relationship with metabolic syndrome

Given the causal association between the MEMN and many metabolic traits, we rank-ordered genes on the basis of the number of metabolic traits for which they were supported having a causal relationship with (Supplementary Table 5) as an alternative to replication as a way to prioritize genes for validation. Four genes ranked at the top of the list: *Fgd6*, *Mmp27*, *BC032204* and *Ppm1l*. However, not only is *Ppm1l* a classically 'druggable' gene, but a knockout mouse for this gene was available from Deltagen, so we selected this gene for validation. *Ppm1l* is a newly discovered protein phosphatase, the function of which is not well characterized.

Weight, fat mass, insulin and glucose levels, blood pressure and other biochemical measures in blood were recorded in *Ppm1l*<sup>-/-</sup> and wild-type littermate controls. The growth curves for the knockout mice were significantly different from those of wild-type controls (Fig. 3d); at the final weight measurement, the knockout mice weighed 19.3% more than wild-type mice (Table 2). *Ppm1l*<sup>-/-</sup> mice also exhibited increased fat mass compared to wild-type controls, with an overall 46.7% increase in fat mass at 20 weeks of age (Table 2). At 21 weeks of age, an oral glucose tolerance test (OGTT) was performed on all mice. Baseline plasma glucose levels were observed to be 11.5% higher in *Ppm1l*<sup>-/-</sup> mice relative to wild-type mice. Male knockout mice demonstrated an improved glucose tolerance, with a 33.3% decrease in the area under the curve (AUC) relative to male wild-type mice (Table 2). In contrast, although glucose levels for females at the 60, 90 and 180 min time points were significantly increased (*P* value = 0.0077, 0.050 and 0.0043, respectively), the difference in AUC was not statistically significant (*P* value = 0.11). At the 30-min OGTT time point, insulin levels in male and female *Ppm1l*<sup>-/-</sup> mice were more than 100% increased compared to those of controls (Table 2). Blood was also collected in all mice at 29 weeks of age, and total cholesterol, triglycerides and free fatty acids were recorded. A significant decrease in free fatty acids was recorded in *Ppm1l*<sup>-/-</sup> mice relative to controls (Table 2), but no other major changes were observed for the other parameters (data not shown). Finally, given that the MEMN is supported as having a causal relationship with a number of traits associated with metabolic syndrome, and given the presence of genes such as ACE in this network, non-invasive blood pressure was monitored in all mice at 25 weeks of age. Overall, the blood pressure in *Ppm1l*<sup>-/-</sup> mice was significantly increased compared to that of controls (Table 2).

### Discussion

By integrating co-expression networks and genotypic data from an F<sub>2</sub> intercross population, we identified a liver and adipose macrophage-enriched sub-network that was associated with disease traits comprising the metabolic syndrome and enriched for expression traits supported as having a causal relationship with these traits. Unlike classic genetics approaches that aim to identify genes underlying genetic loci associated with disease, the approach developed here seeks to identify whole gene networks that respond *in trans* to genetic loci driving disease, and that in turn lead to variations in the disease

traits. Our results demonstrate that there may in fact be thousands of genes capable of increasing susceptibility to metabolic disease traits such as obesity, diabetes and atherosclerosis. Because the causal predictions made in this study rely on conditional dependency arguments that are statistical in nature, experimental validation is critical. Towards that end, *Lpl* and *Lactb* were identified and validated *in vivo* as previously unknown obesity genes, whereas *Ppm1l* was identified and validated as a gene capable of modulating multiple obesity, diabetes and hypertension traits.

Network-based approaches for elucidating the complexity of disease traits cast a broad net for genes that drive disease relative to classic genetic linkage or association studies that limit the search to genes that harbour DNA variations that associate with disease in the population under study. As a result, predictive networks provide the potential to identify hundreds of genes that drive disease and that could serve as points for therapeutic intervention. Our results support the idea that common forms of disease may be emergent properties of networks, where the networks associated with disease are highly interconnected, with many genes in the network potentially having a causal relationship with disease if perturbed strongly enough. With large-scale molecular profiling, genotypic and clinical data collected from large-scale populations, studying how a network of gene interactions affects disease will come to complement more strongly the classic focus of how a single protein or RNA affects disease. The integration of genetic, molecular profiling and clinical data has the potential to paint a more detailed picture of the particular network states that drive disease, and this in turn has the potential to lead to more progressive treatments of disease that may ultimately involve the targeting of whole networks as opposed to current therapeutic strategies focused on targeting one or two genes<sup>46</sup>.

### METHODS SUMMARY

Liver and adipose tissue were extracted from 334 F<sub>2</sub> animals in the B × H cross and profiled on an Agilent custom murine gene expression microarray<sup>17</sup>. All F<sub>2</sub> animals were genotyped at more than 1,300 single nucleotide polymorphism markers and clinically characterized with respect to obesity-, diabetes- and atherosclerosis-related traits<sup>17</sup>. The gene expression and genotype data were combined to construct co-expression networks comprised of the most highly connected nodes from each tissue and sex using previously described methods<sup>47</sup>. Highly interconnected sub-networks were then detected from each co-expression network using an iterative search algorithm<sup>47,48</sup>. QTL were detected for each of the expression and metabolic traits using a forward stepwise regression procedure<sup>17,49</sup>. QTL with pleiotropic effects on expression and metabolic traits were identified using a multivariate likelihood test<sup>11,50</sup>. The B × H QTL, expression and metabolic trait data were then integrated to assess whether each expression trait in each tissue was supported as having a causal relationship with each of the metabolic traits, with respect to QTL detected with pleiotropic effects on the expression and metabolic traits<sup>11</sup>. To identify sub-networks as having a causal relationship with the metabolic traits, each sub-network was tested for enrichment of expression traits supported as having a causal association with the metabolic traits using the Fisher Exact Test. Genes comprising the sub-network supported as having a causal relationship with all metabolic traits scored in the B × H cross were selected for validation on the basis of one of two criteria: the gene was supported as having a causal relationship with the metabolic traits in an



independent, previously published study, or the gene was supported as having a causal relationship with the most metabolic traits scored in the B × H cross. The three genes chosen for validation using these criteria were validated by constructing gene-knockout mouse strains (*Lpl* and *Pmp1l*) or transgenic mouse strains overexpressing the gene of interest (*Lactb*). Full Methods are provided in the Supplementary Information.

Received 5 October 2007; accepted 28 January 2008.

Published online 16 March 2008.

1. Edwards, A. O. *et al.* Complement factor H polymorphism and age-related macular degeneration. *Science* **308**, 421–424 (2005).
2. Haines, J. L. *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308**, 419–421 (2005).
3. Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385–389 (2005).
4. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
5. Frayling, T. M. *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894 (2007).
6. Strittmatter, W. J. *et al.* Apolipoprotein E: high-avidity binding to  $\beta$ -amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc. Natl Acad. Sci. USA* **90**, 1977–1981 (1993).
7. Brem, R. B., Yvert, G., Clinton, P. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
8. Bystrykh, L. *et al.* Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nature Genet.* **37**, 225–232 (2005).
9. Chesler, E. J. *et al.* Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nature Genet.* **37**, 233–242 (2005).
10. Monks, S. A. *et al.* Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.* **75**, 1094–1105 (2004).
11. Schadt, E. E. *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genet.* **37**, 710–717 (2005).
12. Schadt, E. E. *et al.* Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**, 297–302 (2003).
13. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C52 (1999).
14. Schadt, E. E., Sachs, A. & Friend, S. Embracing complexity, inching closer to reality. *Sci. STKE* **2005**, pe40 (2005).
15. Paigen, B., Albee, D., Holmes, P. A. & Mitchell, D. Genetic analysis of murine strains C57BL/6J and C3H/HeJ to confirm the map position of *Ath-1*, a gene determining atherosclerosis susceptibility. *Biochem. Genet.* **25**, 501–511 (1987).
16. Yang, X. *et al.* Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Res.* **16**, 995–1004 (2006).
17. Wang, S. *et al.* Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. *PLoS Genet.* **2**, e15 (2006).
18. Paigen, B. *et al.* *Ath-1*, a gene determining atherosclerosis susceptibility and high density lipoprotein levels in mice. *Proc. Natl Acad. Sci. USA* **84**, 3763–3767 (1987).
19. Castellani, L. W., Goto, A. M. & Lusis, A. J. Studies with apolipoprotein A-II transgenic mice indicate a role for HDLs in adiposity and insulin resistance. *Diabetes* **50**, 643–651 (2001).
20. Wang, X., Korstanje, R., Higgins, D. & Paigen, B. Haplotype analysis in multiple crosses to identify a QTL gene. *Genome Res.* **14**, 1767–1772 (2004).
21. Wang, X. *et al.* Positional identification of *TNFSF4*, encoding OX40 ligand, as a gene that influences atherosclerosis susceptibility. *Nature Genet.* **37**, 365–372 (2005).
22. Warden, C. H., Hedrick, C. C., Qiao, J. H., Castellani, L. W. & Lusis, A. J. Atherosclerosis in transgenic mice overexpressing apolipoprotein A-II. *Science* **261**, 469–472 (1993).
23. Welch, C. L. *et al.* Novel QTLs for HDL levels identified in mice by controlling for *Apoa2* allelic effects: confirmation of a chromosome 6 locus in a congenic strain. *Physiol. Genomics* **17**, 48–59 (2004).
24. Doolittle, M. H., LeBoeuf, R. C., Warden, C. H., Bee, L. M. & Lusis, A. J. A polymorphism affecting apolipoprotein A-II translational efficiency determines high density lipoprotein size and composition. *J. Biol. Chem.* **265**, 16380–16388 (1990).
25. Pajukanta, P. *et al.* Familial combined hyperlipidemia is associated with upstream transcription factor 1 (*USF1*). *Nature Genet.* **36**, 371–376 (2004).
26. Babinska, A. *et al.* F11-receptor (F11R/JAM) mediates platelet adhesion to endothelial cells: role in inflammatory thrombosis. *Thromb. Haemost.* **88**, 843–850 (2002).
27. Huertas-Vazquez, A. *et al.* Familial combined hyperlipidemia in Mexicans: association with upstream transcription factor 1 and linkage on chromosome 16q24.1. *Arterioscler. Thromb. Vasc. Biol.* **25**, 1985–1991 (2005).
28. Ezzahiri, R. *et al.* *Chlamydia pneumoniae* infections augment atherosclerotic lesion formation: a role for serum amyloid P. *APMIS* **114**, 117–126 (2006).
29. Adams, L. D., Geary, R. L., Li, J., Rossini, A. & Schwartz, S. M. Expression profiling identifies smooth muscle cell diversity within human intima and plaque fibrous cap: loss of RGS5 distinguishes the cap. *Arterioscler. Thromb. Vasc. Biol.* **26**, 319–325 (2006).
30. Chang, Y. P. *et al.* Multiple genes for essential-hypertension susceptibility on chromosome 1q. *Am. J. Hum. Genet.* **80**, 253–264 (2007).
31. Mehrabian, M. *et al.* Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits. *Nature Genet.* **37**, 1224–1233 (2005).
32. Austyn, J. M. & Gordon, S. F4/80, a monoclonal antibody directed specifically against the mouse macrophage. *Eur. J. Immunol.* **11**, 805–815 (1981).
33. Ramprasad, M. P., Terpstra, V., Kondratenko, N., Quehenberger, O. & Steinberg, D. Cell surface expression of mouse macrophage and human CD68 and their role as macrophage receptors for oxidized low density lipoprotein. *Proc. Natl Acad. Sci. USA* **93**, 14833–14838 (1996).
34. Wright, S. D., Ramos, R. A., Tobias, P. S., Ulevitch, R. J. & Mathison, J. C. CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein. *Science* **249**, 1431–1433 (1990).
35. Kubota, N. *et al.* PPAR $\gamma$  mediates high-fat diet-induced adipocyte hypertrophy and insulin resistance. *Mol. Cell* **4**, 597–609 (1999).
36. Hu, Y., Liu, W., Huang, R. & Zhang, X. A systematic review and meta-analysis of the relationship between lipoprotein lipase Asn291Ser variant and diseases. *J. Lipid Res.* **47**, 1908–1914 (2006).
37. Preiss-Landl, K., Zimmermann, R., Hammerle, G. & Zechner, R. Lipoprotein lipase: the regulation of tissue specific expression and its role in lipid and energy metabolism. *Curr. Opin. Lipidol.* **13**, 471–481 (2002).
38. Yost, T. J., Jensen, D. R. & Eckel, R. H. Tissue-specific lipoprotein lipase: relationships to body composition and body fat distribution in normal weight humans. *Obes. Res.* **1**, 1–4 (1993).
39. Liobikas, J. *et al.* Expression and purification of the mitochondrial serine protease LACTB as an N-terminal GST fusion protein in *Escherichia coli*. *Protein Expr. Purif.* **45**, 335–342 (2006).
40. Smith, T. S. *et al.* Identification, genomic organization, and mRNA expression of LACTB, encoding a serine  $\beta$ -lactamase-like protein with an amino-terminal transmembrane domain. *Genomics* **78**, 12–14 (2001).
41. Jacobs, C. Life in the balance: cell walls and antibiotic resistance. *Science* **278**, 1731–1732 (1997).
42. Gaucher, S. P. *et al.* Expanded coverage of the human heart mitochondrial proteome using multidimensional liquid chromatography coupled with tandem mass spectrometry. *J. Proteome Res.* **3**, 495–505 (2004).
43. Mootha, V. K. *et al.* Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* **115**, 629–640 (2003).
44. Taylor, S. W. *et al.* Characterization of the human heart mitochondrial proteome. *Nature Biotechnol.* **21**, 281–286 (2003).
45. Bains, R. K. *et al.* Visceral obesity without insulin resistance in late-onset obesity rats. *Endocrinology* **145**, 2666–2679 (2004).
46. Schadt, E. E. & Lum, P. Y. Reverse engineering gene networks to identify key drivers of complex disease phenotypes. *J. Lipid Res.* **47**, 2601–2013 (2006).
47. Lum, P. Y. *et al.* Elucidating the murine brain transcriptional network in a segregating mouse population to identify core functional modules for obesity and diabetes. *J. Neurochem.* **97** (suppl. 1), 50–62 (2006).
48. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabasi, A. L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
49. Haley, C. S. & Knott, S. A. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324 (1992).
50. Jiang, C. & Zeng, Z. B. Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* **140**, 1111–1127 (1995).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported in part by grants from the NIH/NIDDK and NIH/NHLBI to A.J.L. and T.A.D.

**Author Contributions** S.P., D.J.M. and M.-F.C. constructed and characterized the *Pmp1l* knockout mouse. X.Y., L.W.C., S.W., S.D., A.G., T.A.D. and A.J.L. constructed and characterized the B × H cross, the *Lpl* knockout mouse and the *Lactb* transgenic mouse. S.H., A.G., S.D. and B.Z. assisted in the co-expression network analyses. S.E. and A.J.L. performed bioinformatic analyses. All authors discussed the results and commented on the manuscript. S.K.S. and C.Z. aided in the data analysis. P.Y.L. and J.L. aided in the study design and interpretation of the experimental results. Y.C., J.Z. and E.E.S. designed the study, developed methods, analysed the data and wrote the paper.

**Author Information** The liver and adipose microarray data for the B × H cross have been deposited into the GEO database under accession numbers GSE2814 and GSE3086, respectively. Expression data associated with the diet-induced obesity, *Zfp90* transgenic, *Alox5*<sup>-/-</sup> and roziglitazone-treated animals have been uploaded to the GEO database under accession numbers GSE7028, GSE7029, GSE7026 and GSE7027, respectively. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to E.E.S. ([eric\\_schadt@merck.com](mailto:eric_schadt@merck.com)).