A Gene Expression Network Model of Type 2 Diabetes Links Cell Cycle Regulation in Islets with Diabetes Susceptibility

Mark P Keller, YounJeong Choi, Ping Wang, Dawn Belt Davis, Mary E Rabaglia, Angie T Oler, Donald S Stapleton, Carmen Argmann, Kathryn L Schueler, Seve Edwards, H Adam Steinberg, Elias Chaibub Neto, Robert Kleinhanz, Scott Turner, Marc K Hellerstein, Eric E Schadt, Brian S Yandell, Christina Kendziorski and Alan D Attie

Genome Res. published online Mar 17, 2008;
Access the most recent version at doi:10.1101/gr.074914.107

A Gene Expression Network Model of Type 2 Diabetes Links Cell Cycle Regulation in Islets with Diabetes Susceptibility

Mark P. Keller[1], YounJeong Choi[2,3], Ping Wang[2,3], Dawn Belt Davis[1], Mary E. Rabaglia[1], Angie T. Oler[1], Donald S. Stapleton[1], Carmen Argmann[5], Kathy L. Schueler[1], Steve Edwards[5], H. Adam Steinberg[1], Elias Chaibub Neto[3], Robert Kleinhanz[5], Scott Turner[7], Marc K. Hellerstein[6], Eric E. Schadt[5], Brian S. Yandell[2,3,4], Christina Kendziorski[2] and Alan D. Attie[1]*

University of Wisconsin Departments of [1]Biochemistry [2]Biostatistics & Medical Informatics, [3]Statistics, and [4]Horticulture, Madison, WI 53076. [5]Rosetta Inpharmatics, Seattle, WA 98109. [6]Department of Nutritional Sciences and Toxicology, University of California at Berkeley, CA 94720 and Division of Endocrinology and Metabolism, Department of Medicine, University of California at San Francisco, CA 94110, [7]KineMed, Emeryville, CA 94608.

[*]Corresponding author: adattie@wisc.edu

Abstract

Insulin resistance is necessary but not sufficient for the development of type 2 diabetes. Diabetes results when pancreatic beta-cells fail to compensate for insulin resistance by increasing insulin production through an expansion of beta-cell mass or increased insulin secretion. Communication between insulin target tissues and beta-cells may initiate this compensatory response. Correlated changes in gene expression between tissues can provide evidence for such intercellular communication. We profiled gene expression in six tissues of mice from an obesity-induced diabetes-resistant and a diabetes-susceptible strain before and after the onset of diabetes. We studied the correlation structure of mRNA abundance and identified 105 co-expression gene modules. We provide an interactive gene network model showing the correlation structure between the expression modules within and among the six tissues. This resource also provides a searchable database of gene expression profiles for all genes in six tissues in lean and obese diabetes-resistant and diabetes-susceptible mice, at 4 and 10 weeks of age. A cell cycle regulatory module in islets predicts diabetes susceptibility. The module predicts islet replication; we found a strong correlation between $^{2}H_2O$ incorporation into islet DNA *in vivo* and the expression pattern of the cell cycle module. This pattern is highly correlated with that of several individual genes in insulin target tissues, including IGF2, which has been shown to promote beta-cell proliferation, suggesting that these genes may provide a link between insulin resistance and beta-cell proliferation.

Introduction

Type 2 diabetes is a disorder that involves an increased demand for insulin brought about by insulin resistance, together with a failure to compensate with sufficient insulin production. Although Insulin resistance occurs in most obese individuals, diabetes is generally forestalled through compensation with increased insulin. This increase in insulin occurs through an expansion of beta-cell mass and/or increased insulin secretion by individual beta-cells. Failure to compensate for insulin resistance leads to type 2 diabetes.

One way to understand the pathophysiology of diabetes is to examine the coordinate changes in gene expression that occur in insulin-responsive tissues and pancreatic islets in obese animals that either compensate for insulin resistance or progress to type 2 diabetes. In each case, there are groups of genes that undergo changes in expression in a highly correlated fashion. By identifying groups of correlated transcripts (gene expression modules) during the compensation and development of diabetes, we can gain insight into potential pathways and regulatory networks in obesity-induced diabetes.

We study two strains of mice that differ in obesity-induced diabetes susceptibility. In this study, we surveyed gene expression in six tissues of lean and obese C57BL/6 (B6) and BTBR mice aged 4 wks and 10 wks. B6 mice remain essentially non-diabetic at all ages, irrespective of obesity. When obese, BTBR mice become severely diabetic by 10 weeks of age.

By analyzing the correlation structure of the genes under three contrast conditions, obesity, strain, and age, we identified gene expression modules associated with the onset of diabetes and provide an interactive co-expression network model of type 2 diabetes. We found a key module that is comprised of cell cycle regulatory genes. In the islet, the expression profile of these transcripts accurately predicts diabetes and is highly correlated with islet cell proliferation.

Results

Diabetes results from the convergence of age, strain, and obesity

To understand the relative contribution of age, strain, and obesity to the etiology of type 2 diabetes, we compared mice across three primary axes: time; 4 *vs*. 10 wks of age, body mass; lean *vs*. obese, and strain; C57BL/6 (B6) *vs*. BTBR mouse strain (**Fig. 1A**). B6 and BTBR mice differ in obesity-induced diabetes susceptibility. The B6 strain is essentially non-diabetic when it carries the *leptin*$^{ob/ob}$ (*ob*) mutation, whereas, by 10 weeks of age, the BTBR *ob* mouse is severely diabetic (**Fig. 1B**). Since the BTBR *ob* mouse is not yet diabetic at 4 wks of age, changes in gene expression at this time are potential causes rather than consequences of hyperglycemia. The ability of the B6 mouse to maintain euglycemia when challenged with morbid obesity is due to a >50-fold increase in circulating insulin at 10 wks of age (**Fig. 1C**). This contrasts with the relative hypoinsulinemia of the 10 wk old BTBR *ob* mouse. The difference in insulin at 10 wks of age is correlated with the number of islets harvested from the pancreas (**Fig. 1D**), suggesting that the ability to continue compensating for insulin resistance is a function of the ability to continue expanding β-cell mass.

We observed significant differences between B6 and BTBR mice in circulating adipose-derived hormones (adipokines). Adiponectin has been shown to regulate peripheral insulin sensitivity. There was a ~50% decrease in circulating adiponectin in

BTBR mice relative to B6 mice, irrespective of age and obesity (**Fig. 1E**). Plasminogen activator inhibitor-1 (PAI-1), a bio-marker for inflammation, showed an obesity-dependent increase at 4 and 10 wks only in BTBR mice (**Fig. 1F**). In rodents, resistin is highly expressed in adipose tissue and is thought to negatively regulate hepatic insulin sensitivity (Haluzik and Haluzikova 2006). Resistin showed an obesity-dependent increase in all but the diabetic BTBR mice (**Fig. 1G**). In summary, three key adipokines showed significant differences between the various groups of mice, consistent with a potential role for adipose tissue in the diabetes susceptibility of BTBR mice**.**

Differential expression of individual genes among 6 key tissues

Since age and obesity are necessary to unmask diabetes, we sought to deconstruct their relative contribution in B6 and BTBR mice by gene expression profiling of lean and obese male mice at 4 and 10 wks of age. In each strain, average expression levels from the four groups of mice can be sorted into 15 distinct theoretical patterns (**Table S1**). Using EBarrays, the empirical Bayes method described in Methods, for each transcript in each strain, we calculated the posterior probability for each of the 15 patterns and assigned the transcript to the expression pattern with maximum posterior probability (MPP). Differentially expressed (DE) transcripts are defined as those with MPP > 0.7 (MPP > 0.5 for hypothalamus) for one of the DE patterns (patterns 2-15 in Table S1) in at least one mouse strain. **Figure S1** plots the MPP for DE transcripts in B6 *vs*. the MPP for DE transcripts in BTBR. We found that > 96% of the DE transcripts were confined to 7 of the 14 possible DE patterns (shaded in **Table S1**). This approach has enabled us to identify primary and secondary drivers of changes in gene expression in the six tissues profiled. It is important to note that thresholds were chosen to balance FDR and number of genes identified (see legend of Figure S1). To ensure that modules identified were robust to more stringent thresholds for DE transcript identification, we varied the MPP cutoff and remapped the modules onto those identified at lower thresholds. For all tissues, except hypothalamus, the modules identified with MPP > 0.9 dendrogram tree were remapped onto the MPP > 0.8 and MPP > 0.7 dendrogram trees, maintaining their color designation. For hypothalamus the modules from the 0.6 tree were remapped onto the 0.5 tree, as the 0.7 tree had only one module (turquoise). **Figure S7** shows the result from islet (hypothalamus is similar). As shown, most clusters are highly conserved.

Primary *vs.* secondary drivers of differential gene expression at the individual transcript level

There were three variables in our experiment: obesity, age and strain. Using a non-supervised hierarchical algorithm, the 40 mice, 5 in each of 8 groups, were clustered based solely on the DE transcripts for each tissue. Differential gene expression in islet, liver, and adipose tissue was primarily driven by obesity, whereas differential expression in the two muscle tissues and hypothalamus was driven primarily by age (**Fig. S2**). Secondarily, age determined the clustering of the mice in islet, liver, and adipose tissue, whereas obesity drove the secondary changes in muscle and hypothalamus.

Each tissue contains modules of highly-correlated differentially expressed transcripts enriched for specific biological functions

To create a framework to explore strain, obesity and age-dependent determinants of gene expression, we restricted ourselves to the DE transcripts in each tissue and calculated the correlation coefficient among all transcripts, and partitioned them into

color-coded modules by the method of Zhang and Horvath (Zhang and Horvath 2005) (**Table S2**). Such modules often contain transcripts of related biological function (Carlson et al. 2006; Gargalovic et al. 2006; Ghazalpour et al. 2006; Horvath et al. 2006). Each tissue yielded 19 distinct co-expression network modules, except hypothalamus, which had 10, for a total of 105 modules (**Fig. S3**). The modules were largely cohesive, as quantified by measuring the pairwise correlations of transcripts within each module. Average absolute pairwise correlations exceeded 0.7 for 84%, 100%, 79%, 95%, 90% and 30% of the modules in islet, adipose, liver, solues, gastrocnemius and hypothalamus, respectively (data not shown).

To assess the biological relevance of the co-expression network modules we asked if modules contained genes enriched for specific biological processes. A substantial number of modules in each tissue were enriched for genes with specific gene ontology (GO) classifications. For example, in each tissue, except hypothalamus, we identified a single module significantly enriched with genes involved in cell cycle regulation ($p < 10^{-14}$ for each tissue) (**Table S4**). Thus, unsupervised clustering of genes with highly correlated expression profiles yields modules enriched for biologically relevant processes (Horvath et al. 2006).

We examined the contrast condition(s) responsible for differential expression for all DE transcripts in each of the 6 tissues: obesity, age and/or strain (**Fig. 2**; see also http://diabetes.wisc.edu/kelleretal2008/fig2.php for hyper-linked data page). Strain-dependent differences are evident when the pattern distribution for a particular color-coded module is shifted in the two strains. For example, the cell cycle module in islets (greenyellow, arrowhead) is predominantly "all different" in B6, reflecting a combination of obesity and age as primary drivers of DE for this module. However, a large fraction of these genes change to an "age only" pattern in BTBR, resulting from the loss in an obesity-dependent signal present in B6. In contrast to the regulation of cell cycle genes observed in the islets, a similar set of genes in adipose tissue (lightgreen, arrow head), also enriched in cell cycle regulatory transcripts, showed an "all different" expression pattern in both mouse strains. Table S3 lists the number of transcripts contained within each co-expression module as a function of expression pattern, strain and tissue (see also http://diabetes.wisc.edu/kelleretal2008/tabs3.php). In short, an unsupervised analysis of expression modules identified a key change in cell cycle gene expression in islets that distinguishes a diabetes-susceptible from a non-diabetes-susceptible mouse strain.

We identified a module of highly correlated genes in liver (turquoise module) that had a strain-specific expression pattern**.** Within this module is the glucagon receptor (*GcgR*), which shows an age-dependent increase in lean and obese B6 and BTBR mice and an obesity-dependent decrease in all mice except 10 wk B6. Hepatic glucagon receptor mRNA expression has been shown to increase in diabetic animals and in fasting conditions and its expression may be regulated by serum glucose levels via glucokinase (Burcelin et al. 1998). Inhibition of glucagon receptor mRNA expression in liver in *db/db* animals improves glucose tolerance and normalizes serum glucose, so it is clear that hepatic glucagon signaling plays a crucial role in the pathophysiology of diabetes (Liang et al. 2004)**.** As glucagon signaling in the liver leads to increased nutrient mobilization, we looked in the same module for genes involved in these processes**.** We identified genes involved in carbohydrate metabolism (isocitrate dehydrogenase 2, aldolase 1A, phosphoglycerate mutase-2, and 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3),

ketone body production (aldehyde dehyrogenase 1b1), and oxidative phosphorylation (succinate dehydrogenase A, mitochondrial F_0 complex, subunit c)**.** This suggests that by examining correlated genes within a module, we can identify strain-specific differences in key metabolic signals and their downstream effects on cellular metabolism**.**

The cell cycle modules predict adipose & beta-cell replication and obesity & diabetes

To maintain euglycemia, insulin-resistant animals must increase insulin production, which can occur through an expansion of pancreatic beta-cell mass. We focus here on two of the five tissue-specific modules that were highly enriched in genes controlling the cell cycle. The islet cell cycle module had 217 transcripts, which showed an age-dependent decrease in expression in both mouse strains and an obesity-dependent increase only in B6 mice (**Fig. 3A**). We calculated the 1st principle component (PC1) as a single descriptor of the expression pattern for the entire module (**Fig 3B**). The PC1 shows an obesity-dependent increase in cell cycle gene expression in the islets of B6 mice at 4 and 10 weeks of age. This induction fails to occur in BTBR mice. The same strain-specific differences in obesity-dependent induction of the islet cell cycle module were seen at 4 weeks, when the animals were still euglycemic. Thus, the islet cell cycle module changes precede the onset of diabetes.

The cell cycle module profile predicts that obesity induces islet cellular proliferation in B6 but not BTBR mice. To measure islet cellular proliferation, we exposed the mice to 8% ^2H_2O in the drinking water for two weeks and measured the enrichment in ^2H in DNA extracted from their islets. By normalizing the islet enrichment values to ^2H enrichment in bone marrow DNA, which undergoes complete turnover during this period, we estimated the percent new cells in the islets (Neese et al. 2002). B6 *ob* mice showed a 2.6-fold increase in the percent new islet cells relative to B6 lean mice (**Fig. 3C**). In contrast, BTBR islets showed no significant increase in the proportion of new islet cells in response to obesity.

In contrast to islets, the cell cycle module in adipose did not show a strain difference in expression pattern: both B6 and BTBR show an obesity-dependent up-regulation of expression for these genes at 10 weeks of age (**Fig. 3A, B)**. Consistent with the gene expression data, cellular proliferation measured in adipose tissue showed no strain difference in the fraction of newly replicated cells, with mice in both strains showing an obesity-dependent increase in cellular growth (**Fig. 3C**). Thus, the adipose cell cycle module PC1 correctly predicted adipose tissue proliferation and obesity, just as the islet cell cycle module predicted islet proliferation and diabetes.

Gene expression modules are highly correlated with glucose and insulin levels

Modules in islet, adipose and soleus most strongly correlated with plasma glucose (**Fig. S4**) contained genes previously shown to have a role in glucose homeostasis. For example, the cyan module in islet contains the branched chain aminotransferase 1 enzyme (*BCAT1*), which on its own had a 0.85 correlation with plasma glucose. *BCAT1* catalyzes the transamination of alpha-ketoisocaproate (KIC) and glutamate to yield leucine and alpha-ketoglutarate (alpha-KG). We have previously shown that BTBR islets are hyper-responsive to KIC-induced insulin secretion and that alpha-KG can directly stimulate insulin release from isolated pancreatic islets (Rabaglia et al. 2005).

The magenta islet module, also enriched with transcripts highly correlated with plasma glucose (**Fig. S4**), contains transcripts from a number of genes previously

highlighted for their potential role in glucose homeostasis, including: gamma-subunit of the gamma-aminobutyric acid A-receptor, *Gabrd* (plasma glucose correlation = 0.93); peroxisome proliferator activated receptor-alpha, *Ppar*-alpha (0.92); and cocaine and amphetamine regulated transcript, *Cart* (0.86). Glucose suppresses glucagon release by activation of the *GABA-A* receptors on alpha-cells, mediated by GABA released from neighboring beta-cells (Rorsman et al. 1989; Wendt et al. 2004). More recent evidence suggests that glucose may directly suppress glucagon release from alpha-cells (Vieira et al. 2007). Chronic treatment of Ins-1 cells or primary rat islets with a high glucose medium has been shown to decrease *Ppar*-alpha expression (Roduit et al. 2000). Finally, *Cart* expression is up-regulated in the beta-cells of type 2 diabetic models in rat (Wierup et al. 2006) and *Cart* knockout mice have impaired GSIS and are glucose intolerant (Wierup et al. 2005). Taken together, our results indicate that modules highly correlated with plasma glucose may identify compensatory changes in gene expression elicited by hyperglycemia**.**

Only islet, liver and adipose contained modules correlated with plasma insulin. Adipose contained the greatest number of modules correlated with insulin (11 of 19), whereas liver contained the module with the greatest absolute correlation with plasma insulin (**Fig. S4)**. Obesity is a well-known driver of changes in plasma insulin, due to an obesity-induced increase in insulin resistance. The lack of correlation for insulin in the muscle tissues and hypothalamus is consistent with their "age-responsive" expression patterns, with few obesity-related transcripts (**Fig 2**). For islets, liver and adipose, modules with a high correlation with insulin were generally not correlated with glucose.

We identified a module of correlated transcripts in adipose tissue enriched with genes involved in inflammation (the brown module, **Fig. S3**). These transcripts show increased expression in the obese animals, and older animals, of both strains. These data are consistent with the previous finding that adipose tissue from the B6 *ob/ob* mouse is enriched in transcripts found in macrophages (Weisberg et al. 2003; Xu et al. 2003). These cells are believed to secrete cytokines that affect peripheral insulin resistance. Up-regulated transcripts in our module include: *F4/80*, the macrophage-specific surface marker; *CD68*, a macrophage transmembrane protein; *CD11b*, an integrin found primarily on macrophages and other inflammatory cells; and *ADAM8*, a monocyte metalloproteinase. Transcripts in this module showed a high correlation (MS = 0.7) with plasma glucose (**Fig. S4**).

An inter-tissue gene expression network model

A major challenge in diabetes research is to better understand the potential relationships between gene expression changes among various tissues. For example, there are dramatic changes in gene expression in liver, muscle, and adipose tissue associated with insulin resistance and consequent changes in islets. Insulin-resistant tissues may communicate with islets through the production and secretion of blood-borne factors.

One way to search for molecules that mediate inter-organ communication is to identify relationships among the 105 modules across all 6 tissues. For each strain we represented each module with its strain-specific 1st principal component, ($PC1_{strain}$), which is highly correlated with all the module-specific transcripts (**Fig. S5**). Each module corresponds to a node within the network (**Fig. 4,** see also http://diabetes.wisc.edu/kelleretal2008/fig4.php for hyper-linked data page). Strain-specific edges are shown when two modules have significant partial correlation (PaCor).

PaCor is distinct from an ordinary correlation in that PaCor reveals the *direct* correlation between two PC1s after adjusting for the effects of all other module PC1s, as well as plasma glucose and insulin (see Methods). Adjusting for these two physiological traits has allowed us to more clearly establish direct gene-to-gene networks in B6 and BTBR**.** The strain-specific networks contained only 66 and 62 of 105 module nodes, forming just 125 and 161 significant edges out of potentially 5,460 edges, in B6 and BTBR, respectively (**Fig. 4**). Few connections were formed out of the total possible number (~2%), which is consistent with the assumptions made in the network construction algorithm (Schafer and Strimmer 2005a). Thus, in the diabetic strain, fewer network nodes were associated with a greater number of inter-tissue connections.

Strain-specific PaCor among the 105 $PC1_{strain}$ variables, plasma glucose and insulin, revealed remarkable strain differences (**Fig. S6**). In B6, there were 14 connections between glucose and gene modules, whereas in BTBR plasma glucose made no connections with any module after adjusting for insulin and other module nodes. For insulin, there were significant PaCor connections formed with 6 and 7 modules in B6 and BTBR mice, respectively, with only one in common. Many modules contain transcripts highly correlated with plasma glucose or insulin (**Fig. S4**).

There were substantial differences between tissues in the degree of intra-tissue connectivity in the co-expression network (**Fig. 4**). The two muscle tissues had the most intra-tissue connections, whereas liver and hypothalamus had the fewest, in both strains. Gastrocnemius showed a 2-fold increase in the number of intra-tissue connections in BTBR *vs.* B6 mice. The number of connections present in the other tissues was similar between the strains. There was 2-8-fold more inter-tissue than intra-tissue connections in all tissues, except gastrocnemius, where there were ~60% fewer inter-tissue connections**.** All tissues showed an increase in inter-tissue connectivity in BTBR *vs.* B6. Soleus and liver had the greatest number of inter-tissue connections, whereas gastrocnemius had the fewest. Overall, connection strength was greater for intra-tissue connections than for inter-tissue connections throughout the network. Taken together, our results suggest that obesity-dependent diabetes results in dramatic changes in the intra-tissue correlation structure of the co-expression network.

There were several nodes in each of the strain-specific networks that appear to be “hot spots”, or nodes that form the greatest number of connections within the network (**Fig. 4**, asterisks). These hot spots were found in both strains and in all tissues, excluding gastrocnemius, consistent with the relatively low inter-tissue connectivity of this tissue. Several of the hot spot nodes form as many as 10 or more connections with other nodes. It is interesting to note that these hot spots are highly inter-connected among themselves, as well as with other nodes.

The islet cell cycle module, $Islet_{greenyellow}$ (**Fig. 4**, arrow head) showed dramatic strain-dependent changes in its interconnectivity within the network. In the B6 network there is one connection between this node and a node derived from another islet module, $Islet_{cyan}$, with a negative association (**Fig. 4**, arrow head). Interestingly, the $Islet_{cyan}$ module contains transcripts with expression patterns highly correlated with plasma glucose (**Fig. S4**), suggesting that genes positively correlated with glucose, have negative association with the genes within the cell cycle module in B6 islets.

In the BTBR network, the intra-islet connection was lost, being replaced with two new connections: a negative association with the $liver_{magenta}$ node and a positive

association with the $soleus_{midnightblue}$ node (**Fig. 4**, arrow heads). Both of these nodes are highly interconnected within the network, forming 16 and 10 connections, respectively. The soleus node contains transcripts with expression patterns highly correlated with plasma glucose (**Fig. S4**), suggesting glucose-regulated genes in soleus may positively associate with cell cycle regulatory genes in BTBR islets.

The islet cell cycle module is highly correlated with individual transcripts in insulin target tissues

The correlations between modules do not allow the identification of instances where there are strong correlations between a module and a transcript not belonging to a module. To identify transcripts highly correlated with the islet cell cycle module, we calculated the correlation between the module's PC1 and all transcripts in each of the five non-islet tissues. We identified a few transcripts with high correlations to the islet cell cycle PC1, including: *Gdf10* (0.85, gastrocnemius), *BMP-1* (0.84, soleus), *Igf2* (0.86, soleus), *Igf2bp1* (0.85, adipose) and *NGF-beta* (0.8, soleus). These genes could potentially mediate a signal that promotes islet cellular proliferation. These results offer a hypothesis that one or more of these proteins mediates beta-cell proliferation in insulin-resistant mice.

A searchable type 2 diabetes gene expression database

We have created a searchable resource (http://diabetes.wisc.edu) of the gene expression data that was used to generate the network model described herein. This search tool allows the user to enter one to multiple genes and will display the gene expression profiles our 8 experimental groups (lean and obese B6 and BTBR mice at 4 and 10 weeks of age) in any of 6 tissues (islets, adipose, liver, gastrocnemius, soleus and hypothalamus). In addition, we have incorporated a transcript-to-transcript correlation tool that can be used to identify groups of genes with highly correlated expression profiles. For example, searching the database for the expression profiles for *Ccna2* (cyclin A2) shows that this transcript was included in 4 co-expression gene modules ($Gastrocnemius_{cyan}$, $Islet_{greenyellow}$, $Liver_{midnightblue}$ and $Solues_{green}$), all of which were significantly enriched with GO terms associated with cell cycle regulation (see **Table S4**). Plots of the expression profiles reveal, in general, that *Ccna2* decreases with age in all tissues and increases with obesity in adipose for both mouse strains and in islet for B6 mice only. Searching the mlratio database in islets for transcripts that correlate with *Ccna2* yields a list of 100 transcripts with a Pearson's correlation, $r > 0.94$. Many of these highly correlated transcripts themselves are involved in cell cycle regulation, including *FoxM1* ($r = 0.99$), *Ccnb1* ($r = 0.98$), *Ccnb2* ($r = 0.98$), *Brca1* ($r = 0.97$) and *Aurka* ($r = 0.96$), and are included in the cell cycle regulatory modules. In addition to positively correlated transcripts, the most negatively correlated transcripts are displayed, which in the case of *Ccna2*, included *Ccnk* ($r = -0.77$), a cyclin that has been shown previously to suppress cellular proliferation when over-expressed in human glioblastoma cells (Mori et al. 2002). This brief example has illustrated the utility of our searchable database. It can be used to survey gene expression in 6 key tissues as a function of obesity, strain and age in a mouse model of type II diabetes. We believe it will be a powerful resource tool that will benefit many in the diabetes research community.

Discussion

During the transition to diabetes, tissues undergo coordinated changes in gene expression and arrive at a new highly-regulated, although pathological, steady-state. These changes are highly correlated and thus enable us to identify modules of coordinately regulated genes. In this study, we exploited three primary variables that converge to cause type 2 diabetes in our model: obesity, age, and mouse strain, to study the correlation structure of the changes in gene expression in 6 tissues. We obtained a gene expression network model containing 105 modules and establish the modules' inter and intra-tissue relationships in the non-diabetic and diabetic state.

A major finding in our study is that in 5 of the 6 tissues profiled, at least one of the tissue-specific modules was significantly enriched with cell cycle regulatory genes. Of direct relevance to diabetes pathogenesis, the islet cell cycle module exhibited a striking strain difference in expression pattern; B6, but not BTBR, showed an obesity-dependent increase in expression at both 4 and 10 weeks of age.

These results predict a strain difference in beta-cell proliferation during the onset of obesity that would correlate with resistance or susceptibility to diabetes. This prediction was tested with a direct measure of cellular proliferation. We found that B6 islets have a robust, obesity-dependent increase in islet cell replication, consistent with previous reports. In contrast, BTBR islets failed to increase proliferation in response to obesity. Thus, the islet cell cycle module predicted diabetes.

Given the dramatic islet growth phenotypes that have been reported for the *Cdk4* knockout (Rane et al. 1999; Tsutsui et al. 1999) and constitutively-active transgenic (Rane et al. 1999) mouse, we asked whether *Cdk4* or any of the D-type cyclins were present in the islet and other tissue-specific cell cycle modules. These genes were not present in the cell cycle modules, suggesting that age, obesity and strain were not factors that alter their expression in our experimental model. However, cyclins (Ccn) A, B and E were identified in several cell cycle modules, suggesting these molecules play a critical role in obesity and age-related changes in cell cycle progression. *CcnA* was present in the cell cycle modules in all tissues except adipose, whereas *CcnE* was unique to adipose.

There are a number of Cdk inhibitors (Cdkn), but only two that were identified in the cell cycle modules: *Cdkn2c* (*p18*) in the liver module and *Cdkn3* in both liver and islet modules. Similar to the partnership that is formed between *Cdk4/6* and *CcnD*, *Cdk2* partners with either *CcnE* or *CcnA* to mediate phosphorylation of retinoblastoma tumor suppressor protein (*pRb*) at sites distinct from those phosphorylated by *Cdk4/6* (Cozar-Castellano et al. 2006). Remarkably, *Cdk2* was uniquely included in the cell cycle module of the islets, suggesting that *Cdk2*-dependent phosphorylation of *pRb* may be a key regulatory step that mediates the strain difference in islet cell proliferation between B6 and BTBR mice. *pRb* is widely regarded as the molecular "brake" that controls transition from G1 into S phase of cellular growth (Cozar-Castellano et al. 2006). Once relieved of *pRb*-dependent inhibition, a family of E2F transcription factors mediates coordinate regulation of gene expression that is required for cellular replication. However, it is important to note that beta-cell specific ablation of *pRb* does not lead to pancreatic beta-cell mass or glucose homeostatic phenotypes, suggesting that other factors (*e.g.*, the additional pocket proteins, *p107* or *p130*), can compensate to achieve growth arrest in the absence of *pRb* (Vasavada et al. 2007).

E2F1 and *E2F2* were found in soleus and liver, whereas *E2F8* was in adipose and islet cell cycle modules. *E2F7* was found exclusively in the islet cell cycle module. Our results reveal key molecular components of the cell cycle regulatory machinery that form expression pathways in most of the tissues profiled. Some of these components are uniquely contained in islet pathways and may play a critical role in islet cell proliferation during aging or under the stimulus evoked by obesity-induced insulin resistance.

The cell cycle module results also predicted an obesity-dependent increase in adipose cell proliferation in both mouse strains. The prediction was tested *in vivo* and again, the results are entirely consistent with the cell cycle module data; both mouse strains showed an obesity-dependent increase in adipose cell proliferation.

What is the link between obesity and an increase in islet cellular proliferation? One possibility is expression and secretion of mitogenic factors from peripheral insulin-resistant tissues that circulate in the blood, and stimulate beta-cells to proliferate, a mechanism supported by transplantation studies (Flier et al. 2001). Perhaps B6 mice express a beta-cell mitogenic signal in peripheral tissues and this signal is missing or not functional in BTBR mice.

To search for possible candidates for these factors, we identified genes in insulin target tissues tissues with expression profiles highly correlated to the islet cell cycle regulatory module and whose products were putative secreted peptides. Several candidates were identified, including *NGF*-beta, *IGF2*, *IGF2bp1, Gdf10 and BMP-1*.

Several studies have provided compelling evidence that these molecules play a critical role in regulating beta-cell mass. Widespread (Petrik et al. 1999) or local (Devedjian et al. 2000; Okamoto et al. 2006) over-expression of *IGF2* has been shown to promote islet cell hyperplasia, and application of *IGF2* to beta-cells in culture, induces proliferation (Milo-Landesman and Efrat 2002). Recent genome-wide association studies have identified genetic variants in *IGF2bp2* that are associated with type 2 diabetes in human patients (Saxena et al. 2007; Scott et al. 2007; Zeggini et al. 2007). While distinct from *IGF2bp2, IGF2bp1* functions similarly to *IGF2bp2* to stabilize *IGF2* mRNA, resulting in increased synthesis and secretion of *IGF2* protein. Removal of *NGF*-beta from isolated islets in culture induces apoptosis (Pierucci et al. 2001). Finally, recent work has shown that *BMP-3*, a member of the BMP ligand super-family that includes *Gdf10* and *BMP-1*, may play a critical role in regulating beta-cell mass. *BMP-3* knockout mice have decreased islet *Mki67*-positive immunoreactivity, reduced beta-cell mass and increased random-fed blood glucose, suggesting a role for *BMP-3* in regulating beta-cell proliferation (Lee et al. 2007).

Recent evidence suggests that beta-cell replication is the primary means by which animals increase beta-cell mass during adulthood and under conditions of islet regeneration (Dor et al. 2004). All beta-cells have equal capacity to replicate (Brennand et al. 2007) and this replicative capacity requires functional *cyclin-D2* (Georgia and Bhushan 2004). Mice harboring beta-cell specific ablations of the transcript factor *FoxO1* (Okamoto et al. 2006) or *InsR* (Okada et al. 2007) fail to show islet cell hyperplasia in response to severe insulin resistance. Taken together, these data strongly support a model whereby the expansion of beta-cell mass in response to insulin resistance is due to replication of pre-existing beta-cells. Our finding of obesity-dependent differences in the expression of genes critical for cell cycle regulation in isolated islets, coupled with our direct *in vivo* measure of islet cellular proliferation corroborates this model.

These studies provide the scientific community with the first gene expression network model of type 2 diabetes across multiple tissues. In addition, it provides a large database inferring intra- and inter-tissue connections between gene expression modules across a wide array of cellular functions. The modules can be broadly used to make predictions about the regulation of numerous pathways as we did with the cell cycle module.

Methods

Animals

All mice used in this study were male, bred from our in-house colonies at the University of Wisconsin Biochemistry Department and housed in an environmentally-controlled facility on a 12 hr light/dark cycle (6 AM – 6 PM, respectively). Mice were provided free access to water at all times and to a standard rodent chow (Purina #5008) *ab libitum*, except during a fasting period (8 AM – noon) in order to obtain plasma at 4 or 10 wk of age, after which they were sacrificed by decapitation. For each animal, the following tissues were collected in order: left lateral lobe of the liver, hypothalamus, right gonadal fat pad (adipose), pancreas, soleus and gastrocnemius. For the two muscle tissues, both the right and left were combined for gene expression profiling. All tissues, except the pancreas, were flash frozen in liquid nitrogen. All procedures were approved by University of Wisconsin Animal Care and Use Committee.

Materials

Collagenase Type XI, RIA-grade BSA, dextrose, and Ficoll Type 400-DL were purchased from Sigma. Hanks' balanced salt solution was from GIBCO. RNeasy Mini Kit was from Qiagen. DEPC-treated water was from Ambion.

Plasma measurements

Glucose was measured by the glucose oxidase method using commercially available kits (Sigma-Aldrich, St. Louis MO). For lean mice, insulin was measured by radioimmunoassay (RIA; RI-13K, Linco Research, St. Charles, MO). For *ob* mice, insulin was measured by an in-house developed ELISA using a pair of anti-insulin/proinsulin antibodies (clones D6C4 and D3E7-BT) purchased from Research Diagnostics (Concord, MA). Briefly, 96-well high-binding plates (Corning #3690) were coated (50 µl/well) overnight with 3 µg/ml of D6C4. After removal of D6C4, plates were blocked with PBS containing 4% RIA-grade BSA (Sigma) for 1 hour (100 µl/well) and then incubated for 1 hour with insulin standards (Linco Research, 0.1 to 10 ng/ml), whole plasma or whole pancreas extract (25 µl/well). D3E7-BT (25 µl/well), 1 µg/ml in PBS/1% BSA was added, gently mixed and incubated for an additional hour. After washing each well 3 times (50 mM Tris, 0.2% Tween-20, pH 8), 1 µg/ml of streptavidin-HRP (Pierce, Rockford, IL) in PBS/0.1% BSA was added (50 µl/well) and incubated for 30 minutes. Following an additional 3 washes, 16 µmol/ml of *o*-phenylenediamine (Sigma), dissolved in citrate buffer (0.1 M citrate-phosphate, 0.03% H_2O_2, pH 5), was added (50 µl/well) and incubated for 30 minutes, followed by an equal volume of 0.18 M sulphuric acid to quench the reaction. Absorbance at 492 nm was determined by a plate reader (Ultra 384 TECAN). Insulin contents in plasma were calculated by comparison to known standards. Adiponectin, PAI-1 and resistin were determined by commercial ELISA services at Linco Research (St. Charles, MO).

Islet isolation and RNA purification

Intact pancreatic islets were isolated from mice using a collagenase digestion procedure as previously described (Rabaglia et al. 2005). Islets were carefully hand-picked under a stereo microscope to remove contaminating acinar tissue, after which the islets were washed twice with phosphate buffered saline (PBS) and centrifuged at 2,500 rpm, 5

minutes, 4°C. The PBS supernatant was removed and 200 µl RLT buffer (Qiagen) was added. Islets were homogenized by hand for 1 minute with a plastic micro-pestel (USA Scientific), and stored at -80°C until RNA purification. RNA was purified using the Qiagen RNeasy Mini Kit, according to manufacturer directions. An Agilent Bioanalyzer 2100 was used to assess RNA quality for all islet samples, which typically showed a 28/18S ratio of ~1.5 or greater.

RNA isolation from non-islet tissues, gene expression profiling, data normalization, and GO term enrichment analysis.

RNA preparations (liver, muscles, adipose and hypothalamus) and all array hybridizations were performed at Rosetta Inpharmatics (Merck & Co., Seattle, WA, USA). The custom ink-jet microarrays used in this study were manufactured by Agilent Technologies (Palo Alto, CA) and consisted of 4,732 control probes and 39,558 non-control oligonucleotides extracted from mouse Unigene clusters and combined with RefSeq sequences and RIKEN full-length cDNA clones. Mouse tissues were homogenized and total RNA extracted using Trizol reagent (Invitrogen, CA, USA) according to manufacturer's protocol. Total RNA was reverse transcribed and labeled with either Cy3 or Cy5 flurochrome. For a given strain, labeled complementary RNA (cRNA) from each animal of that strain was hybridized against a pool of labeled cRNAs constructed from equal aliquots of RNA from all of the animals for that strain (over both time points). All hybridizations were performed in fluor-reversal for 48 hours in a hybridization chamber, washed, and scanned using a confocal laser scanner. Arrays were quantified on the basis of spot intensity relative to background, adjusted for experimental variation between arrays using average intensity over multiple channels, and fitted to a previously described error model to determine significance (type I error) (He et al. 2003). Gene expression measures are reported as the ratio of the mean log10 intensity (mlratio). Gene expression data that was used for the trait-gene correlations was generated using the ratio splitter pair wise ratio builder function in Resolver 6.0 (Rosetta Biosoftware, Seattle, WA) to account for the strain-specific reference pools. This pipeline allows the creation of new experiments based on comparisons of intensity channels from existing ratio hybridizations without having to prepare new hybridizations. The ratio splitting operation which generates intensity profiles includes error modeling of the channels of the ratio scan, group normalization, forward transformation of intensities, group de-trending and inverse transformations. Experiments are then rebuilt by ratioing each sample to a new baseline value, represented here as a super-pool (average of all array hybridizations in the experiment). The statistical significance of the overlap between input sets from the co-expression networks, and GO biological process gene sets was assessed using the hypergeometric distribution and a multiple test correction (Bonferroni).

In vivo islet proliferation measurement

The proliferation rate of islet cells was measured using a recently-developed heavy water (^2H_2O) labeling technique (Busch et al. 2004; Shankaran et al. 2006; Shankaran et al. 2007). Briefly, the incorporation of deuterium (^2H) from ^2H_2O into the deoxyribose moiety of deoxyribonucleotides in cells replicating their DNA is measured by gas chromatography/mass spectrometry (GC/MS). To rapidly attain stable ^2H_2O enrichment in body water, mice were given an IP injection of ^2H_2O in 0.9% NaCl at 6 wk of age. The

volume, V (mL) of the IP injection was calculated for each animal according the formula for lean and *ob* animals respectively, V_{lean} = 0.03 * body weight (gm) and V_{ob} = 0.015 * body weight (gm). On the same day of the IP injection, mice were placed on drinking water containing 8% ^2H_2O for a period of two weeks *ad libitum*. Mice were sacrificed at 8 weeks of age, at which time plasma was collected and islets and adipose tissue isolated as described above. This procedure yielded average ^2H_2O enrichment in plasma of 5.7% ± 0.5% in all mice except BTBR *ob* where 7.8% ± 0.5% enrichment was achieved, owing to polyuria, in turn due to hyperglycemia, in these diabetic mice. Hind legs were collected in order to determine the ^2H_2O enrichment in the DNA of bone marrow, a cellular population considered to have completely, or nearly completely, turned over during the 2 week labeling period (Neese et al. 2002). DNA was extracted from islets, whole adipose tissue or bone marrow using Qiagen DNEasy tissue kits (Qiagen Inc, Valencia, CA) and hydrolyzed to deoxyribonucleosides. The deoxyribose moiety of purine deoxyribonucleosides was then converted to the pentafluorobenzyl triacetate derivative by reaction with excess pentafluorobenzyl hydroxylamine under acidic conditions, followed by acetylation with acetic anhydride. GC/MS analysis was performed in negative chemical ionization mode using an Agilent (Palo Alto, CA) model 5973 mass spectrometer and a 6890 gas chromatograph fitted with a db-225 column. Selected ion monitoring was performed on ions with mass-to-charge ratios (*m/z*) 435 and 436, Incorporation of ^2H into purine deoxyribose was quantified as the molar excess fraction M_1 (EM_1), correcting for injected amount of material as described (Neese et al. 2002).The fraction of newly replicated cells in islet or adipose was calculated as the ratio of the ^2H-enrichment for each tissue to that observed in bone marrow.

In vivo adipose proliferation measurement

Measurement of adipose cell proliferation was performed as described above for islets. The whole epididmal adipose tissue was removed and frozen. Genomic DNA was isolated and ^2H enrichment was determined by GC/MS as above.

Identification of differentially expressed (DE) genes

To classify genes into differential expression patterns, we used an empirical Bayes hierarchical modeling approach called EBarrays (Kendziorski et al. 2006; Newton et al. 2001; Yuan and Kendziorski 2006), which is implemented in R, a publicly-available statistical analysis environment (R Development Core Team. 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria) and available through Bioconductor (www.bioconductor.org)**.** EBarrays describes the probability distribution of a set of expression measurements**.** It accounts generally for differences among genes in their true underlying expression levels, measurement fluctuations, and distinct expression patterns for a given gene among cell types or conditions. An expression pattern is an arrangement of a gene's true underlying intensities (μ) in each condition. The number of patterns possible depends on the number of conditions from which the expression measurements were obtained. For example, when measurements are taken from two conditions, two patterns of expression are possible: equivalent expression ($\mu_1 = \mu_2$) and differential expression ($\mu_1 \neq \mu_2$). Given the 4 conditions within each strain (4wk or 10wk; lean or obese), 15 expression patterns are possible (**Table S1**). Since we do not know a priori which genes are in which patterns,

the marginal distribution of the data is a mixture over the possible patterns with model parameters determined by the full set of array data. In this way, the approach utilizes information across a set of arrays to optimize model fit and is thus more efficient than a number of methods that make gene inferences one gene at a time**.** Posterior probabilities for each of the 15 patterns are calculated for every transcript and used for transcript classification. For each tissue, a transcript is assigned to the expression pattern with maximum posterior probability (MPP). Differentially expressed (DE) transcripts are defined as those with MPP > 0.7 (MPP > 0.5 for hypothalamus) in at least one mouse strain. For a given threshold, FDR is estimated by averaging the posterior probabilities of equivalent expression for each transcript on the list (Newton 2006). Primary expression data for all arrays used in this study have been uploaded to the GEO repository (accession no. GSE10785).

Identification of co-expression modules

We used a previously-developed method to identify transcript co-expression modules (Zhang and Horvath 2005). For transcripts identified as DE by EBarrays, an adjacency matrix was constructed. Each entry in the matrix is the absolute value of Pearson's correlation, adjusted so that the overall network is approximately scale-free. Connection strength between two genes (x_i and x_j) in the network was determined according to the adjacency function, $a_{ij} = |cor(x_i,x_j)|^{\beta}$, using the estimated power parameter β, resulting in a weighted network (Zhang and Horvath 2005). We note that this allows for all correlations to be used, unlike approaches that invoke arbitrary thresholds. For a discussion of the advantages of weighted *vs*. unweighted networks, see (Zhang and Horvath 2005) and references therein. The 8,000 most connected transcripts were used in the topological overlap matrix (TOM) calculation, and 1 – TOM was used as a distance matrix in the hierarchical clustering of the transcripts for module identification. When there were fewer than 8,000 DE transcripts in a particular tissue (adipose, soleus and hypothalamus), all were used for module identification. We found that clusters were robust to more stringent thresholds of 0.8 or 0.9 (MPP) for DE transcript identification (Fig. S7).

Partial correlation based networks

A Gaussian graphical modeling framework was used for gene-gene network construction (Schafer and Strimmer 2005a), (Schafer and Strimmer 2005b). Briefly, the method assumes a linear relationship among variables that can be described by a multivariate normal distribution. In this setting, the partial correlation (PaCor) matrix completely prescribes dependence relationships among variables since a non-zero PaCor between two variables indicates conditional dependence given all other variables; and a zero PaCor indicates that the variables are conditionally independent. More precisely, given $(X_1, X_2, \ldots, X_n)$, the partial correlation between X_1 and X_2 is defined as the correlation of X_{1r} and X_{2r} where X_{ir} denotes the residuals obtained after regressing X_i upon $(X_3, \ldots, X_n)$ (i=1,2). In contrast to Pearson's correlation coefficient between two variables, which can be high if those two variables are both related to a third variable, the PaCor quantifies the *direct* correlation between two variables since effects from all other variables are *adjusted* for, or more specifically regressed away. Significant PaCor's were identified as previously described (Schafer and Strimmer 2005b), with FDR controlled at 0.005. For

each mouse strain, 1,000 simulations of multi-variate normal data were generated for 20 mice and 107 nodes, using the strain-specific empirical covariance matrix, to verify that FDR was well-controlled and to evaluate power and the positive predictive value (percent of correctly detected edges) for each network.

Acknowledgements

This work was supported by the National Institute of Diabetes and Digestive Kidney Diseases grants 58037 and 66369 (ADA), the National Institute of General Medical Sciences grants 76274 (CK), 69430 and PA-02-110 (BSY), Brazil CNPq (ECN), a Hatch Grant from the College of Natural Resources, University of California at Berkeley and grant Bio 04-10445 from the UC Discovery-BioStar program (MKH) and from in-kind support from Rosetta Inpharmatics and KineMed, Inc. We thank Karl Broman, Roger Davis, Wes Pike and Victoria Browning for their critical comments on earlier versions of the manuscript, Yuerong Zhu for construction of the hyperlinked figures and database management and personnel at the Gene Expression Laboratory at Rosetta for carrying out all of the gene expression profiling.

Figure legends

Figure 1

10 week old BTBR *ob* mice are severely diabetic. (A) Schematic representation of experimental model depicting a gene expression network connecting key tissues in a mouse when examined over 3 primary axes: obesity, strain and age. Clinical phenotypes are shown for 5 to 7 animals for each of the 8 groups of mice used for study. Plasma glucose (**B**), insulin (**C**), total number of islets harvested per pancreas (**D**), adiponectin (**E**), PAI-1 (**F**) and resistin (**G**) are plotted. Open (lean) and closed (*ob/ob*) circles represent individual mice. Horizontal bars show mean values for each group (± S.D.).

Figure 2

Co-expression modules can be deconstructed to show strain-dependent changes in transcript expression patterns. The strain-specific expression pattern for each co-expression module is illustrated for all six tissues profiled. The color of a particular module within one tissue is not related to that same color for a module of another tissue, but is preserved across strains. The vertical size of the lines used to illustrate the module transcripts is proportional to the strain-specific posterior probability determination illustrated in **Fig S1**. A decrease in the size of the symbols is evident in the hypothalamus compared to the other tissues, reflecting the decreased posterior probability cut-off (0.5) that was used for DE transcript identification in hypothalamus. For each strain and all tissues, every transcript has a unique expression pattern. Filled arrow heads highlight the cell cycle regulatory modules in islet and adipose tissue. Strain-dependent differences in expression pattern are evident when the pattern distribution for a particular color-coded module is shifted in the two strains. For example, the cell cycle regulatory gene set in islets largely shifts from pattern 15 in B6 to pattern 7 in BTBR (see arrow heads). This figure is hyperlinked to our micro array gene expression data base at http://diabetes.wisc.edu/kelleretal2008/fig2.php.

Figure 3

Co-expression modules enriched with cell cycle regulation accurately predict diabetes and obesity. Expression heat maps (**A**) and the 1st principal component (PC1) on $\log_{10}$ scale (**B**) of the cell cycle regulatory modules in islets (217 transcripts) and adipose (96 transcripts) are shown. For the heat maps, red shows increased expression, green shows decreased expression and black is neutral. Bar plots in **B** show the PC1 for individual mice and correspond to an expressed decrease for negative values and increased expression for positive values. The percentage of new cells, derived from an *in vivo* measure of ^{2}H incorporation into newly synthesized DNA, is shown for islets and adipose tissue (**C**). Where significant obesity-dependent differences were observed, p-values are shown. Arrows are used to show influence of obesity. *NS*, not significant.

Figure 4

A gene-gene network model is distinct between B6 and BTBR mice. A gene-gene network was constructed based on the partial correlation (PaCor) between the strain-specific PC1 calculated between all modules identified in the 6 tissues profiled. Modules are illustrated as colored bricks along the inside and outside of the network wheels and preserve the color scheme illustrated in **Fig 2**. Inter-tissue edges within the network are

shown as lines connecting inside modules; intra-tissue edges are depicted as arcs connecting the outside modules. The cell cycle regulatory module in islet and those modules that form a direct connection to the cell cycle islet module are highlighted with open arrow heads. Network hot spots are indicated with asterisks. Line thickness is proportional to the magnitude of the PaCor, which ranged from 0.487 to 0.093 in B6 and from 0.303 to 0.086 in BTBR, for maximum and minimum respectively. Positive predictive value for edge accuracy, obtained from simulations (see methods), were on average 78% in B6 and 77% in BTBR. Red, negative PaCor; Green, positive PaCor. Significance is set to control FDR at 0.5%. This figure is hyperlinked to our micro array gene expression data base at http://diabetes.wisc.edu/kelleretal2008/fig4.php.

Supplementary material

Figure S1

Each tissue yielded a unique collection of differentially expressed genes. The maximum posterior probability (MPP) for B6 is plotted against MPP for BTBR for each tissue. Only transcripts that are differentially expressed (DE) in at least one strain are shown. Vertical and horizontal lines illustrate the MPP threshold that was used to classify a transcript as DE. Each dot represents a transcript, the total of which is the sum of the two numbers listed under the tissue type. Transcripts plotted in black have their MPP for the same expression pattern, whereas those plotted in red do not. For each tissue except hypothalamus, DE transcripts are defined as those with an MPP exceeding 0.7 in all tissues except hypothalamus, yielding a false discovery rate (FDR) of islet 0.2, adipose 0.17, liver 0.19, soleus 0.2 and gastrocnemius 0.2. The threshold was chosen to balance FDR with the total number of genes identified. For hypothalamus, a slightly lower MPP threshold of 0.5 was used, which yielded an FDR of 0.42 and 3,704 genes identified; this was done as the threshold of 0.7 yielded too few genes (357) for meaningful clustering and module construction.

Figure S2

Non-supervised hierarchical clustering of the mice results in obesity *vs.* age-driven gene expression tissue groups. For each tissue, gene expression was determined in 8 groups of 5 individuals per group. The mice were hierarchically clustered for each tissue based solely on the transcripts determined to be differentially expressed as illustrated in **Fig. S1**. Mice in Islet, adipose and liver segregate first by obesity and then by age, whereas in the muscles and hypothalamus, the reverse was observed. Obese animals are indicated by shading.

Figure S3

Transcript modules with highly correlated expression profiles are present in all six tissues. Dendrogram trees were produced for each tissue by average linkage hierarchical clustering of the DE transcript sets illustrated in **Fig. S1**. Co-expression gene modules are indicated by the vertical branches and were assigned colors as shown. Modules showing significant enrichment with gene ontology are annotated. **Figure S7** indicates that the

modules constructed are largely robust to the thresholds chosen. Height represents distance between clusters, with distance measured by topological overlap dissimilarity.

Figure S4

Co-expression modules can be highly correlated with plasma glucose or insulin. The absolute value of the Pearson's correlation coefficient was calculated between all transcripts and fasting plasma glucose or insulin. The mean absolute correlation for each module is reported as the module significance (MS) (Ghazalpour et al. 2006) and is defined as the mean of the glucose (top panel) or insulin (bottom panel) absolute correlation values for all transcripts contained within a given module. Arrow heads indicate modules with greatest glucose and insulin MS values. Significance is shown for p-values at 0.05, 0.01 and 0.001. Brightly colored modules are those that achieve a p-value ≤ 0.05.

Figure S5

Strain-specific 1st principal component can accurate depict all the transcripts contained within a co-expression module. The heat map in **A** illustrates the 452 transcripts that are contained within the brown module in adipose, which was significantly enriched with gene ontology associated with inflammation (see **Fig. S4**). Red indicates increased expression; green, decreased expression; and black is neutral. A strain-specific PC1 (black trace) is shown superimposed on the individual expression profiles for all transcripts within the module plotted as the log ratio of gene expression (**B**). The majority of transcripts show an up-regulation with obesity (orange traces), while a small number show a down-regulation with obesity (blue traces). A Pearson's correlation was calculated between the individual transcripts and the PC1 for each strain. A histogram of the resulting correlation values yielded two discrete populations, at -1.0 and 1.0 (**C**).

Figure S6

Plasma glucose and insulin form significant partial correlations with co-expression modules. A gene-to-clinical trait network was constructed based on the partial correlation (PaCor) between the strain-specific PC1 calculated between all modules and either glucose or insulin. Modules are illustrated as in **Fig. 4**. Line thickness is proportional to the magnitude of the PaCor. Red, negative PaCor; Green, positive PaCor.

Figure S7

Co-expression modules are well preserved across a wide range of posterior probability thresholds. Dendrogram trees are illustrated for DE transcripts identified in islets using moderate (MPP > 0.7), intermediate (MPP > 0.8) or high (MPP > 0.9) posterior probability thresholds. (**A**) The number of transcripts, value of β necessary to achieve a "scale-free" network (see methods) and resulting modules (color-coded for each tree) are shown as a function of MPP. The dendrogram calculated from the moderate threshold is redrawn from that shown for islets in **Fig. S3**. The color designation for transcripts contained within specific modules for the high threshold dendrogram (**A**) is preserved to show the location of those module-specific transcripts in the moderate and intermediate dendrograms illustrated in **B**. Arrow heads highlight the

cell cycle regulatory module identified for each dendrogram. A similar degree of module preservation as a function of MPP threshold was observed for the other tissues profiled (data not shown).

Table S1
There are 15 possible patterns of gene expression in 4 groups of mice.
As detailed in methods, EBarrays was used to classify transcripts into one of 15 patterns. The patterns represent the possible ways in which the average expression levels across groups can equal or differ from one another. Numbers 1 – 4 are used to indicate if one group is the same or different from another. For example, pattern #1 (null hypothesis, none different) occurs when the average expression levels in each group are equal. Pattern #15 (age and obesity, all different) is a pattern consisting of transcripts with different average expression levels across the four groups. Pattern #6 (obesity only) describes the situation where the lean mice are different from the obese mice; but neither the lean mice nor the obese mice differ as a function of age. Shaded rows show the expression patterns exhibited by the vast majority (>96%) of differentially expressed (DE) transcripts.

Table S2
Color table used for reference to co-expression gene modules.

Table S3
Co-expression modules can be used to deconstruct strain-specific changes in expression patterns. Co-expression modules are identified by color name for each of the 6 tissues profiled. There were 19 modules identified in all tissues, except hypothalamus, where there were 10. The number of transcripts contained in each module is shown for both strains and as a function of expression pattern. Only those patterns where >96% of the DE transcripts were confined are shown. The table is a numerical representation of what is graphically illustrated in Fig. 2 and is hyperlinked to our data base at http://diabetes.wisc.edu/kelleretal2008/tabs3.php.

Table S4
Co-expression gene modules in 6 mouse tissues contain transcripts highly enriched with specific gene ontology (GO) categories. Bonferroni-corrected (for multiple testing penalty) and non-corrected p-values are shown. GO terms with Bonferroni p-values ≤ 0.01 are included.

References

Brennand, K., D. Huangfu, and D. Melton. 2007. All beta Cells Contribute Equally to Islet Growth and Maintenance. *PLoS Biol* **5:** e163.

Burcelin, R., C. Mrejen, J.F. Decaux, S.H. De Mouzon, J. Girard, and M.J. Charron. 1998. In vivo and in vitro regulation of hepatic glucagon receptor mRNA concentration by glucose metabolism. *J Biol Chem* **273:** 8088-8093.

Busch, R., D. Cesar, D. Higuera-Alhino, T. Gee, M.K. Hellerstein, and J.M. McCune. 2004. Isolation of peripheral blood CD4(+) T cells using RosetteSep and MACS for studies of DNA turnover by deuterium labeling. *J Immunol Methods* **286:** 97-109.

Carlson, M.R., B. Zhang, Z. Fang, P.S. Mischel, S. Horvath, and S.F. Nelson. 2006. Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC Genomics* **7:** 40.

Cozar-Castellano, I., N. Fiaschi-Taesch, T.A. Bigatel, K.K. Takane, A. Garcia-Ocana, R. Vasavada, and A.F. Stewart. 2006. Molecular control of cell cycle progression in the pancreatic beta-cell. *Endocr Rev* **27:** 356-370.

Devedjian, J.C., M. George, A. Casellas, A. Pujol, J. Visa, M. Pelegrin, L. Gros, and F. Bosch. 2000. Transgenic mice overexpressing insulin-like growth factor-II in beta cells develop type 2 diabetes. *J Clin Invest* **105:** 731-740.

Dor, Y., J. Brown, O.I. Martinez, and D.A. Melton. 2004. Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation. *Nature* **429:** 41-46.

Flier, S.N., R.N. Kulkarni, and C.R. Kahn. 2001. Evidence for a circulating islet cell growth factor in insulin-resistant states. *Proc Natl Acad Sci USA* **98:** 7475-7480.

Gargalovic, P.S., M. Imura, B. Zhang, N.M. Gharavi, M.J. Clark, J. Pagnon, W.P. Yang, A. He, A. Truong, S. Patel et al. 2006. Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. *Proc Natl Acad Sci USA* **103:** 12741-12746.

Georgia, S. and A. Bhushan. 2004. Beta cell replication is the primary mechanism for maintaining postnatal beta cell mass. *J Clin Invest* **114:** 963-968.

Ghazalpour, A., S. Doss, B. Zhang, S. Wang, C. Plaisier, R. Castellanos, A. Brozell, E.E. Schadt, T.A. Drake, A.J. Lusis et al. 2006. Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet* **2:** e130.

Haluzik, M. and D. Haluzikova. 2006. The role of resistin in obesity-induced insulin resistance. *Curr Opin Investig Drugs* **7:** 306-311.

He, Y.D., H. Dai, E.E. Schadt, G. Cavet, S.W. Edwards, S.B. Stepaniants, S. Duenwald, R. Kleinhanz, A.R. Jones, D.D. Shoemaker et al. 2003. Microarray standard data set and figures of merit for comparing data processing methods and experiment designs. *Bioinformatics* **19:** 956-965.

Horvath, S., B. Zhang, M. Carlson, K.V. Lu, S. Zhu, R.M. Felciano, M.F. Laurance, W. Zhao, S. Qi, Z. Chen et al. 2006. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc Natl Acad Sci USA* **103:** 17402-17407.

Kendziorski, C.M., M. Chen, M. Yuan, H. Lan, and A.D. Attie. 2006. Statistical methods for expression quantitative trait loci (eQTL) mapping. *Biometrics* **62:** 19-27.

Lee, N.K., H. Sowa, E. Hinoi, M. Ferron, J.D. Ahn, C. Confavreux, R. Dacquin, P.J. Mee, M.D. McKee, D.Y. Jung et al. 2007. Endocrine regulation of energy metabolism by the skeleton. *Cell* **130:** 456-469.

Liang, Y., M.C. Osborne, B.P. Monia, S. Bhanot, W.A. Gaarde, C. Reed, P. She, T.L. Jetton, and K.T. Demarest. 2004. Reduction in glucagon receptor expression by an antisense oligonucleotide ameliorates diabetic syndrome in db/db mice. *Diabetes* **53:** 410-417.

Milo-Landesman, D. and S. Efrat. 2002. Growth factor-dependent proliferation of the pancreatic beta-cell line betaTC-tet: an assay for beta-cell mitogenic factors. *Int J Exp Diabetes Res* **3:** 69-74.

Mori, T., Y. Anazawa, K. Matsui, S. Fukuda, Y. Nakamura, and H. Arakawa. 2002. Cyclin K as a direct transcriptional target of the p53 tumor suppressor. *Neoplasia* **4:** 268-274.

Neese, R.A., L.M. Misell, S. Turner, A. Chu, J. Kim, D. Cesar, R. Hoh, F. Antelo, A. Strawford, J.M. McCune et al. 2002. Measurement in vivo of proliferation rates of slow turnover cells by 2H2O labeling of the deoxyribose moiety of DNA. *Proc Natl Acad Sci USA* **99:** 15345-15350.

Newton, M., Wang, P., Kendziorski, C. 2006. Hierarchical mixture models for expression profiles. In *Bayesian inference for gene expression and proteomics* (eds. K.-A. Do, P. Mueller and M. Vanucci), pp. 40-52. Cambridge University Press, New York, NY.

Newton, M.A., C.M. Kendziorski, C.S. Richmond, F.R. Blattner, and K.W. Tsui. 2001. On differential variability of expression ratios: improving statistical inference about gene expression changes from microarray data. *J Comput Biol* **8:** 37-52.

Okada, T., C.W. Liew, J. Hu, C. Hinault, M.D. Michael, J. Krtzfeldt, C. Yin, M. Holzenberger, M. Stoffel, and R.N. Kulkarni. 2007. Insulin receptors in beta-cells are critical for islet compensatory growth response to insulin resistance. *Proc Natl Acad Sci USA* **104:** 8977-8982.

Okamoto, H., M.L. Hribal, H.V. Lin, W.R. Bennett, A. Ward, and D. Accili. 2006. Role of the forkhead protein FoxO1 in beta cell compensation to insulin resistance. *J Clin Invest* **116:** 775-782.

Petrik, J., J.M. Pell, E. Arany, T.J. McDonald, W.L. Dean, W. Reik, and D.J. Hill. 1999. Overexpression of insulin-like growth factor-II in transgenic mice is associated with pancreatic islet cell hyperplasia. *Endocrinology* **140:** 2353-2363.

Pierucci, D., S. Cicconi, P. Bonini, F. Ferrelli, D. Pastore, C. Matteucci, L. Marselli, P. Marchetti, F. Ris, P. Halban et al. 2001. NGF-withdrawal induces apoptosis in pancreatic beta cells in vitro. *Diabetologia* **44:** 1281-1295.

Rabaglia, M.E., M.P. Gray-Keller, B.L. Frey, M.R. Shortreed, L.M. Smith, and A.D. Attie. 2005. Alpha-Ketoisocaproate-induced hypersecretion of insulin by islets from diabetes-susceptible mice. *Am J Physiol Endocrinol Metab* **289:** E218-224.

Rane, S.G., P. Dubus, R.V. Mettus, E.J. Galbreath, G. Boden, E.P. Reddy, and M. Barbacid. 1999. Loss of Cdk4 expression causes insulin-deficient diabetes and Cdk4 activation results in beta-islet cell hyperplasia. *Nat Genet* **22:** 44-52.

Roduit, R., J. Morin, F. Masse, L. Segall, E. Roche, C.B. Newgard, F. Assimacopoulos-Jeannet, and M. Prentki. 2000. Glucose down-regulates the expression of the peroxisome proliferator-activated receptor-alpha gene in the pancreatic beta -cell. *J Biol Chem* **275:** 35799-35806.

Rorsman, P., P.O. Berggren, K. Bokvist, H. Ericson, H. Mohler, C.G. Ostenson, and P.A. Smith. 1989. Glucose-inhibition of glucagon secretion involves activation of GABAA-receptor chloride channels. *Nature* **341:** 233-236.

Saxena, R., B.F. Voight, V. Lyssenko, N.P. Burtt, P.I. de Bakker, H. Chen, J.J. Roix, S. Kathiresan, J.N. Hirschhorn, M.J. Daly et al. 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316:** 1331-1336.

Schafer, J. and K. Strimmer. 2005a. An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* **21:** 754-764.

Schafer, J. and K. Strimmer. 2005b. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat Appl Genet Mol Biol* **4:** Article32.

Scott, L.J., K.L. Mohlke, L.L. Bonnycastle, C.J. Willer, Y. Li, W.L. Duren, M.R. Erdos, H.M. Stringham, P.S. Chines, A.U. Jackson et al. 2007. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316:** 1341-1345.

Shankaran, M., C. King, J. Lee, R. Busch, M. Wolff, and M.K. Hellerstein. 2006. Discovery of novel hippocampal neurogenic agents by using an in vivo stable isotope labeling technique. *J Pharmacol Exp Ther* **319:** 1172-1181.

Shankaran, M., M.E. Marino, R. Busch, C. Keim, C. King, J. Lee, S. Killion, M. Awada, and M.K. Hellerstein. 2007. Measurement of brain microglial proliferation rates in vivo in response to neuroinflammatory stimuli: Application to drug discovery. *J Neurosci Res* **85:** 2374-2384.

Tsutsui, T., B. Hesabi, D.S. Moons, P.P. Pandolfi, K.S. Hansel, A. Koff, and H. Kiyokawa. 1999. Targeted disruption of CDK4 delays cell cycle entry with enhanced p27(Kip1) activity. *Mol Cell Biol* **19:** 7011-7019.

Vasavada, R.C., I. Cozar-Castellano, D. Sipula, and A.F. Stewart. 2007. Tissue-specific deletion of the retinoblastoma protein in the pancreatic beta-cell has limited effects on beta-cell replication, mass, and function. *Diabetes* **56:** 57-64.

Vieira, E., A. Salehi, and E. Gylfe. 2007. Glucose inhibits glucagon secretion by a direct effect on mouse pancreatic alpha cells. *Diabetologia* **50:** 370-379.

Weisberg, S.P., D. McCann, M. Desai, M. Rosenbaum, R.L. Leibel, and A.W. Ferrante, Jr. 2003. Obesity is associated with macrophage accumulation in adipose tissue. *J Clin Invest* **112:** 1796-1808.

Wendt, A., B. Birnir, K. Buschard, J. Gromada, A. Salehi, S. Sewing, P. Rorsman, and M. Braun. 2004. Glucose inhibition of glucagon secretion from rat alpha-cells is mediated by GABA released from neighboring beta-cells. *Diabetes* **53:** 1038-1045.

Wierup, N., M. Bjorkqvist, M.J. Kuhar, H. Mulder, and F. Sundler. 2006. CART regulates islet hormone secretion and is expressed in the beta-cells of type 2 diabetic rats. *Diabetes* **55:** 305-311.

Wierup, N., W.G. Richards, A.W. Bannon, M.J. Kuhar, B. Ahren, and F. Sundler. 2005. CART knock out mice have impaired insulin secretion and glucose intolerance, altered beta cell morphology and increased body weight. *Regul Pept* **129:** 203-211.

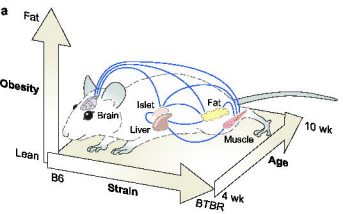
Xu, H., G.T. Barnes, Q. Yang, G. Tan, D. Yang, C.J. Chou, J. Sole, A. Nichols, J.S. Ross, L.A. Tartaglia et al. 2003. Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J Clin Invest* **112:** 1821-1830.

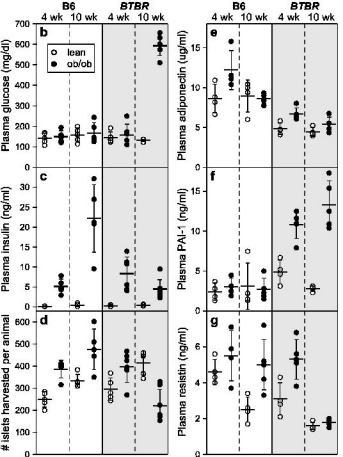
Yuan, M. and C. Kendziorski. 2006. A unified approach for simultaneous gene clustering and differential expression identification. *Biometrics* **62:** 1089-1098.

Zeggini, E. M.N. Weedon C.M. Lindgren T.M. Frayling K.S. Elliott H. Lango N.J. Timpson J.R. Perry N.W. Rayner R.M. Freathy et al. 2007. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316:** 1336-1341.

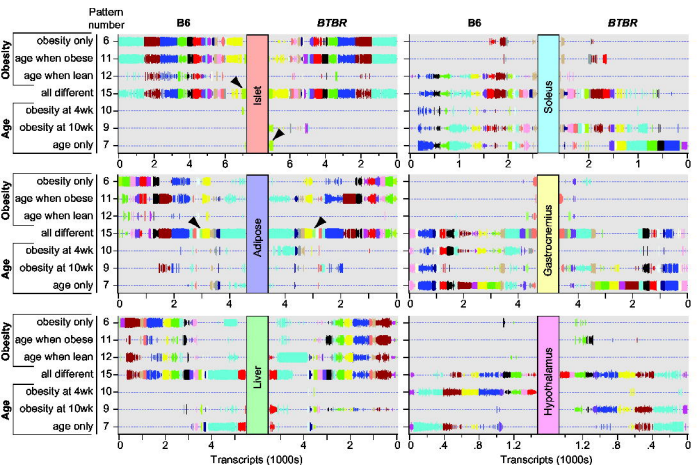
Zhang, B. and S. Horvath. 2005. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4:** Article17.

074914
Keller et al., 2007
Figure 1

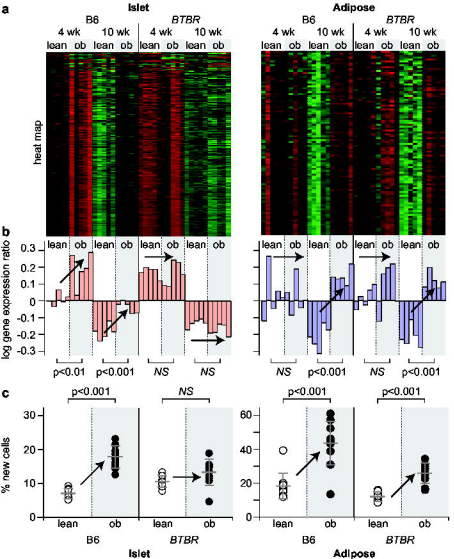




074914
Keller et al., 2007
Figure 2



074914
Keller et al., 2007
Figure 3



074914
Keller et al., 2007
Figure 4