A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort

Danielle M. Greenawalt, Radu Dobrin, Eugene Chudin, et al.

Genome Res. published online May 20, 2011
Access the most recent version at doi:10.1101/gr.112821.110

A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort

Danielle M. Greenawalt,[1,5,9] Radu Dobrin,[2,5,6,9] Eugene Chudin,[3] Ida J. Hatoum,[4] Christine Suver,[3,7] John Beaulaurier,[3] Bin Zhang,[3,7] Victor Castro,[4] Jun Zhu,[3,7] Solveig K. Sieberts,[3,7] Susanna Wang,[3] Cliona Molony,[1] Steven B. Heymsfield,[2] Daniel M. Kemp,[2] Marc L. Reitman,[2] Pek Yee Lum,[3] Eric E. Schadt,[3,5,8,9] and Lee M. Kaplan[4,5,9]

[1]*Merck Research Laboratories, Boston, Massachusetts 02115, USA;* [2]*Merck Research Laboratories, Rahway, New Jersey 07065, USA;* [3]*Genetics, Rosetta Inpharmatics LLC, a wholly owned subsidiary of Merck & Co., Inc., Seattle, Washington 98109, USA;* [4]*Massachusetts General Hospital, Boston, Massachusetts 02114, USA*

To map the genetics of gene expression in metabolically relevant tissues and investigate the diversity of expression SNPs (eSNPs) in multiple tissues from the same individual, we collected four tissues from approximately 1000 patients undergoing Roux-en-Y gastric bypass (RYGB) and clinical traits associated with their weight loss and co-morbidities. We then performed high-throughput genotyping and gene expression profiling and carried out a genome-wide association analyses for more than 100,000 gene expression traits representing four metabolically relevant tissues: liver, omental adipose, subcutaneous adipose, and stomach. We successfully identified 24,531 eSNPs corresponding to about 10,000 distinct genes. This represents the greatest number of eSNPs identified to our knowledge by any study to date and the first study to identify eSNPs from stomach tissue. We then demonstrate how these eSNPs provide a high-quality disease map for each tissue in morbidly obese patients to not only inform genetic associations identified in this cohort, but in previously published genome-wide association studies as well. These data can aid in elucidating the key networks associated with morbid obesity, response to RYGB, and disease as a whole.

[Supplemental material is available for this article. The expression data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo) under super series accession no. GSE24335. Genotyping data will be made available at dbGaP (http://www.ncbi.nlm.nih.gov/gap) and from the Massachusetts General Hospital at http://www.samscore.org.]

Obesity has been identified by the Centers for Disease Control and Prevention (CDC) as a disease of epidemic proportions, with >30% of Americans classified as obese (National Health and Nutrition Examination Survey 2008). Although there are rare monogenic causes for obesity, such as mutations in the leptin signaling pathway (for review, see Farooqi and O'Rahilly 2008), on the whole, obesity is a complex disease with many contributing genetic and environmental factors (Willer et al. 2008b). For subjects with extreme obesity caused by monogenic or complex factors (BMI $\geq$ 40 kg/m^2), Roux-en-Y gastric bypass (RYGB) is a remarkably effective surgical intervention that results not only in weight loss but also in the correction of diabetes and other co-morbidities (Buchwald et al. 2004; Aslan et al. 2010). RYGB entails anastomosis of the jejunum to a small pouch of the stomach near the esophago-gastric junction, causing food to bypass a large portion of the stomach and the duodenum. Other types of bariatric surgery, such as vertical banding, that do not include bypass of the stomach and duodenum, are not as effective in reducing body weight or reversing diabetes (Tice et al. 2008). Given the extreme nature of the RYGB cohort and the dramatic effects of this treatment on obesity and associated co-morbidities, genes comprising the molecular networks identified in adipose, liver, and stomach tissue may be expected to associate with obesity phenotypes and treatment response, given that previous studies in mice and human have identified liver and adipose tissue networks enriched for genes that causally associate with metabolic disease traits (Chen et al. 2008; Emilsson et al. 2008). By identifying SNPs that associate with the gene expression that comprise these networks, we can attempt to establish a connection between genetics, gene expression, and disease in human and increase the power to identify genes associated with human disease or response to treatment. Furthermore, understanding of how genetic differences in the RYGB cohort affect molecular responses to the surgery will aid in identifying a path to ordering the molecular processes that underlie the physiological changes occurring during RYGB.

In this study, we collected four tissues (liver, omental adipose [OA], subcutaneous adipose [SA], and stomach) from 1008 patients at the time of RYGB (referred to here as the RYGB cohort) and scored a number of clinical traits relating to their co-morbidities. We then proceeded to identify significant associations between the SNP genotypes and gene expression traits in the four profiled tissues, producing a high-quality disease map composed of 24,531

[5]These authors contributed equally to this work.
Present addresses: [6]Johnson & Johnson, Radnor, PA 19087, USA; [7]Sage Bionetworks, Seattle, WA 98109, USA; [8]Pacific Biosciences, Menlo Park, CA 94025, USA.
[9]Corresponding authors.
E-mail danielle_greenawalt@merck.com.
E-mail RDobrin@ITS.JNJ.COM.
E-mail eric.schadt@gmail.com.
E-mail lmkaplan@partners.org.
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.112821.110.

expression-associated SNPs (eSNPs) corresponding to 9931 distinct genes. This represents the greatest number of eSNPs to our knowledge identified by any study to date and the first study to identify eSNPs from stomach tissue. Most importantly, we were able to identify unique sets of eSNPs in each tissue as well as a set of eSNPs that were common to multiple tissues. The large proportion of eSNPs that were replicated over different tissue types collected from the RYGB cohort indicates that DNA polymorphisms often affect gene expression in a tissue-independent manner. With the clinical data available, we also performed a genome-wide association study (GWAS) for extreme obesity, using the gene expression data to facilitate the identification of genes that are affected by the SNP genotypes of interest.

In addition to finding sets of eSNPs that differentiate each tissue, the eSNPs identified in this study have an additional utility to functionally inform previously published and future GWAS. The eSNP data generated from this study of four metabolically relevant tissues will provide a path forward to pinpointing genes and associated gene networks of interest responding to genetic perturbations associated with disease, which, in turn, may lead to a better understanding of disease etiology and ultimately the design of therapeutics for metabolic disorders.

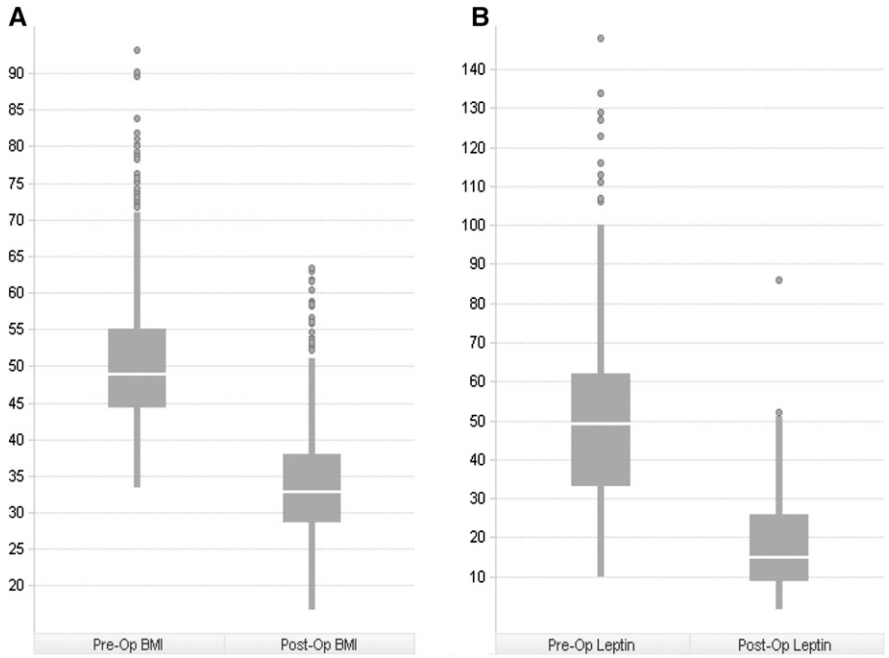

Figure 1. Effect of RYGB on BMI and Leptin. (*A*) Change in BMI (kg/m^2) from pre-surgery to post-surgery of the RYGB cohort. (*B*) Change in leptin (ng/mL) levels from pre-surgery to post-surgery of the RYGB cohort. Pre-op readings are taken as an average of all records per patient for a trait up to 1 yr prior to surgery. Post-op readings are an average of post-op records recorded between 6 mo and 36 mo post-surgery.

Results

Gastric bypass in morbidly obese patients greatly affects clinical endpoints

Tissue was collected from 1008 patients at the time of RYGB. Clinical traits were collected from 1 yr pre-operation to 7 yr post-operation to track the response of body weight to surgery. The clinical characteristics of the cohort at time of surgery are described in Table 1. The RYGB procedure had a large effect on body weight (Fig. 1).

Table 1. Baseline clinical characteristics of the RYGB cohort

Trait	Mean	StdDev	Max	Min	Percent
Age (yr)	44.7	11.5	75.0	18.0	—
Pre-op WT (kg)	141.7	30.7	280.5	85.0	—
BMI	50.5	8.8	93.2	33.3	—
Insulin (uIU/mL)	23.0	19.5	210.0	2.0	—
Glucose (mg/dL)	119.9	45.6	345.0	61.5	—
HbA1c (%)	6.4	1.4	13.6	4.5	—
Leptin (ng/mL)	50.8	24.7	148.0	10.0	—
Triglycerides (mg/dL)	161.7	100.4	1192.0	45.0	—
Cholesterol (mg/dL)	189.8	37.6	420.0	73.0	—
HDL (mg/dL)	48.1	12.0	100.5	17.5	—
LDL (mg/dL)	110.2	32.3	261.0	16.0	—
Type 2 diabetic	—	—	—	—	46
Statin use	—	—	—	—	28
Insulin use	—	—	—	—	13
Metformin use	—	—	—	—	5

Individuals studied lost an average of 32.7% of their body weight 2 yr after surgery. Mean BMI dropped from 50.5 to 33.9 kg/m^2 after surgery. In 10% of the individuals in this cohort, the BMI was reduced to <25 kg/m^2, which is considered normal. Leptin levels were also greatly reduced by surgery; pre-op leptin levels ranged from 10 to 148 ng/mL, with a mean of 51 ng/mL, while post-op leptin ranged from 1.5 to 86 ng/mL, with a mean of 19 ng/mL, constituting a 63% decrease in mean leptin levels. Information on T2D status, metformin, insulin, and statin use was extracted from text medical records and used as covariates in eQTL and gene expression analysis.

Genome-wide association analysis for expression traits

RNA was isolated from each tissue with 651 liver, 701 SA, and 848 OA successfully profiled using an Agilent whole-genome gene expression microarray targeting 39,303 transcripts. RNA was extracted from 118 stomach samples and profiled on an Affymetrix whole-genome expression array. In addition, DNA was isolated from the liver samples for each patient and genotyped using a whole-genome, high-density SNP panel. We identified *cis*- and *trans*-acting expression quantitative trait loci (eQTLs) using a method based on Schadt et al. (2008). Briefly, *cis* eQTLs for gene expression traits were determined by identifying the SNP most strongly associated with each expression trait profiled on the array over all genotyped SNPs typed within 1 Mb of the structural gene of interest, similar to the method reported in other eQTL studies (Schadt et al. 2008; Dimas et al. 2009). Results are reported at a 10% false discovery rate (FDR) based on permutations of the association between SNP genotypes and expression levels for any given expression trait. The 1-Mb threshold was chosen because extending the distance out further, while bringing to light additional eSNPs, did so at the expense of raising the FDR at the given *P*-value threshold. *Trans* eQTLs were determined using a similar approach,

but instead considering all SNPs >1 Mb from the structural gene of interest. Representative eSNPs for each *trans* eQTL were then selected as the most statistically significantly associated SNP within the eQTL region. A summary of the eSNPs identified in liver, OA, SA, and stomach tissues is reported in Table 2 (eSNP data are available at http://www.samscore.org). A total of 24,513 distinct *cis* and *trans* eSNPs associated with 15,241 transcripts or 9931 distinct genes, were identified, representing >58% of the 17,052 human protein-coding genes (Pruitt et al. 2009). Compared to liver, ~25% more *cis* and *trans* eSNPs were identified in the adipose tissues. The increase in eSNPs could be a consequence of greater power in the adipose samples, due to the larger sample size (*n*) in the SA and OA tissues ($n = 701$ and 848, respectively) compared to the liver tissue ($n = 651$) or due to tissue type. Stomach tissue could not be collected for all patients. Gene expression data were available for 118 stomach samples. Despite this small number of samples, we were still able to identify 878 *cis* eSNPs.

Previous studies have shown an ~30% overlap between *cis* eSNPs identified in different tissues (Schadt et al. 2008) and ~12% percent overlap between two cell populations derived from a single donor (Dimas et al. 2009). To compare the number of eSNPs in our study present in multiple tissues collected from the same individual, we first used the comparison method employed by Dimas et al. (2009) by which we determined, if an expression trait had an eSNP in one tissue, was it at all present in either of the other two tissues? As Dimas notes, this method may overestimate the overlap because eQTLs for the same expression trait may not represent identical genetic variants. When comparing eQTLs from the RYGB cohort, we found that 72% of the *cis* eSNPs identified in the liver, 79% of those found in OA, and 80.5% from SA were also found in at least one of the other two tissues. However, because of the overestimation that can result from this method, we attempted to test directly whether an eSNP identified in one tissue could be identified in either of the other two tissues. Using this method, we found that 46% of the eSNPs identified at a genome-wide significance level in either liver or SA were replicated in the other tissue at the 0.05 significance level, while 48% of the *cis* eSNPs identified in liver or OA were replicated in the other tissue at this same significance level. For eSNPs identified in either OA or SA at the genome-wide significance level, 71% were replicated in the other tissue at the 0.05 significance level. The increased overlap identified using both overlap approaches (compared to the overlaps reported in previously published studies [Emilsson et al. 2008; Schadt et al. 2008; Dimas et al. 2009]) could potentially reflect the increased power achieved with a larger sample size, the fact that the tissues are matched to each patient in the RYGB cohort, or the fact that tissues were collected at the same time point. Adipose tissues taken from two different depots of the same individual in the RYGB cohort were found to have a much higher overlap of *cis* eQTLs (71%) than *cis* eQTLs identified in multiple cell types derived from the same individual (Dimas et al. 2009).

Table 2. Number of eSNPs identified in each tissue at 10% FDR

Tissue	Number of *cis* eSNPs	Number of *trans* eSNPs
Liver	7902	785
Omental adipose	10,730	1006
Subcutaneous adipose	10,247	1138
Stomach	878	38

eSNPs that were specific to liver were found to be enriched for the KEGG pathway *drug metabolism–cytochrome P450 genes* ($E = 1.63 \times 10^{-7}$, where *E* denotes the Bonferroni-corrected Fisher's Exact Test [FET] *P*-value). This result agrees with a recently published study indicating association of liver eSNPs to expression and enzymatic activity of a number of important drug-metabolizing enzymes (Yang et al. 2010). In addition, liver-specific eSNPs were enriched for the Gene Ontologies *lipid, fatty acid, steroid metabolism* ($E = 1.79 \times 10^{-4}$), and *inflammatory response* ($E = 3.3 \times 10^{-3}$). SA-specific eSNPs were enriched for the Gene Ontologies *developmental processes* ($E = 9 \times 10^{-4}$) and *anatomical structure development* ($E = 0.01$). While these categories are not very specific, they may be representative of genes involved in adipocyte differentiation. OA-specific eSNPs were not enriched at the adjusted 0.05 significance level for any KEGG pathway or GO category.

When comparing the liver *cis* eSNPs identified in the RYGB cohort to an independent set of liver *cis* eSNPs (Schadt et al. 2008), profiled on the same gene microarray, by directly comparing matched expression trait to SNP pairs, there was a 66% overlap between the two studies. When the significance of the association between the SNP and the expression trait was compared in the two liver data sets, 84% of the eSNPs identified in the RYGB cohort were found to have more significant *P*-values, at least partially reflecting increased power from the larger sample size in the RYGB cohort. Comparing the RYGB SA *cis* eSNPs to an independent set of SA *cis* eSNPs revealed a 57% overlap when expression traits present on the two arrays were compared (Emilsson et al. 2008).

To determine whether the increase in number of eSNPs and eSNP tissue overlap identified in this cohort reflects the study size, we randomly selected a smaller subsample of 107 individuals from the RYGB cohort in which liver, SA, and OA were successfully profiled. As expected, we confirmed that the number of eQTLs detected and the overlap between eQTL sets for each tissue increased with increasing sample size (Fig. 2). Greater than 15% more *cis* eSNPs were identified in the adipose samples than liver. When we compared the overlap of eQTLs identified in the 107-sample subset, we found that 32% (and 33%) of the *cis* eQTLs identified in either liver or SA (or OA) at a genome-wide significance level were identified in the other tissue at the 0.05 significance level, whereas this overlap between SA and OA was 49%. The number of *cis* eQTLs identified in multiple tissues decreased with a decrease in sample size. We believe that the large overlap observed between the two different adipose depots, regardless of the sample size, is a reflection of similar biology, their respective overlaps with liver *cis* eQTLs being much smaller. It has also previously been shown that eQTLs that are common among different cell types are located closer to their associated transcript, compared with eQTLs found to be specific to a single cell type (Dimas et al. 2009). This behavior was observed in our data set regardless of the sample size used for the analysis (Fig. 2C). However, we should acknowledge that in both studies, eSNPs that are farther from the transcript of interest may be more likely to be false positives, which is why they are not replicated in multiple tissues.

eSNP identification in four tissues from a morbidly obese cohort has for the first time informed a number of previously identified SNP associations, many of which are to metabolic disorders (Supplemental Table 1). rs1532085 and rs3764261 have both been associated with HDL cholesterol (Willer et al. 2008a; Aulchenko et al. 2009; Sabatti et al. 2009) and are eSNPs to *LIPC* and *CETP*, respectively, in our cohort. We found that rs1805081, previously associated with obesity (Meyre et al. 2009), is an eSNP to *NPC1*, and that rs10486607, associated with diabetes-related insulin traits (Meigs et al. 2007), is an eSNP to *CPVL*. A common mis-sense

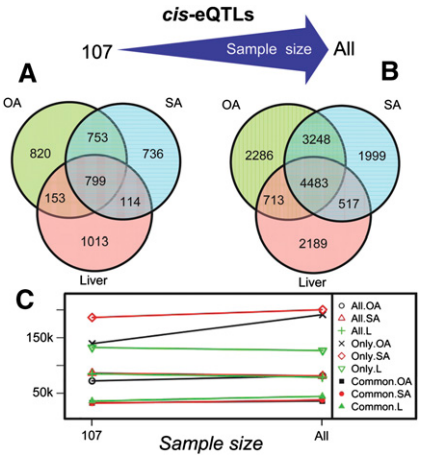

Figure 2. *Cis*-eQTLs overlaps. (*A*) Venn diagram for the *cis* eSNPs detected in three tissues using a subpopulation of 107 individuals. (*B*) Overlaps of the *cis* eSNPs identified in three tissues at maximum sample size in the RYGB cohort. Regardless of sample size, the OA and SA tissues share many more *cis* eQTLs in comparison to liver versus adipose. (*C*) Plot of mean distance of the eSNP to the gene. Symbols marked in the legend with "Only" represent the subset of the *cis* eSNPs specific to a particular tissue, "All" for all *cis* eSNPs detected in a tissue, and "Common" for the subset of *cis* eSNPs common to all three tissues. *Cis* eSNPs common to all three tissues are significantly closer to the associated genetic signal regardless of the sample size or tissue in which they were detected compared to *cis* eSNPs unique to a single tissue. The mean distance for all *cis* eQTLs detected in each tissue show little change between the small subpopulation and the full sample.

mutation, P446L, in *GCKR* has been associated with a number of metabolic disorders including T2D traits, obesity, and triglycerides (Orho-Melander et al. 2008; Vaxillaire et al. 2008; Shen et al. 2009). Our study found that rs1260326 is an eSNP to *GCKR* in the liver. Carriers of the minor allele were found to have increased expression of *GCKR* in the liver.

To further explore the utility of eSNPs identified in the RYGB cohort in enhancing interpretations of GWAS data, we examined the open access GWAS results database (OPGWAS DB) (Johnson and O'Donnell 2009) to assess whether disease-associated SNPs for a number of different disease traits are enriched for eSNPs. There are 87 different primary phenotype disease categories represented in the OPGWAS DB. To reduce artifactual enrichments that may arise in small sample sets, we required that a disease category contain more than 10 disease-associated SNPs or that the number of eSNPs corresponding to the disease-associated SNPs be more than five. This resulted in a set of 59 disease categories (from the original 87). At a 10% FDR, there were 299,242 SNPs out of the 2,529,766 SNPs represented in HapMap (v2) that were either associated with expression in the RYGB cohort or that were significantly correlated (correlation $P < 10 \times 10^{-6}$) with SNPs that were significantly associated with expression traits in the RYGB cohort. We expanded the set of SNPs considered for this enrichment beyond the set of SNPs considered above in the eSNP analyses because the HapMap (v2) set captures the majority of disease SNPs in the OPGWAS DB database.

We identified a number of disease categories with disease SNPs that were significantly enriched for eSNPs in the RYGB cohort (Table 3). For example, for the "Early onset extreme obesity" category, 42 SNPs were represented in OPGWAS DB, and 11 of these were associated with expression traits in the RYGB cohort, whereas only five would have been expected by chance (a 2.21-fold enrichment, FET, $P = 0.0082$). For the type I diabetes disease (T1D) category, 364 SNPs were represented in OPGWAS DB, and 316 of these (87%) were associated with expression, whereas only 43 were expected by chance (a greater than sevenfold enrichment, FET, $p \ll 10 \times 10^{-15}$). While a great majority of the T1D-associated SNPs are located within the MHC region on chromosome 6, the association of different HLA adipose and liver expression traits in this region is extremely useful for elucidating the genes and mechanisms perturbed by the causal variants at this locus (Schadt et al. 2008). The eSNPs in the MHC region are associated with 42 different genes, including *NOTCH4*, *DDR1*, *HLA-B*, *HLA-C*, and *HLA-DRB1*. These genes and SNPs have been associated with many other diseases as well. For example, SNPs associated with *HLA-DPA1* and *HLA-DPB1* expression have not only associated with T1D, but with chronic hepatitis B (Kamatani et al. 2009). In addition, a number of the other disease areas in the OPGWAS DB database are associated with the MHC region as well, including childhood asthma, rheumatoid arthritis, multiple sclerosis, and Crohn's disease, all of which had disease-associated SNP sets that were very enriched for eSNPs. These results agree with those recently published by Nicolae et al. (2010), which looked at the enrichment of GWAS SNPs eQTLs identified from lymphoblastoid cell lines (LCLs) in comparison to minor allele frequency-matched SNPs. Both Nicolae et al. (2010) and Nica et al. (2010) have recently reported that GWAS SNPs for immune-related disorders, such as T1D and Crohn's disease, are highly enriched for eQTLs identified from LCLs. While we do find enrichment of eQTLs from the RYGB cohort for disease categories such as *Early Onset Extreme Obesity*, immunity-associated diseases are also enriched for in this eQTL data set.

To further elucidate the biological relevance of the genetics and gene expression data from a morbidly obese cohort, we first compared the liver eSNPs identified from the RYGB cohort to a previously published liver eSNP set identified from a randomly selected cohort of 427 livers (Schadt et al. 2008). We found that the RYGB-specific eSNP genes were enriched for *lipid*, *fatty acid*, and *steroid metabolism genes* ($E = 0.01$), and *lipid biosynthetic process* ($E = 0.04$), including a number of genes that have been linked to obesity, type II diabetes (T2D), and cholesterol biosynthesis. *CPT1B* and *DGAT1*, two genes known to be involved in obesity (Smith et al. 2000; Robitaille et al. 2007), were found to have liver eSNPs specific to the RYGB cohort (liver eSNP, *P*-value = 8.91×10^{-17}, 1.57×10^{-12}, respectively). *FABP2*, *FABP3*, *PLA2G4A*, *APOC3*, *CETP*, *DHCR7*, *DBI*, and *SREBF2* were found to be RYGB-specific eSNPs, and each of these genes has been linked to T2D, metabolic syndrome, or cholesterol biosynthesis previously (Swenson 1991; Klos et al. 2006; Vock et al. 2008). The drug metabolism-cytochrome p450s were enriched ($E = 0.01$) in the SA eSNPs identified in the RYGB cohort. The differences between the eSNPs identified in the RYGB and random liver or adipose cohorts (Emilsson et al. 2008; Schadt et al. 2008) may be due to sample size, as we have explored previously, or to differences that exist in the activity of genes in these individuals due to their weight and co-morbidities. However, we recalculated eSNPs from a subset of 400 liver samples from the RYGB cohort, and the RYGB-specific eSNPs identified above could be replicated in the smaller cohort. Therefore, due to the extreme obesity represented in the RYGB cohort, the eSNPs identified in this study may be more relevant than those identified from non-obese cohorts for informing GWAS involving metabolically relevant diseases.

We then went on to explore the gene expression data through gene trait correlations between liver, OA, SA expression, and obesity traits of interest in the cohort including pre-op BMI, pre-op

Table 3. Disease-associated SNPs enriched for eSNPs

Disease	Total disease	Total loci represented	Total eSNPs	Total loci detected	eSNP fold-enrichment	Enrichment *P*-value[a]
Childhood asthma	81	8	73	1	7.62	$\ll 1 \times 10^{-16}$
Type I diabetes	364	9	316	3	7.34	$\ll 1 \times 10^{-16}$
Height	40	7	33	1	6.97	$\ll 1 \times 10^{-16}$
Rheumatoid arthritis	580	64	429	11	6.25	$\ll 1 \times 10^{-16}$
Multiple sclerosis	416	112	226	16	4.59	$\ll 1 \times 10^{-16}$
Hair, eye, and skin pigmentation	60	6	31	1	4.37	6.16×10^{-14}
Serum LDL cholesterol levels	14	4	7	1	4.23	5.20×10^{-4}
Systemic lupus erythematosus (SLE)	24	16	12	5	4.23	5.10×10^{-6}
Serum uric acid levels	45	4	20	1	3.76	4.65×10^{-8}
Neuroticism	20	18	6	4	2.54	0.024
Crohn's disease	583	151	166	16	2.41	$\ll 1 \times 10^{-16}$
Early onset extreme obesity	42	31	11	7	2.21	0.008
Pulmonary function phenotypes	68	48	17	7	2.11	0.002
Lipid level measurements	59	41	14	10	2.01	0.008
Minor histocompatibility antigenicity	53	44	12	4	1.91	0.019
Parkinson's disease	1345	334	269	88	1.69	$\ll 1 \times 10^{-16}$
Blood phenotypes	72	50	14	4	1.64	0.041
HIV-1 disease progression	659	208	124	28	1.59	1.40×10^{-7}
Bipolar disorder	215	134	38	27	1.49	0.008
Amyotrophic lateral sclerosis	1548	356	234	79	1.28	6.27×10^{-5}
Coronary artery disease	2013	265	142	51	1.20	0.014
Alzheimer's disease	1003	383	304	93	1.19	0.001
Type II diabetes mellitus	2166	372	272	95	1.14	0.012

[a]The fold enrichment is computed as the observed overlap count divided by the expected overlap count, estimated by dividing the observed number of SNPs in the RYGB cohort (299,242) by the number of SNPs in the HapMap (v2) set (2,529,766). The nominal *P*-values represent the significance of the Fisher's Exact Test statistic under the null hypothesis that the frequency of the indicated disease SNP set is the same between a reference set of all SNPs in HapMap and the set of eSNPs.

leptin, nadir BMI, and excess weight loss at 1 yr (EWL) (Supplemental Table 2). *XKR4* and *TIMP4* were found to be the top correlated genes to pre-op BMI and nadir BMI in both OA and SA. *TIMP4* was positively correlated in both tissues, while *XRK4* was negatively correlated. In both tissues the gene set correlated to pre-op BMI at FDR <5% was enriched for genes present in the *macrophage enriched metabolic network* (MEMN; $E = 7.59 \times 10^{-32}$), which has been linked to obesity previously (Chen et al. 2008; Emilsson et al. 2008). *TIMP4* is a metallopeptidase inhibitor, which is regulated in adipocytes and has been found to be up-regulated in inflammatory cardiovascular disorder and during adipogenesis in SA (Koskivirta et al. 2006; MacLaren et al. 2010). *RDH5* and *DHRS7B* were found to be positively correlated to pre-op leptin levels in both tissues. In the liver, *OAT* and *PALLD* were found to be the most highly correlated genes to pre-op BMI, again the correlated genes at FDR <5% were enriched for genes present in the MEMN ($E = 7.59 \times 10^{-32}$). *F2RL1* was the most highly correlated gene to nadir BMI, and *ANXA5* was correlated to EWL. Both gene sets identified at FDR <10% were enriched for *immunoglobulins* ($E = 2 \times 10^{-8}$, $E = 3.34 \times 10^{-5}$). The nadir BMI gene set was also enriched for *B-cell and antibody mediated immunity* ($E = 1.05 \times 10^{-8}$).

We next investigated the connectivity of the genes in each tissue profiled in the RYGB cohort through the construction of gene–gene coexpression networks (Zhang and Horvath 2005). Coexpression networks were built independently for each tissue (Supplemental Fig. 1, described further in Methods). Each module of the networks was assessed for association with clinical traits by computing the Pearson correlation coefficient (*R*) and *P*-value (*P*) between the first Principal Component (PC) of the module in the mRNA expression space and the clinical traits of interest. We then queried the eSNPs identified as specific to the RYGB cohort in the coexpression network modules to identify modules whose expression was at least partially explained by eSNPs specific to this cohort, to further allow us to focus our analysis on relevant modules. The subcutaneous adipose "salmon" and "blue" network modules (Supplemental Fig. 1C; Supplemental Table 3) were the most enriched for RYGB cohort-specific eSNP-associated genes (FET, *P*-value = 2.1×10^{-5} and 1.4×10^{-4}, respectively). We found that the "salmon" and "blue" modules were also significantly associated with a number of traits of interest in this cohort, including pre-op BMI ("salmon": $R = 0.19$, $P = 3.1 \times 10^{-9}$; "blue": $R = 0.17$, $P = 2.7 \times 10^{-4}$), pre-op white blood cell count ("salmon": $R = 0.20$, $P = 1.1 \times 10^{-8}$; "blue": $R = 0.30$, $P = 1.9 \times 10^{-12}$), post-op BMI ("salmon": $R = 0.15$, $P = 1.2 \times 10^{-6}$; "blue": $R = 0.10$, $P = 0.032$), and post-op leptin levels ("salmon": $R = 0.28$, $P = 0.012$). Genes in both the SA "salmon" and "blue" network modules significantly overlapped the MEMN that had been associated with a number of metabolic traits including obesity, diabetes, and atherosclerosis ("salmon": $E = 4.1 \times 10^{-39}$; "blue": $E = 8.1 \times 10^{-65}$) (Chen et al. 2008; Emilsson et al. 2008). Focusing only on the eSNP-associated genes in these modules, this enrichment remains significant ("salmon": $E = 2.2 \times 10^{-16}$; "blue": $E = 5.9 \times 10^{-22}$). These results replicate previous findings implicating adipose inflammation as a central process in obesity (Emilsson et al. 2008), and we have found an enrichment of genes in this module whose expression in adipose tissue is driven by eSNPs specific to an obese cohort.

Three modules in the liver (the "sienna," "cyan," and "blue" modules in Supplemental Fig. 1A; Supplemental Table 4) were found to be significantly correlated to pre- and post-op weight traits as well as leptin levels. The strongest correlations were observed between the liver "sienna" module and the pre-op BMI ($R = 0.24$, $P = 4.9 \times 10^{-11}$), post-op BMI ($R = 0.17$, $P = 3.3 \times 10^{-5}$), and post-op leptin levels ($R = 0.26$, $P = 2.7 \times 10^{-3}$). The "sienna" module contains only 26 genes, which were enriched for oxygen transport ($E = 1.04 \times 10^{-10}$) and defense response to bacterium ($E = 2.35 \times 10^{-9}$). Defensin alpha 1, 3, and 4 are the most highly connected genes in

this module and connect to a number of other genes associated with immune and inflammation response. The "blue" module had significant correlations to pre-op BMI ($R = 0.18$, $P = 2.7 \times 10^{-6}$) and post-op BMI ($R = 0.15$, $P = 1.3 \times 10^{-5}$), while the "cyan" module showed similar correlations to both pre-op BMI ($R = 0.19$, $R = 1.9 \times 10^{-5}$) and post-op BMI ($R = 0.15$, $P = 4.8 \times 10^{-4}$). Notably, both the "blue" and "cyan" modules were also highly enriched for genes involved in *immune response* ($E = 6.85 \times 10^{-50}$) and *inflammatory response* ($E = 1.18 \times 10^{-24}$), respectively. These modules were not significantly enriched for cohort-specific eSNPs.

Genome-wide association analysis

The availability of high-density SNP genotyping and clinical endpoint data from individuals in the RYGB cohort also allowed us to conduct SNP–clinical trait association analyses. We first looked for association between SNPs, BMI, and leptin in the RYGB cohort at a number of time points (for a full description, see Methods). A single significant association at 10% FDR to pre-op BMI was identified, rs1916803 ($P = 1.69 \times 10^{-7}$) (Table 4). To our knowledge, this region has not been associated with BMI previously; however, we were able to use the RYGB eSNP data to inform this association. rs1916803 was identified as an eSNP in SA to the gene *WDR69* ($P = 8.66 \times 10^{-8}$). *WDR69* is not well studied, but it has been identified as a target of *SUMO1* (Ganesan et al. 2007), which could indicate a role for the gene in cell cycle and growth control. Bayesian networks that have been derived from human female liver data identify an association between *WDR69* and *HTR2C*, the serotonin receptor 2c (Schadt et al. 2008). *HTR2C* is a target of the weight-loss drug dexfenfluramine, and polymorphisms in *HTR2C* have been linked to obesity and weight gain (Pooley et al. 2004). *WDR69* was also found to be differentially expressed in an independent analysis of pre- and post-RYGB skeletal muscle biopsies in a separate cohort (Park et al. 2006).

A single association was identified between post-op BMI in the RYGB cohort and rs17485462. rs17485462 is not an eSNP in the RYGB cohort; it is located in the gene *C12orf11* and near *ITPR2*. *ITPR2* is a ligand-sensitive Ca^{2+}-release channel that mediates extracellular Ca^{2+} influx. Both *ITPR2* and *ITPR3*, associated with extreme obesity through case-control analysis (Cotsapas et al. 2009), have been linked to calcium and nutrient signaling in the pancreas and salivary glands and are important in the ability to digest nutrients. However, *ITPR2* double knockout mice weigh less than controls even when fed a wet mash so that digestion is not impeded (Futatsugi et al. 2005). *ITPR2* subnetwork extracted from the Bayesian network derived from this human tissue cohort is enriched for genes involved in *lipid metabolism* ($E = 4.4 \times 10^{-4}$) and *fatty acid metabolism* ($E = 6.6 \times 10^{-4}$).

Pre-op leptin levels in the RYGB cohort were suggestively associated with two SNPs, rs661577 ($P = 1.64 \times 10^{-6}$, 20% FDR) and rs2343 ($P = 1.36 \times 10^{-6}$, 20% FDR), located on two different chromosomes (Table 4). rs661577 is an eSNP in OA for the gene *TPP2* ($P = 1.55 \times 10^{-6}$), tripeptidyl peptidase II, which is believed to regulate apoptosis and cell proliferation. Individuals with the minor allele were found to have a mean pre-op leptin level 27 ng/dL higher than individuals who were homozygous for the major allele. rs661557 was also found to be associated with the ratio of pre-op leptin to BMI ($P = 1.23 \times 10^{-6}$). In the mammalian hypothalamus, *TPP2* has a proposed anti-satiety role and is believed to stimulate adipogenesis, pointing to an important role for *TPP2* in obesity and diet control. *TPP2* levels have been found to be lower in adipose tissue from obese individuals, but then increased during adipogenesis of 3T3-L1 cells and in response to metabolic cues in mouse models (McKay et al. 2007). The second association to pre-op leptin identified, rs2343, is in the gene *TTC7B*. *TTC7B* has been shown to be differentially expressed in lean versus obese B6 and BTBR mice (Keller et al. 2008), and a haplotype including *TTC7B* has been associated with T2D previously (Salonen et al. 2007).

Our within-cohort association analysis did not replicate previously identified SNPs associated with BMI or leptin. However, when we looked at other traits, such as HDL, we found significant associations to multiple SNPs within or near *CETP*, cholesteryl ester transfer protein, a region that has been linked to HDL cholesterol in the literature previously (Saxena et al. 2007; Chambers et al. 2008; Kathiresan et al. 2008; Willer et al. 2008a). rs7499892 was found to be significantly associated ($P = 6.51 \times 10^{-8}$, FDR < 10%) with pre-op HDL levels in the RYGB cohort. rs7499892 is in an intron of *CETP* and is an eSNP to *CETP* in the liver ($P = 5.37 \times 10^{-5}$). Variants in *CETP* have been linked to HDL cholesterol levels in the literature previously; however, for the first time, our study links these associations to the expression of *CETP*.

Discussion

Obesity is a complex disease with many contributing genetic and environmental factors. Currently, bariatric surgery is the only effective treatment of extreme obesity. RYGB surgery had a dramatic effect on BMI and circulating leptin levels in this cohort. Genetic analysis of these patients has identified novel polymorphisms associated with extreme BMI and leptin levels, and replicated previous associations between HDL and SNPs within *CETP*. Through our genetics of gene expression analysis, we were then able to link the SNPs to the expression of specific genes of interest, including *WDR69* and *TPP2*, and for the first time linked an SNP associated with HDL to the expression of *CETP*. As may be expected given the extreme obesity of this cohort, genetic associations to BMI and leptin differ from those that have been identified through case-control studies or analysis of individuals with a random range of BMI. We have attempted to correct for biological and technical cofounders of the data set through the use of technical corrections as well as the use of T2D status and common medications as covariates in our analysis. However, this may not capture cofounders such as diet, smoking status, and other unknown characteristics of the cohort. These factors may have an effect on expression of genes and should be considered when interpreting these results. It is also unknown whether individuals in this cohort were *MC4R* carriers of monogenic forms of obesity. However, recent literature has shown that there is no difference in response to RYGB for *MC4R* carriers (Aslan et al. 2010).

Table 4. Associations to BMI and leptin identified in the RYGB cohort

Trait	dbSNP ID	*P*-value	Chr	BP position	MAF	eSNP gene	Closest gene
Pre-op BMI	rs1916803	1.69×10^{-7}	2	228,461,276	0.47	*WDR69*	
Post-op BMI	rs17485462	4.06×10^{-7}	12	26,976,961	0.13		*C12orf11; ITPR2*
Pre-op leptin	rs661577	1.64×10^{-6}	13	102,124,313	0.18	*TPP2*	
Pre-op leptin	rs2343	1.36×10^{-6}	14	90,164,252	0.40		*TTC7B*

Significant associations to weight traits identified at 20% FDR. If the SNP is an eSNP, the associated gene is reported; for non-eSNPS, the closest gene is reported. (MAF) Minor allele frequency.

High-density genotyping and gene expression analysis in four metabolically relevant tissues from individuals undergoing RYGB has allowed us to identify 24,531 unique eSNPs associated with about 10,000 known genes. The number of eSNPs identified from this cohort is to our knowledge the largest of any study to date. We have for the first time identified eSNPs to stomach, an important tissue in weight loss post-RYGB. We also show that as sample size increases, so does the number of observed eSNPs, as well as the number of tissue-independent eSNPs, leading us to believe that as sample sizes increase for these types of studies and more tissues are accessed, an eSNP to each gene may be identified. Through comparison of eSNPs identified in liver and adipose from random individuals, with the RYGB cohort we found that eSNPs to a number of genes linked to T2D, cholesterol biosynthesis, and obesity were specific to the obese cohort. Therefore, we hypothesize that sets of eSNPs may not only be tissue-specific, but cohort-specific as well, dependent on demographics and gene activity.

Analysis of the gene expression and gene connectivity from each of the tissues has allowed us to identify a number of genes of interest correlated to traits of extreme obesity and weight loss after surgery. Both analyses identified association to gene sets or modules enriched for inflammation-associated genes, indicating that inflammation plays a central role in obesity. While our results do not imply causality between the eQTLs, coexpression network modules, and the traits correlated with the modules, we have previously demonstrated in mouse studies that coexpression modules enriched for eQTLs that associate with phenotypes of interest are causal for the associated phenotypes (Chen et al. 2008), which represents one possible interpretation of the results we report here.

We have used the RYGB eSNPs to inform our association analysis as well as show the utility of eSNPs in interpreting GWAS for disease. In the future, this set of eSNPs from multiple metabolically relevant tissues will be of great use in informing association analysis with the potential to pinpoint genes and associated networks of interest that may lead to the design of therapeutics for metabolic disorders. Comparison to other liver eSNP data sets has shown that a number of the eSNPs associated with genes that have been linked to obesity phenotypes identified in this cohort are distinct to the obese population, reinforcing the importance of this eSNP data set in the study of metabolic disorders.

Methods

Data set description

Tissues were collected from a total of 1008 patients at the time of RYGB surgery at Massachusetts General Hospital between 2000 and 2007 (Supplemental Table 5). Samples were collected in RNAlater (Ambion/Applied Biosystems), stored at −80°, and shipped to the Rosetta Inpharmatics Gene Expression Laboratory in Seattle, Washington for extraction, amplification, labeling, and microarray processing in 2007. The samples processed ranged in size from 100 to 200 mg. Genomic DNA was extracted from liver tissues, and total RNA was extracted from liver, SA, OA, and stomach tissues. RNA from liver, OA, and SA was converted to fluorescently labeled cRNA that was hybridized to custom 44K DNA oligonucleotide microarrays manufactured by Agilent Technologies as described previously (Hughes et al. 2001; Schadt et al. 2008). The custom array consists of 4720 control probes and 39,820 noncontrol oligonucleotides. Samples with a 3′ bias >1 or <−1.5 from the mean of all samples were removed from the analysis, to prevent any bias from cRNA yield. RNA and gene expression profiling results were successfully collected from 651 liver, 848 OA, and 701 SA samples. RNA from 130 stomach samples collected in the RYGB cohort were amplified and labeled using a custom automated version of the NuGEN Ovation WB protocol and profiled on the Affymetrix Human U133A array. Samples with an RIN >5 and a 28S/18S ratio between 0.75 and 3.02 went on for profiling. Hybridization, labeling, and scanning using Affymetrix ovens, fluidics stations, and scanners followed recommended protocols (NuGEN). One hundred eighteen stomach samples were profiled. Each DNA sample was genotyped on the Illumina 650Y BeadChip array. Nine hundred fifty samples were successfully genotyped. Sex was confirmed using PLINK (Purcell et al. 2007). Identity by state (IBS) analysis was performed to identify related individuals. Eighteen parent–offspring, six sibling, and eight second-degree relatives were identified; four of these were related trios. Twenty-eight individuals were removed to eliminate IBS in the data set, leaving 922 samples for use in analysis. EIGENSTRAT was used to confirm reported race (Price et al. 2006).

Demographic information including age, race, gender, height, type of surgery, and year of surgery was collected for each patient. The cohort was predominantly (88%) self-reported "white" and female (75%). Fifty-nine percent of surgeries were performed laproscopically, and the rest were performed open. Weight and blood levels of leptin were collected for each individual at varying times from 1 yr pre-op to 7 yr post-op. The post-op data available on individuals vary depending on when the surgery was performed. BMI was calculated as (weight in kilograms)/(height in meters)^2. Information on diabetic status, insulin use, metformin use, and statin use was extracted from medical records when available and confirmed through independent chart review in one-third of the cohort. Forty-six percent of the cohort was identified as type 2 diabetic, 13% had received prescriptions for insulin, 5% had received a prescription for metformin, and 28% had received a prescription for a statin.

Microarray normalization

Gene expression data were generated using Rosetta Resolver gene expression analysis software (version 7.0, Rosetta Biosoftware) and MATLAB (The MathWorks). To remove bias in expression profiles related to potential latent variables unrelated to underlying biological processes, we implemented a normalization method based on control probes present on the microarrays. We separated the control probes into two classes: (1) specialty probes, such as spike-in probes or other probes designed to monitor the quality of the microarrays; and (2) border probes used to describe the geometry of the microarray. We then identified the Principal Components (PC) explaining the variability of each control probe class and then identified the same components from a randomly permuted data set. We performed 10,000 permutations for each set of control probes and selected Principal Components with *P*-values defined as {number of [var(random-PC_i) > var(PC_i)]}/(number of permutations) < 10^{-4}. Following this procedure, we identified a total of 14, 20, and 19 Principal Components for liver, OA, and SA, respectively. The expression data in each tissue for each probe are thus the residuals from a linear model fitting incorporating the significant PCs.

The next step in data normalization was to analyze the contribution of each of the experimental covariate's age, race, gender, surgery year, surgery type, and BMI to the gene expression. The distribution of *P*-values obtained from an ANOVA model for each covariate strongly suggests that correcting for age, gender, and surgery year was necessary (also confirmed by stepwise regression using Akaike information criteria [AIC]). EIGENSTRAT (Price et al. 2006) Principal Component 1 was also used as a covariate in all analysis as a surrogate for race.

A Matlab implementation of this algorithm is available upon request. The expression data are available at GEO super series Accession ID GSE24335. Genotyping data will be made available at dbGaP (https://dbgap.ncbi.nlm.nih.gov/), and all data are available upon request from Massachusetts General Hospital at http://www.samscore.org.

SNP–trait association analysis

SNP–trait associations were identified using the Kruskal-Wallis test (Kruskal and Wallis 1952). Additional covariates used in this analysis included diabetes status, insulin use, metformin use, and statin use. Association was calculated to BMI, leptin levels, and HDL at each available time point including pre-op (first pre-op reading), baseline (reading closest to time of surgery), early (within 90 d of surgery), post-op readings between 6 mo and 36 mo averaged, post-op 1 yr, post-op 2 yr, and post-op last reported reading. All time points were not available for every subject. Where SNP associations were identified to the same trait in high LD (>0.75) with each other, the SNP with the most significant *P*-value is reported. The false discovery rate (FDR) for significance was determined through permutations of each individual trait and time point. A 10% FDR was used for reporting results as significant, 20% FDR for suggestive.

eSNP identification

Cis- and *trans*-acting expression quantitative trait loci (eQTLs) were identified using a method similar to that previously described (Schadt et al. 2008). Briefly, eQTLs for gene expression traits were determined by identifying the SNP most strongly associated with each expression trait profiled on the array over all genotyped SNPs. An expression trait in our study is the basic measure of gene expression for a given transcript of interest monitored via the Agilent microarrays using oligonucleotide probes specific to the transcript of interest. The expression trait measure for a given sample is formed by taking the mean of the log ratio of the intensity measure for the corresponding probe in the sample divided by the intensity for the same probe in the reference sample. The mean log ratio statistic is therefore a quantitative measure of the relative abundance of the corresponding target in the sample relative to the reference sample. For a given expression trait, these mean log ratio measures were used as input into the R-based kruskal.test procedure, along with the genotype vectors for the corresponding SNPs of interest. The kruskal.test procedure converts the quantitative gene expression vector into a rank-ordered list based on the mean log ratio values and then tests whether the population medians are equal among the genotype groups. A *cis* eQTL occurs if the SNP and the transcript are within 1 Mb from each other regardless of the transcriptional start or stop of the transcript. All other eQTLs that are not *cis* are *trans* eQTLs. Association was tested to all expression traits measured on the array. Only associations that exceed 10% FDR are reported. FDR was estimated by permuting expression data.

Gene trait correlations

Gene trait correlations were calculated for pre-op BMI, pre-op leptin, excess weight loss at 1 yr (EWL), and nadir BMI in liver, OA, and SA using the Pearson correlation in R (corr function). The false discovery rate (FDR) was calculated using the R Q-value package (Dabney and Storey).

Coexpression modules

The 10,159 (the top 25%) most varying genes based on standard deviation were selected for constructing a weighted gene coexpression network for each of the three tissues (Zhang and Horvath 2005). The weighted network analysis begins with a matrix of the Pearson correlations between all gene pairs, then converts the correlation matrix into an adjacency matrix using a power function $f(x) = x^{\beta}$. The parameter β of the power function is determined in such a way that the resulting adjacency matrix, i.e., the weighted coexpression network, is approximately scale-free. To measure how well a network satisfies a scale-free topology, we use the fitting index proposed by Zhang and Horvath (2005), i.e., the model fitting index R^2 of the linear model that regresses $\log[p(k)]$ on $\log(k)$, where k is connectivity and $p(k)$ is the frequency distribution of connectivity. The fitting index of a perfect scale-free network is 1. For each data set, we select the smallest β that leads to an approximately scale-free network with the truncated scale-free fitting index $R^2 > 0.8$.

To explore the modular structures of the coexpression network, the adjacency matrix is further transformed into a topological overlap matrix (Ravasz et al. 2002). As the topological overlap between two genes reflects not only their direct interaction but also their indirect interactions through all the other genes in the network, previous studies (Ravasz et al. 2002; Zhang and Horvath 2005) have shown that topological overlap leads to more cohesive and biologically more meaningful modules. To identify modules of highly co-regulated genes, we used average linkage hierarchical clustering to group genes based on the topological overlap of their connectivity, followed by a dynamic cut-tree algorithm to dynamically cut clustering dendrogram branches into gene modules (Langfelder et al. 2008).

To distinguish between modules, each module was assigned a unique color identifier, with the remaining, poorly connected genes colored gray. Supplemental Figure 1 shows the topological overlap matrix (TOM) heatmaps for each module from each tissue. In this type of map, the rows and the columns represent genes in a symmetric fashion, and the color intensity represents the interaction strength between genes.

Acknowledgments

The authors acknowledge Abby Cange for her work in making the data accessible, and the Rosetta Gene Expression Laboratory.

References

Aslan IR, Campos GM, Calton MA, Evans DS, Merriman RB, Vaisse C. 2010. Weight loss after Roux-en-Y gastric bypass in obese patients heterozygous for MC4R mutations. *Obes Surg* doi: 10.1007/s11695-010-0295-8.

Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, Penninx BW, Janssens AC, Wilson JF, Spector T, et al. 2009. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* **41:** 47–55.

Buchwald H, Avidor Y, Braunwald E, Jensen MD, Pories W, Fahrbach K, Schoelles K. 2004. Bariatric surgery: A systematic review and meta-analysis. *JAMA* **292:** 1724–1737.

Chambers JC, Elliott P, Zabaneh D, Zhang W, Li Y, Froguel P, Balding D, Scott J, Kooner JS. 2008. Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat Genet* **40:** 716–718.

Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, et al. 2008. Variations in DNA elucidate molecular networks that cause disease. *Nature* **452:** 429–435.

Cotsapas C, Speliotes EK, Hatoum IJ, Greenawalt DM, Dobrin R, Lum PY, Suver C, Chudin E, Kemp D, Reitman M, et al. 2009. Common body mass index–associated variants confer risk of extreme obesity. *Hum Mol Genet* **18:** 3502–3507.

Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M, et al. 2009. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325:** 1246–1250.

Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, et al. 2008. Genetics of gene expression and its effect on disease. *Nature* **452:** 423–428.

Farooqi IS, O'Rahilly S. 2008. Mutations in ligands and receptors of the leptin–melanocortin pathway that lead to obesity. *Nat Clin Pract Endocrinol Metab* **4:** 569–577.

Futatsugi A, Nakamura T, Yamada MK, Ebisui E, Nakamura K, Uchida K, Kitaguchi T, Takahashi-Iwanaga H, Noda T, Aruga J, et al. 2005. IP3 receptor types 2 and 3 mediate exocrine secretion underlying energy metabolism. *Science* **309:** 2232–2234.

Ganesan AK, Kho Y, Kim SC, Chen Y, Zhao Y, White MA. 2007. Broad spectrum identification of SUMO substrates in melanoma cells. *Proteomics* **7:** 2216–2221.

Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, et al. 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* **19:** 342–347.

Johnson AD, O'Donnell CJ. 2009. An open access database of genome-wide association results. *BMC Med Genet* **10:** 6. doi: 10.1186/1471-2350-10-6.

Kamatani Y, Wattanapokayakit S, Ochi H, Kawaguchi T, Takahashi A, Hosono N, Kubo M, Tsunoda T, Kamatani N, Kumada H, et al. 2009. A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. *Nat Genet* **41:** 591–595.

Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, et al. 2008. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* **40:** 189–197.

Keller MP, Choi Y, Wang P, Davis DB, Rabaglia ME, Oler AT, Stapleton DS, Argmann C, Schueler KL, Edwards S, et al. 2008. A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility. *Genome Res* **18:** 706–716.

Klos KL, Sing CF, Boerwinkle E, Hamon SC, Rea TJ, Clark A, Fornage M, Hixson JE. 2006. Consistent effects of genes involved in reverse cholesterol transport on plasma lipid and apolipoprotein levels in CARDIA participants. *Arterioscler Thromb Vasc Biol* **26:** 1828–1836.

Koskivirta I, Rahkonen O, Mayranpaa M, Pakkanen S, Husheem M, Sainio A, Hakovirta H, Laine J, Jokinen E, Vuorio E, et al. 2006. Tissue inhibitor of metalloproteinases 4 (TIMP4) is involved in inflammatory processes of human cardiovascular pathology. *Histochem Cell Biol* **126:** 335–342.

Kruskal W, Wallis W. 1952. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc* **47:** 583–621.

Langfelder P, Zhang B, Horvath S. 2008. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut library for R. *Bioinformatics* **24:** 719–720.

MacLaren RE, Cui W, Lu H, Simard S, Cianflone K. 2010. Association of adipocyte genes with ASP expression: a microarray analysis of subcutaneous and omental adipose tissue in morbidly obese subjects. *BMC Med Genomics* **3:** 3. doi: 10.1186/1755-8794-3-3.

McKay RM, McKay JP, Suh JM, Avery L, Graff JM. 2007. Tripeptidyl peptidase II promotes fat formation in a conserved fashion. *EMBO Rep* **8:** 1183–1189.

Meigs JB, Manning AK, Fox CS, Florez JC, Liu C, Cupples LA, Dupuis J. 2007. Genome-wide association with diabetes-related traits in the Framingham Heart Study. *BMC Med Genet* (Suppl 1) **8:** S16. doi: 10.1186/1471-2350-8-S1-S16.

Meyre D, Delplanque J, Chevre JC, Lecoeur C, Lobbens S, Gallina S, Durand E, Vatin V, Degraeve F, Proenca C, et al. 2009. Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat Genet* **41:** 157–159.

National Health and Nutrition Examination Survey. 2008. Prevalence of overweight, obesity and extreme obesity among adults: United States, trends 1976–80 through 2005–2006. NCHS Health E-Stats. http://www.cdc.gov/nchs/data/hestat/overweight/overweight_adult.pdf.

Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, Dermitzakis ET. 2010. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet* **6:** e1000895. doi: 10.1371/journal.pgen.1000895.

Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. 2010. Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLoS Genet* **6:** e1000888. doi: 10.1371/journal.pgen.1000888.

Orho-Melander M, Melander O, Guiducci C, Perez-Martinez P, Corella D, Roos C, Tewhey R, Rieder MJ, Hall J, Abecasis G, et al. 2008. Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. *Diabetes* **57:** 3112–3121.

Park JJ, Berggren JR, Hulver MW, Houmard JA, Hoffman EP. 2006. GRB14, GPD1, and GDF8 as potential network collaborators in weight loss-induced improvements in insulin action in human skeletal muscle. *Physiol Genomics* **27:** 114–121.

Pooley EC, Fairburn CG, Cooper Z, Sodhi MS, Cowen PJ, Harrison PJ. 2004. A 5-HT2C receptor promoter polymorphism (HTR2C–759C/T) is associated with obesity in women, and with resistance to weight loss in heterozygotes. *Am J Med Genet B Neuropsychiatr Genet* **126:** 124–127.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38:** 904–909.

Pruitt KD, Harrow J, Harte RA, Wallin C, Diekhans M, Maglott DR, Searle S, Farrell CM, Loveland JE, Ruef BJ, et al. 2009. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res* **19:** 1316–1323.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81:** 559–575.

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. 2002. Hierarchical organization of modularity in metabolic networks. *Science* **297:** 1551–1555.

Robitaille J, Houde A, Lemieux S, Perusse L, Gaudet D, Vohl MC. 2007. Variants within the muscle and liver isoforms of the carnitine palmitoyltransferase I (CPT1) gene interact with fat intake to modulate indices of obesity in French-Canadians. *J Mol Med* **85:** 129–137.

Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J, Jones CG, Zaitlen NA, Varilo T, Kaakinen M, et al. 2009. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* **41:** 35–46.

Salonen JT, Uimari P, Aalto JM, Pirskanen M, Kaikkonen J, Todorova B, Hypponen J, Korhonen VP, Asikainen J, Devine C, et al. 2007. Type 2 diabetes whole-genome association study in four populations: The DiaGen consortium. *Am J Hum Genet* **81:** 338–345.

Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, et al. 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316:** 1331–1336.

Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, et al. 2008. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* **6:** e107. doi: 10.1371/journal.pbio.0060107.

Shen H, Pollin TI, Damcott CM, McLenithan JC, Mitchell BD, Shuldiner AR. 2009. Glucokinase regulatory protein gene polymorphism affects postprandial lipemic response in a dietary intervention study. *Hum Genet* **126:** 567–574.

Smith SJ, Cases S, Jensen DR, Chen HC, Sande E, Tow B, Sanan DA, Raber J, Eckel RH, Farese RV Jr. 2000. Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking Dgat. *Nat Genet* **25:** 87–90.

Swenson TL. 1991. The role of the cholesteryl ester transfer protein in lipoprotein metabolism. *Diabetes Metab Rev* **7:** 139–153.

Tice JA, Karliner L, Walsh J, Petersen AJ, Feldman MD. 2008. Gastric banding or bypass? A systematic review comparing the two most popular bariatric procedures. *Am J Med* **121:** 885–893.

Vaxillaire M, Cavalcanti-Proenca C, Dechaume A, Tichet J, Marre M, Balkau B, Froguel P. 2008. The common P446L polymorphism in GCKR inversely modulates fasting glucose and triglyceride levels and reduces type 2 diabetes risk in the DESIR prospective general French population. *Diabetes* **57:** 2253–2257.

Vock C, Doring F, Nitz I. 2008. Transcriptional regulation of HMG-CoA synthase and HMG-CoA reductase genes by human ACBP. *Cell Physiol Biochem* **22:** 515–524.

Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM, et al. 2008a. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* **40:** 161–169.

Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM, Berndt SI, Elliott AL, Jackson AU, Lamina C, et al. 2008b. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* **41:** 25–34.

Yang X, Zhang B, Molony C, Chudin E, Hao K, Zhu J, Gaedigk A, Suver C, Zhong H, Leeder JS, et al. 2010. Systematic genetic and genomic analysis of cytochrome P450 enzyme activities in human liver. *Genome Res* **20:** 1020–1036.

Zhang B, Horvath S. 2005. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4:** Article 17. doi: 10.2202/1544-6115.1128.

Received July 13, 2010; accepted in revised form April 4, 2011.