Classification; BIOLOGICAL SCIENCES: Medical Sciences

Expression-based genome-wide association study links *CD44* in adipose tissue with type 2 diabetes

Keiichi Kodama[a,b], Momoko Horikoshi[c], Kyoko Toda[d,1], Satoru Yamada[e,1], Kazuo Hara[c,1], Junichiro Irie[e,f,1], Marina Sirota[a,b], Alexander A. Morgan[a,b], Rong Chen[a,b], Hiroshi Ohtsu[g], Shiro Maeda[h], Takashi Kadowaki[c], and Atul J. Butte[a,b,2]

[a]Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, 1265 Welch Road, Stanford, CA 94305, USA

[b]Lucile Packard Children's Hospital, 725 Welch Road, Palo Alto, CA 94304 USA

[c]Department of Metabolic Diseases, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

[d]Division of Basic Research, Biomedical Laboratory, Kitasato Institute Hospital, Kitasato University, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8642, Japan

[e]Diabetes Center, Kitasato Institute Hospital, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8642, Japan

[f]Department of Internal Medicine, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan

[g]Department of Clinical Trial Data Management, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan

[h]Laboratory for Endocrinology and Metabolism, Center for Genomic Medicine, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan

[1]K.T., S.Y., K.H. and J.I. contribute equally to this work.

[2]To whom correspondence should be addressed. E-mail: abutte@stanford.edu

Correspondence to: Atul J. Butte

Division of Systems Medicine, Department of Pediatrics, Stanford University School of Medicine, 1265 Welch Road, Room X-163 MS-5415, Stanford, CA, 94305-5415

E-mail: abutte@stanford.edu

Phone: +1-650-723-3465 Fax: +1-650-723-7070

Key Words: bioinformatics, microarray, diabetes, CD44, secreted phosphoprotein 1, SPP1, osteopontin, OPN

[Manuscript information]

Main manuscript

Abstract

Text

Materials and Methods

Acknowledgments

References; 42

Figure legends

Figure count; 5

Supporting information (SI)

SI Methods

SI Figure legends

SI figure count; 6

SI table count; 3

ABSTRACT

Type 2 diabetes (T2D) is a complex, polygenic disease affecting nearly 300 million people world-wide. T2D is primarily characterized by insulin resistance, and growing evidence has indicated the causative link between adipose tissue inflammation and the development of insulin resistance. Genetic association studies have successfully revealed a number of important genes consistently associated with T2D to date. However, these robust T2D-associated genes do not fully elucidate the mechanisms underlying the development and progression of the disease. Here we report an alternative approach, gene-expression based genome-wide association study (eGWAS): searching for genes repeatedly implicated in functional microarray experiments (often publicly-available). We performed an eGWAS across 130 independent experiments (totally 1,175 T2D case-control microarrays) to find additional genes implicated in the molecular pathogenesis of T2D, and identified the immune-cell receptor *CD44* as our top candidate ($P = 8.5 \times 10^{-20}$). We found *CD44* deficiency in a diabetic mouse model ameliorates insulin resistance and adipose tissue inflammation, and also found that anti-CD44 antibody treatment decreases blood glucose levels and adipose tissue macrophage accumulation in a high-fat-diet fed mouse model. Further, in humans, we observed CD44 is expressed in inflammatory cells in obese adipose tissue, and discovered serum CD44 levels were positively correlated with insulin resistance and glycemic control. *CD44* likely plays a causative role in the development of adipose tissue inflammation and insulin resistance in rodents and humans. Genes repeatedly implicated in publicly-available experimental data may have novel functionally important roles in T2D and other complex diseases.

\body

INTRODUCTION

Type 2 diabetes (T2D) is a common multifactorial disease characterized by hyperglycemia primarily resulting from peripheral insulin resistance, and growing functional evidence has indicated the causative link between adipose tissue inflammation and the development of insulin resistance (1, 2). In the past decade, a number of genome-wide genetic association studies (GWAS) have revealed forty loci consistently associated with susceptibility to T2D and have rapidly expanded our knowledge of the genetic architecture of this disease (3-13). However, the genes located in or near these loci do not fully elucidate the tissue-specific molecular mechanisms underlying the development of T2D.

A large number of experiments using genome-wide gene-expression microarray measurements have been also performed over the past decade; however, there has been very little success in fully identifying functionally important genes in the pathogenesis of T2D. Since a large number of genes are often detected as significant in each microarray experiment, it may be hard to sub-select optimal candidates from individual studies for further verification. Combination of genome-wide data from two or more experiments has been performed for obesity (14-16), T2D (14), and other multifactorial disorders (17-20). Several of these methods have used microarray technology to focus on candidate genes already implicated in a region of a congenic or model animal (*accelerated positional candidate identification*). More recently, investigators have applied microarrays to genetics by considering gene expression levels as quantitative traits (eQTLs) and finding relations between gene variants and transcripts, with successful

application to the identification of genes and targets for T2D (21, 22). The need for large numbers of simultaneously acquired genetic and gene-expression measurements within a single study makes this approach less scalable.

We suspected that the large number of molecular measurements from experimental results that are now publicly-available, due to requirements from journals and funding agencies (23), could be used as an alternative scalable source of data. The strategy of finding commonly implicated genes across related -- but deliberately varied -- experimental conditions has been theorized to yield less overfit, potentially more generalizable causal factors (24), and publicly-available data could be used as a source of these varied experimental vantage points for a condition. In this report, we propose the application of a gene-expression based genome-wide association study (eGWAS), a meta-analysis method for computing the likelihood of finding repeated differential expression for every gene across a large number of case and control microarray experiments, compared to expected. Our hypothesis is that those genes most repeatedly implicated across a large set of experimental representations of T2D can serve as data-driven causal T2D genes, and candidates for validation. This approach is only feasible as many of these source raw experimental results are publicly-available; here, we integrated 130 independent microarray experiments for T2D. In this case, our T2D candidates were found independent of any knowledge about insulin signaling, glucose or lipid metabolism. For our top candidate gene, identification was followed by confirmatory functional studies using mouse models and samples from human subjects. (**Fig. S1**).

RESULTS

eGWAS identifies *CD44* as a functional candidate gene for T2D.

We carried out an eGWAS for T2D using 130 independent microarray experiments, totaling 1,175 samples collected from public repositories (**Fig. 1**, **Table S1** and **S2** and **SI Methods**). We ranked all 24,898 genes by the likelihood that repeated differential expression for that gene was due to chance, then controlled for multiple-hypothesis testing. To overview which molecular functions are most shared in the highest-ranked genes in our T2D eGWAS, we took the top 127 genes (**Table S3**; Bonferroni threshold, $P < 2.0 \times 10^{-6}$) from our eGWAS, and then estimated the enrichment of Gene Ontology (GO) terms. Interestingly, "receptor activity" and "receptor binding" functions were the most implicated of the top-ranked genes ("receptor activity"; 38%, "receptor binding"; 19%) (**Fig. S2**). This suggested that a number of top-ranked genes on our list are involved in intra- and inter-tissue signaling cascades in the development of T2D (1).

Our top-most T2D candidate gene was *CD44* (**Fig. 1:** χ^2 analysis; $P = 8.5 \times 10^{-20}$, **Fig. S3:** Fisher's exact test; $P = 6.1 \times 10^{-17}$, **Fig. S4:** weighted Z-method), markedly differentially expressed in experiments studying diabetes in adipose tissue, compared to other tissues (**Fig. S5**). *CD44* is located on chromosome 11p13 and codes for a cell-surface glycoprotein, an immunological cell (macrophage/T cell) receptor, involved in inflammatory cell migration and activation. Interestingly, one of the known ligands for *CD44*, secreted phosphoprotein 1 (SPP1; also known as osteopontin (OPN)), a Th1 cytokine secreted by immunological cells (macrophages), was also included in the top-ranked genes (**Fig. 1:** χ^2 analysis; $P = 1.3 \times 10^{-11}$, **Fig. S3:** Fisher's exact test; $P = 3.8 \times 10^{-10}$). Recent studies have indicated that obese adipose tissue is hallmarked with chronic, low-grade inflammation, and that inflammation plays a central role in the development of

insulin resistance (1, 2). Although the contributions of the *CD44* encoded protein to the molecular pathogenesis of T2D have not yet been reported, SPP1 was previously reported as a link between adipose tissue inflammation (stromal infiltration by inflammatory cells) and the development of insulin resistance in a murine model of diet-induced obesity (25). Furthermore, the expression profile of *CD44* and *SPP1* are coordinately dysregulated, especially in adipose tissue (**Fig. S6**; coordinate dysregulation rate = 0.90). These findings suggest that CD44 might have a key role in mediating obesity-induced adipose tissue inflammation and insulin resistance.

***CD44* expression increases in obese adipose tissue.**

High-fat feeding in C57BL/6J mice leads to the development of obesity, adipose inflammation and insulin resistance (25, 26). To examine whether the *CD44* mRNA transcript is expressed in adipose tissue and modulated by obesity, C57BL/6J mice were maintained either on a normal-fat diet (NFD; 12% of total calories from fat) or high-fat diet (HFD; 60% of total calories from fat) for 16 weeks (n = 8/group). Compared with the NFD group, mice fed a HFD gained 37% more weight after a 16-week feeding period (29.9 ± 0.5 versus 40.9 ± 1.6 g, $P = 5.8 \times 10^{-8}$). Epididymal white adipose tissue (EWAT) was removed from these mice to analyze *CD44* mRNA expression levels. Feeding a HFD resulted in a significant 11.3-fold increase of *CD44* mRNA levels in adipose tissue compared to NFD (**Fig. 2*A***). In order to establish the presence of CD44 protein in adipose tissue, an immunohistochemical localization of CD44 was performed on EWAT isolated from HFD mice. We found that CD44 was abundantly expressed in inflammatory cells within adipose tissue (**Fig. 2*B***). These results clearly indicate that $CD44^+$ cells accumulated in EWAT of diet-induced obese mice. In addition, we

confirmed that the expression levels of *SPP1* mRNA in adipose tissue in the HFD group was significantly higher than that in the NFD mice (0.005 ± 0.003 versus 0.06 ± 0.02, $P < 0.05$), similar to previous reports (25). Interestingly, we also found that *CD44* mRNA expression level was positively correlated with the *SPP1* mRNA expression level in the HFD mice (**Fig. 2*C*:** r = 0.78, $P = 0.02$), suggesting that CD44 and SPP1 may be closely related in obese adipose tissue.

***CD44* deficiency ameliorates adipose tissue inflammation and insulin resistance.**

To next determine the contribution of CD44 to the development of adipose tissue inflammation and insulin resistance, we fed male *CD44* $^{-/-}$ and diabetes-prone C57BL/6J (*CD44* $^{+/+}$) mice with either a HFD (n = 16/group) or a NFD (n = 10/group) for 12 weeks, and performed immunohistochemical analysis and metabolic measurements on these mice. There were no significant differences in body weights between *CD44* $^{-/-}$ and *CD44* $^{+/+}$ mice after feeding a NFD or a HFD (NFD: *CD44* $^{-/-}$ 28.5 ± 0.5 g versus *CD44* $^{+/+}$ 29.5 ± 0.5 g; HFD: 38.1 ± 1.5 g versus 39.5 ± 1.2 g). In *CD44* $^{+/+}$ mice fed a HFD, we frequently observed the accumulation of inflammatory cells (macrophages) forming crown-like structures (CLSs) surrounding adipocytes in obese visceral adipose tissue. However, *CD44* $^{-/-}$ mice fed a HFD exhibited strikingly less macrophage infiltration into the stroma of adipose tissue compared to *CD44* $^{+/+}$ mice fed a HFD (**Fig. 3*A***). Fasting blood glucose levels were significantly lower in *CD44* $^{-/-}$ mice fed a HFD compared with the diabetes-prone *CD44* $^{+/+}$ mice fed a HFD (**Fig. 3*B***). Glucose tolerance tests also indicated that *CD44* $^{-/-}$ mice fed a HFD were significantly more efficient in their ability to clear intraperitoneally injected glucose than *CD44* $^{+/+}$ mice fed a HFD (**Fig. 3*C***; solid lines) despite similar insulin secretory responses after the injection of glucose.

Furthermore, insulin sensitivity, as measured by insulin tolerance test, showed that *CD44* $^{-/-}$ mice fed a HFD were significantly more efficient at insulin-mediated suppression of blood glucose than *CD44* $^{+/+}$ mice fed a HFD (**Fig. 3*D***; solid lines). Interestingly, even in the mice fed a NFD, the ameliorative effects of *CD44* deficiency on insulin sensitivity was observed in both glucose and insulin tolerance tests (**Fig. 3 *C* and *D***; dashed lines). These results that we obtained with CD44-deficient mice confirm that CD44 molecules are essential for macrophage recruitment and inflammation in adipose tissue and the development of insulin resistance in diet-induced obese mice.

CD44 blockade decreases blood glucose levels and adipose macrophage infiltration.

Several *in vivo* studies have shown that anti-CD44 monoclonal antibody (CD44 mAb) treatment exhibits robust anti-inflammatory effects in animal models of immune-mediated diseases (27-30). We therefore sought to investigate whether CD44 blockade might demonstrate a therapeutic effect on T2D. We performed daily intraperitoneal injections of CD44 mAb in diabetic model mice for one week, and found that blood glucose levels and adipose macrophage accumulation were significantly reduced in CD44 mAb treated mice compared to isotype-control treated mice, despite continuing on the HFD and similar body weight increase during the treatment (**Fig. 4 *A*** and ***B***). When adipose inflammation was quantified as the average number of CLSs per low power field, EWAT from CD44 mAb treated mice contained significantly fewer inflammatory cells compared with control treated mice (2.4 ± 0.5 versus 5.4 ± 0.7; $P = 0.0005$). Collectively, these effects of CD44 mAb clearly show that CD44 molecules are required for the recruitment of macrophages into obese adipose tissue and the maintenance of inflammatory reactions there, and the CD44 receptor may be useful as a therapeutic target

for T2D.

To gain additional insight into the clinical importance of CD44 in obese fat, we performed an immunohistochemical analysis of CD44 in omental adipose tissue in human obesity. Consistent with our mouse model observation, we discovered that $CD44^+$ cells infiltrated into the stroma of adipose tissue in human obese subjects, suggesting that CD44 molecules may mediate macrophage migration into obese adipose tissue in humans (**Fig. 5*A***).

Soluble CD44 shed from cell surfaces exists in normal human serum. To estimate the relevance between CD44 protein and glucose homeostasis in human subjects, we evaluated the relationship between serum levels of CD44 and metabolic traits in human, and found that serum CD44 was positively correlated with glycemic control and insulin resistance as estimated through HbA1c (n=55, r = 0.49, $P < 0.001$) and HOMA-IR (n=55, r = 0.29, $P = 0.03$) (**Fig. 5 *B*** and ***C***). We then classified the 55 subjects into two groups according to the WHO criteria (31): hyperglycemia (n=21: 'diabetes mellitus' + 'impaired glucose regulation'); and normoglycemia (n=34: 'Normal Glucose Tolerance'), and found that the serum levels of CD44 were significantly higher in the hyperglycemic group than normoglycemic group (246.9 ± 15.0 versus 209.5 ± 9.8 ng/ml, $P = 0.02$). These results suggest that CD44 protein may be released from insulin-resistant and diabetic tissues into circulation in humans.

DISCUSSION

Adipose tissue inflammation is now thought to be a pivotal event leading to the metabolic syndrome, insulin resistance and T2D. Genome-wide experimental methods to identify disease genes, such as association studies (GWAS), linkage studies (GWL),

classical quantitative trait loci (cQTL) and expression QTL (eQTL) analyses, have been performed for T2D by many researchers to date, and these methods have revealed a number of loci to be linked with T2D (3-14, 21, 32-35). However these genetically mapped loci do not fully account for the tissue-specific mechanisms underlying the development of T2D. In this report, we proposed an alternative methodology, gene-expression based genome-wide association study (eGWAS), computing the likelihood of finding repeated differential expression of a gene in disease-related tissues using thousands of case-control microarray samples. To detect additional genes functionally implicated in the molecular pathogenesis of T2D, we successfully performed an eGWAS to identify T2D candidate genes, and verified our top candidate gene, *CD44*, an immunological cell receptor, plays a significant role in the development of adipose tissue inflammation and insulin resistance in mouse models and human subjects.

In our T2D eGWAS, we identified in total 127 genes as significantly repeatedly dysregulated with *P* values less than 2.0×10^{-6} (under the Bonferroni-corrected threshold), and rediscovered several genes that have been shown previously to be important in T2D pathogenesis, including *TCF7L2*, *PPARG*, *KCNQ1*, *IDE*, *CD36*, *GLUT4* and *LEPR*, supporting the validity of our methodology. Of the 127 genes, we found that more than half were implicated in the "receptor" or "ligand" activity by using the Gene Ontology (GO) term enrichment analysis (**Fig. S2**). Interestingly, *SPP1*, encoding a ligand for the CD44 receptor, was shown to be also included in our top-ranked gene list (**Fig. 1** and **Table S3**). Furthermore, our eGWAS provided the prediction of tissue specificity of gene-expression by calculating a distribution of scores for each gene across tissues, indicating that *CD44* mRNA expression was more highly up-regulated in adipose tissue

in diabetes than other tissues (**Fig. S5)**. These analysis results led us to the speculation that our top-most candidate gene, *CD44*, encoding an immune-cell surface receptor, may be implicated in adipose tissue inflammation causing insulin resistance in obesity, given the fact that the CD44 receptor is known to regulate immune-cell migration and activation and its ligand SPP1 has been previously reported to mediate macrophage infiltration into obese adipose tissue (25). When interpreting this hypothesis, however, we need to consider the possibility that the dysregulation of *CD44* expression in relevant tissues can result secondarily from hyperglycemia and diabetes. We therefore performed verification experiments to see if the dysregulated expression of *CD44* gene in obese adipose tissue can be accepted as the cause of insulin resistance.

In our functional tests for the *CD44* gene products, we showed that knocking out the receptor *CD44* leads to striking reduction in immune-cell infiltration into visceral adipose tissue and improvements in insulin sensitivity in mouse models, and anti-CD44 monoclonal antibody treatment decrease the blood glucose levels and visceral adipose tissue macrophages in diabetic obese mice. In addition, we showed that higher serum levels of the soluble-form of CD44 correlate with increasing hyperglycemia and insulin resistance in humans. Although the expression of CD44 in macrophages and T cells in obese adipose tissue has been previously reported (36, 37), to our knowledge, the present study is the first to directly address the functional role of CD44 in adipose tissue inflammation. The findings that systemic glucose intolerance is ameliorated by CD44 depletion and blockade using CD44 mAb strongly suggest that CD44-dependent adipose inflammation has an impact on systemic metabolism. However, further studies are needed to determine which immune-cell population primarily expresses CD44 receptor

(macrophages or T cells?) and how the CD44 molecules initiate and maintain immune-cell infiltration into adipose tissue (through the SPP1 signals?). Even so, these results indicate the significance of CD44 immune-receptor as a possible therapeutic target for T2D and a novel biomarker for insulin resistance.

In conclusion, we discovered that an immune-cell receptor gene, *CD44*, is pathogenetically implicated in the development of adipose tissue inflammation and insulin resistance, using a data-driven candidate gene approach by using an eGWAS method of integrating over a thousand of publicly-available genome-wide functional microarrays related to T2D. Application of the eGWAS methodology to publicly-available data can yield promising candidate genes that are differentially expressed in T2D-relevant tissues, independently of knowledge about insulin signaling, glucose or lipid metabolism. We suggest that a data-driven approach can enable investigators to consider glucose homeostasis phenotypes from a different point of view, and notice new pathways that could be involved in the development of T2D. Though GWAS and other genetic analyses will continue as the method-of-choice for the next few years, an eGWAS approach could complement these studies to yield additional pathogenetically important genes for many other complex diseases using this integrated data-driven approach.

MATERIALS AND METHODS

See *SI Methods* for further descriptions.

Expression-based genome-wide association study (eGWAS).

All T2D-related genome-wide microarray experiments used for this study were collected from three public data sources; the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/), the Diabetes Genome Anatomy Project (DGAP; http://www.diabetesgenome.org/) and the Nuclear Receptor Signaling Atlas (NURSA; http://www.nursa.org/). There were a total of 1,175 samples (591 T2D cases and 584 controls) in 130 independent data sets. To estimate differences between groups of samples from diabetic subjects and groups representing control, raw post-quantitation microarray data were re-analyzed using Significance Analysis of Microarrays software (SAM) (38). For each gene in every microarray experiment with three or more samples in each group, we calculated a D-score (*di*), which denotes the standardized change in gene expression:

$$d_i = \frac{\bar{x}_{i-t2d} - \bar{x}_{i-control}}{S_i + S_0}$$

where $\bar{x}_{i-t2d}$ is the mean expression level of *gene i* in group T2D, $\bar{x}_{i-control}$ is the mean expression level of *gene i* in group control, S_i is the standard deviation for the numerator calculation, and S_0 is a small positive constant. We considered genes to be significantly dysregulated with either an absolute value of the D-score ≥ 2 or a fold change ≥ 2 between controls and cases. We then converted all probe identifiers across the various microarray platforms for mouse, rat and human to the latest human Entrez Gene identifiers using our previously published AILUN system (39). Gene expression profiles

were assigned in our eGWAS database according to the standardized (human) Entrez Gene ID. There were 24,898 genes in the database in total. For every one of the 24,898 genes, we counted the observed number of microarray experiments in which each gene was significantly dysregulated. We then calculated *P* values from the number of positive/negative experiments for every one of genes and sum of the number of positive/negative experiments for all other genes, using a 2×2 χ^2 analysis (**Fig. 1**) or a Fisher's exact test as an alternative (**Fig. S3**), and ranked all the genes according to their *P* values ($-\log_{10}(P)$). A third method, Liptak-Stouffer's weighted Z-method (40), provided additional support for *CD44* (**Fig. S4**).

Animal experiments.

Mice for breeding, C57BL/6J wild-type (diabetes-prone) and *CD44*-deficient mice backcrossed to C57BL/6J for at least 10 generations (B6.Cg-*Cd44*^{tm1Hbg}/J), were obtained from The Jackson Laboratory. Mice were housed in a barrier facility under specific pathogen-free conditions. The Animal Care and Use Committee of Kitasato University (Tokyo, Japan) approved all animal experiments.

Human studies.

Venous peripheral blood samples were collected from human subjects who went through a 75g oral glucose tolerance test after an overnight fast (n=55: gender (M/F); 36/19). HbA1c was measured in Japan Diabetes Society (JDS)-HbA1c units, and then converted to National Glycohemoglobin Standardization Program (NGSP) levels by the formula HbA1c (%) (NGSP) = HbA1c (JDS) (%) + 0.4% (41). We then calculated homeostasis model assessment as an index of insulin resistance (HOMA-IR = fasting plasma insulin (μU/ml) × fasting plasma glucose (mg/dl) / 405) as previously described

(42). Serum sCD44std (standard soluble CD44) and SPP1 concentrations were determined using a quantitative enzyme-linked immunosorbent assay (ELISA) technique (sCD44std ELISA; Bender MedSystems, Vienna, Austria, Human Osteopontin Quantikine ELISA; R&D Systems, Minneapolis, MN).

Informed consent was obtained from all of the subjects enrolled in this study and the protocol was approved by the ethics committee of the University of Tokyo.

Immunohistochemistry.

For histological analysis of CD44 expression in adipose tissue, EWAT was removed from mouse models, and omental adipose tissue obtained from consented donors undergoing elective gastric bypass surgery (lot number. OM020304B) was purchased from Zen-Bio, Inc. (Research Triangle Park, NC). Formalin-fixed paraffin-embedded sections were stained with mouse monoclonal antibody against CD44 at 1:50 dilution (DF1485/sc-7297; Santa Cruz Biotechnology, Inc.), followed by reactions with anti-mouse immunoglobulins-HRP, and anti-fluorescein-HRP. 3,3'-Diaminobenzidine (DAB) was used as a chromogen.

Analysis of inflammatory cell (macrophage) content in EWAT was performed on tissue pads isolated from model mice. Formalin-fixed paraffin-embedded sections were incubated overnight with primary antibody: Purified Anti Mouse MAC-2 Monoclonal Antibody (CL8942AP, 1:100; Cedarlane laboratories, Ontario, Canada), and stained using Histofine® Simple Stain Mouse MAX-PO (R) secondary antibody (Nichirei Biosciences, Inc., Tokyo, Japan) with a DAB solution.

Statistics.

For verification studies in mouse models and human subjects, comparisons

between two groups were performed using the 2-tailed Welch's t-test. Two-way repeated measures ANOVA was used to examine the treatment (antibodies type) × time interaction on blood glucose changes from baseline. *P* values of less than 0.05 were considered significant. All experimental data are represented as mean ± standard error unless otherwise noted.

Acknowledgments

We thank Saori Ohta of Kitasato University School of Pharmacy for her support of animal experiments. We thank Dr. Shojiro Morinaga of Department of Pathology, Kitasato Research Institute Hospital, for his assistance in the histological analysis. We thank Dr. Damon Tojjar of Department of Clinical Sciences, Lund University, Scania University Hospital, for his suggestions in preparing the manuscript. We thank Dr. Purvesh Khatri in the Division of Systems Medicine, Stanford University School of Medicine, for his suggestions in preparing the manuscript. This work was supported by the grants from the Howard Hughes Medical Institute, National Library of Medicine (R01 LM009719 and K22 LM008261), and Lucile Packard Foundation for Children's Health.

Author Contributions

K.K., M.H., K.H., S.M., T.K., and A.J.B. designed research. K.K., M.H., K.T., S.Y., J.I., K.H. S.M., T.K., and A.J.B performed research. M.S., A.A.M., and R.C. contributed new computational analysis tools. K.K., M.H., M.S., A.A.M., R.C., H.O., K.H., S.M., T.K., and A.J.B. analyzed data. All authors participated in data interpretation. K.K., M.H., K.H., S.M., T.K., and A.J.B. prepared the report. All authors provided critical review of the draft and approved the final version.

REFERENCES

1. Hotamisligil GS (2006) Inflammation and metabolic disorders. *Nature* 444(7121):860-867.
2. Shoelson SE, Lee J, & Goldfine AB (2006) Inflammation and insulin resistance. *J Clin Invest* 116(7):1793-1801.
3. The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145):661-678.
4. Scott LJ*, et al.* (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316(5829):1341-1345.
5. Saxena R*, et al.* (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316(5829):1331-1336.
6. Sladek R*, et al.* (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 445(7130):881-885.
7. Zeggini E*, et al.* (2008) Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 40(5):638-645.
8. Unoki H*, et al.* (2008) SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet* 40(9):1098.
9. Yasuda K*, et al.* (2008) Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* 40:1092.
10. Dupuis J*, et al.* (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 42(2):105-116.

11. Saxena R, *et al.* (2010) Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* 42(2):142-148.

12. Voight BF, *et al.* (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 42(7):579-589.

13. Yamauchi T, *et al.* (2010) A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. *Nat Genet* 42(10):864-868.

14. Aitman TJ, *et al.* (1999) Identification of Cd36 (Fat) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nat Genet* 21(1):76-83.

15. Schadt EE, *et al.* (2005) An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* 37(7):710-717.

16. Chen Y, *et al.* (2008) Variations in DNA elucidate molecular networks that cause disease. *Nature* 452(7186):429-435.

17. Karp CL, *et al.* (2000) Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma. *Nat Immunol* 1(3):221-226.

18. Klein RF, *et al.* (2004) Regulation of bone mass in mice by the lipoxygenase gene Alox15. *Science* 303(5655):229-232.

19. Yagil C, *et al.* (2005) Identification of hypertension-related genes through an integrated genomic-transcriptomic approach. *Circ Res* 96(6):617-625.

20. Hsu YH, *et al.* (2010) An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility Loci for osteoporosis-related traits. *PLoS Genet* 6(6):e1000977.

21. Zhong H, *et al.* (2010) Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet* 6:e1000932.

22. Derry JM, *et al.* (2010) Identification of genes and networks driving cardiovascular and metabolic phenotypes in a mouse F2 intercross. *PLoS One* 5(12):e14319.

23. Perou CM (2001) Show me the data! *Nat Genet* 29(4):373.

24. Paylor R (2009) Questioning standardization in science. *Nat Methods* 6(4):253-254.

25. Nomiyama T, *et al.* (2007) Osteopontin mediates obesity-induced adipose tissue macrophage infiltration and insulin resistance in mice. *J Clin Invest* 117(10):2877-2888.

26. Surwit RS, Kuhn CM, Cochrane C, McCubbin JA, & Feinglos MN (1988) Diet-induced type II diabetes in C57BL/6J mice. *Diabetes* 37(9):1163-1167.

27. Hutas G, *et al.* (2008) CD44-specific antibody treatment and CD44 deficiency exert distinct effects on leukocyte recruitment in experimental arthritis. *Blood* 112(13):4999-5006.

28. Mikecz K, Brennan FR, Kim JH, & Glant TT (1995) Anti-CD44 treatment abrogates tissue oedema and leukocyte infiltration in murine arthritis. *Nat Med* 1(6):558-563.

29. Katoh S, *et al.* (2003) A role for CD44 in an antigen-induced murine model of pulmonary eosinophilia. *J Clin Invest* 111(10):1563-1570.

30. Brocke S, Piercy C, Steinman L, Weissman IL, & Veromaa T (1999) Antibodies to CD44 and integrin alpha4, but not L-selectin, prevent central nervous system

inflammation and experimental encephalomyelitis by blocking secondary leukocyte recruitment. *Proc Natl Acad Sci U S A* 96(12):6896-6901.

31. Alberti KG & Zimmet PZ (1998) Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. *Diabet Med* 15(7):539-553.

32. Hanis CL*, et al.* (1996) A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nat Genet* 13(2):161-166.

33. Mahtani MM*, et al.* (1996) Mapping of a gene for type 2 diabetes associated with an insulin secretion defect by a genome scan in Finnish families. *Nat Genet* 14(1):90-94.

34. Ghosh S*, et al.* (1999) Type 2 diabetes: evidence for linkage on chromosome 20 in 716 Finnish affected sib pairs. *Proc Natl Acad Sci U S A* 96(5):2198-2203.

35. Watanabe RM*, et al.* (2000) The Finland-United States investigation of non-insulin-dependent diabetes mellitus genetics (FUSION) study. II. An autosomal genome scan for diabetes-related quantitative-trait loci. *Am J Hum Genet* 67(5):1186-1200.

36. Nishimura S*, et al.* (2009) CD8+ effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. *Nat Med* 15(8):914-920.

37. Zeyda M*, et al.* (2011) Osteopontin Is an Activator of Human Adipose Tissue Macrophages and Directly Affects Adipocyte Function. *Endocrinology* .

38. Tusher VG, Tibshirani R, & Chu G (2001) Significance analysis of microarrays

applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98(9):5116-5121.

39. Chen R, Li L, & Butte AJ (2007) AILUN: reannotating gene expression data automatically. *Nat Methods* 4(11):879.

40. Whitlock MC (2005) Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J Evol Biol* 18(5):1368-1373.

41. Seino Y*, et al.* (2010) Report of the Committee on the Classification and Diagnostic Criteria of Diabetes Mellitus. *J Diabetes Invest* 1(5):212-228.

42. Matthews DR*, et al.* (1985) Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 28(7):412-419.

FIGURE LEGENDS

Fig. 1: Expression-based genome-wide association study (eGWAS) for T2D using a χ^2 analysis.

Plot of -$\log_{10}$ (*P* value) (y-axis) by chromosomal position (x-axis). *P* values for each gene were calculated from our eGWAS across 130 microarray experiments with 1,175 T2D case-control microarray samples (591 T2D cases and 584 controls) as the likelihood of finding repeated differential expression compared to expected using a χ^2 analysis, or a Fisher's exact test (**Fig. S3**). Our top gene, *CD44*, showed a significant differential expression in 78 experiments ($P = 8.5 \times 10^{-20}$). The red line indicates the Bonferroni threshold ($P = 2.0 \times 10^{-6}$). The green dots indicate several well known T2D-susceptibility genes.

Fig. 2: CD44 expression in adipose tissue of obese mice.

(*A*) *CD44* mRNA expression levels in epididymal adipose tissues in C57BL/6J mice fed either a NFD or HFD (n = 8/group). **(*B*)** CD44 immunoreactivity (DAB chromogen; brown) in epididymal adipose tissues from C57BL/6J mice fed a HFD. Sections were counterstained with hematoxylin (blue). **(*C*)** *CD44* and *SPP1* gene expression profiles in the HFD (n = 8; circles) and NFD (n = 8; triangles) groups. Correlation between *CD44* and *SPP1* mRNA expression in the HFD group (circles) was analyzed using the Pearson's correlation test. Gene expression was monitored using real-time RT-PCR and normalized to expression of *GAPDH* mRNA.

Fig. 3: Histological and metabolic analyses of wild-type *CD44* $^{+/+}$ and *CD44* $^{-/-}$ mice.

(*A*) Inflammatory cell (macrophage) content determined by immunohistochemical staining for Mac-2 (DAB, brown; hematoxylin, blue) in epididymal adipose tissues from *CD44* $^{-/-}$ and *CD44* $^{+/+}$ mice fed a HFD. **(*B-D*)** Metabolic measurements on *CD44* $^{+/+}$ (open bars and symbols; diabetes-prone) and *CD44* $^{-/-}$ (black bars and symbols) mice fed either a HFD (n = 16/group; solid lines) or a NFD (n = 10/group; dashed lines). **(*B*)** Fasting blood glucose. **(*C*)** Glucose tolerance tests (intraperitoneal glucose (2 g/kg body weight)) after a 14-hour overnight fast. Venous blood was obtained for measurement of blood glucose at 0, 15, 30, 60, 90 and 120 minutes after the injection. **(*D*)** Insulin tolerance tests (intraperitoneal insulin (1.0 unit/kg body weight)) after a 4-hour fast. Venous blood was obtained for measurement of blood glucose at 0, 30, 45 minutes after the injection.

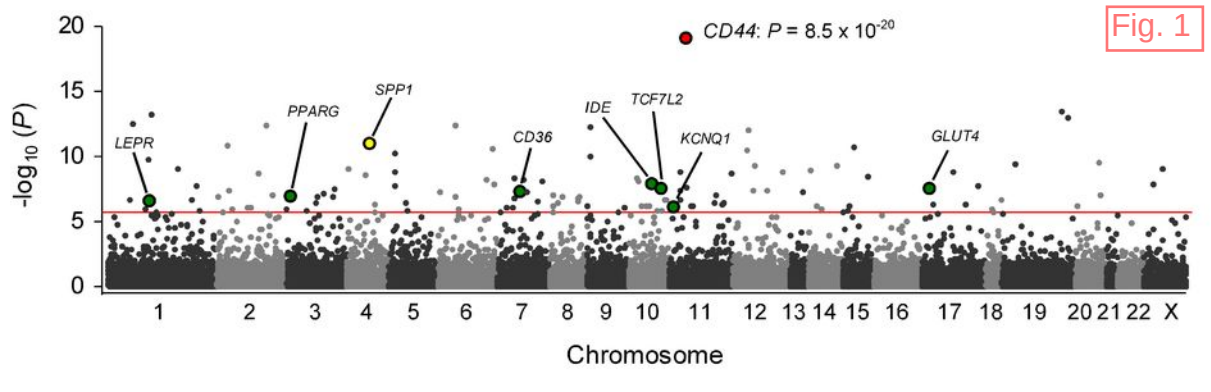
Fig. 4: Anti-CD44 antibody treatment for diabetic mice.

(*A*) HFD fed C57BL/6J mice were injected intraperitoneally with purified rat anti-mouse CD44 (IM7; 553131, BD Pharmingen) (n = 8; black filled circles) or purified rat IgG2b, κ isotype control (A95-1; 559478, BD Pharmingen) (n = 8; open circles) for one week (100 μg at day 0 and 50 μg at day 1-7). Morning blood glucose was measured at day 0, 1, 3, 5 and 7 during the treatment. The effect of anti-CD44 treatment on blood glucose levels was evaluated with two-way repeated measures ANOVA (*P; treatment × time). Comparisons between two groups were performed using t-test. Data are represented as mean ± standard error. **(*B*)** Epididymal adipose tissues from control and anti-CD44 antibody-treated mice were analyzed for inflammatory cell (macrophage) content using a Mac-2 antibody (magnified as indicated).

Fig. 5: CD44 functional experiments in human subjects.

(*A*) Paraffin-embedded omental adipose tissue from an obese woman (age (yr); 57, BMI (kg/m^2); 36.9) was analyzed for CD44 immunoreactivity. **(*B* and *C*)** The correlation between serum levels of standard soluble CD44 (sCD44std) and either an index of glycemic control; HbA1c (***B***) or an index of insulin resistance; HOMA-IR (***C***), determined by using a linear regression model estimated with minimal square method in human subjects (n = 55: mean ± SD age (yr); 60.3 ± 15, BMI (kg/m^2); 23.2 ± 4.3, fasting plasma glucose (mg/dl); 109 ± 13, fasting plasma insulin (μU/ml); 6.22 ± 3.84, HbA1c (%); 5.9 ± 0.34).

Fig. 1


CD44: P = 8.5 x 10-20
LEPR
PPARG
SPP1
CD36
IDE
TCF7L2
KCNQ1
GLUT4
-log10 (P)
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosome

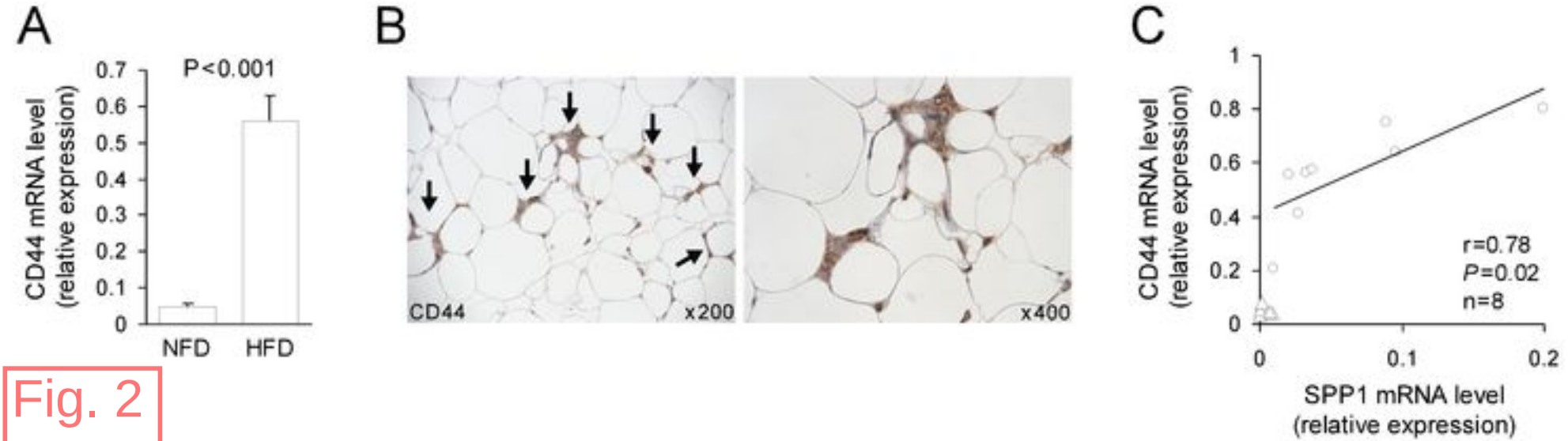

A
P<0.001
CD44 mRNA level
(relative expression)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
NFD
HFD
B
CD44
x200
x400
C
CD44 mRNA level
(relative expression)
1
0.8
0.6
0.4
0.2
0
r=0.78
P=0.02
n=8
0
0.1
0.2
SPP1 mRNA level
(relative expression)

Fig. 2

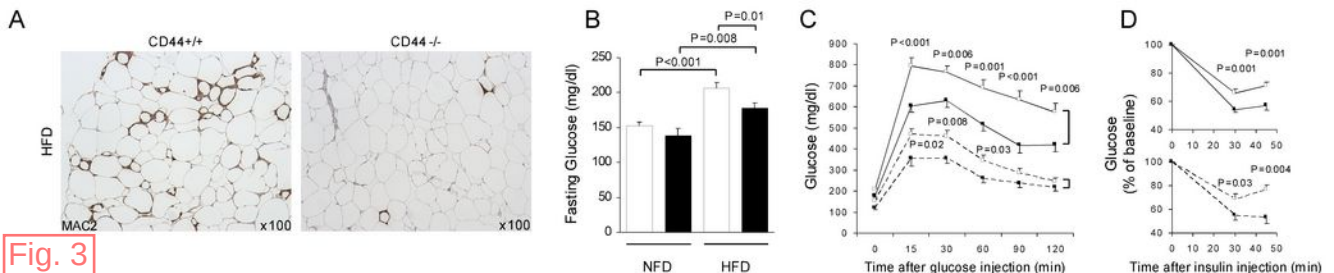
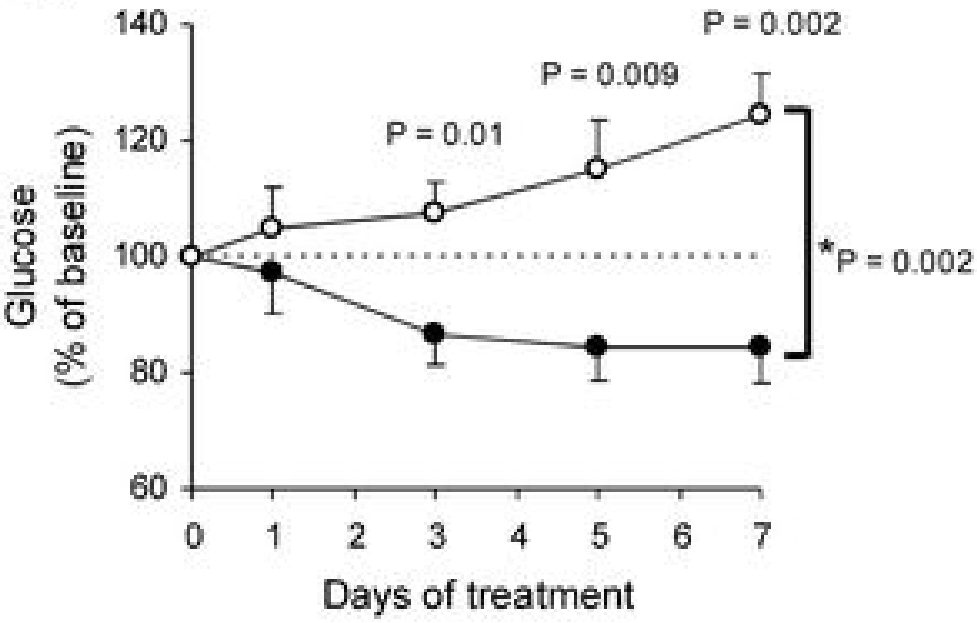

A
CD44+/+
CD44 -/-
HFD
MAC2
x100
x100
B
P=0.01
P=0.008
P<0.001
Fasting Glucose (mg/dl)
250
200
150
100
50
0
NFD
HFD
C
P<0.001
P=0.006
P=0.001
P<0.001
P=0.006
P=0.008
P=0.02
P=0.03
Glucose (mg/dl)
900
800
700
600
500
400
300
200
100
0
0
15
30
60
90
120
Time after glucose injection (min)
D
P=0.001
P=0.001
P=0.004
P=0.03
Glucose (% of baseline)
100
80
60
40
0
10
20
30
40
50
Time after insulin injection (min)

Fig. 3

Fig. 4

A


140
120
100
80
60
Glucose
(% of baseline)
P = 0.002
P = 0.009
P = 0.01
*P = 0.002
0
1
2
3
4
5
6
7
Days of treatment

B

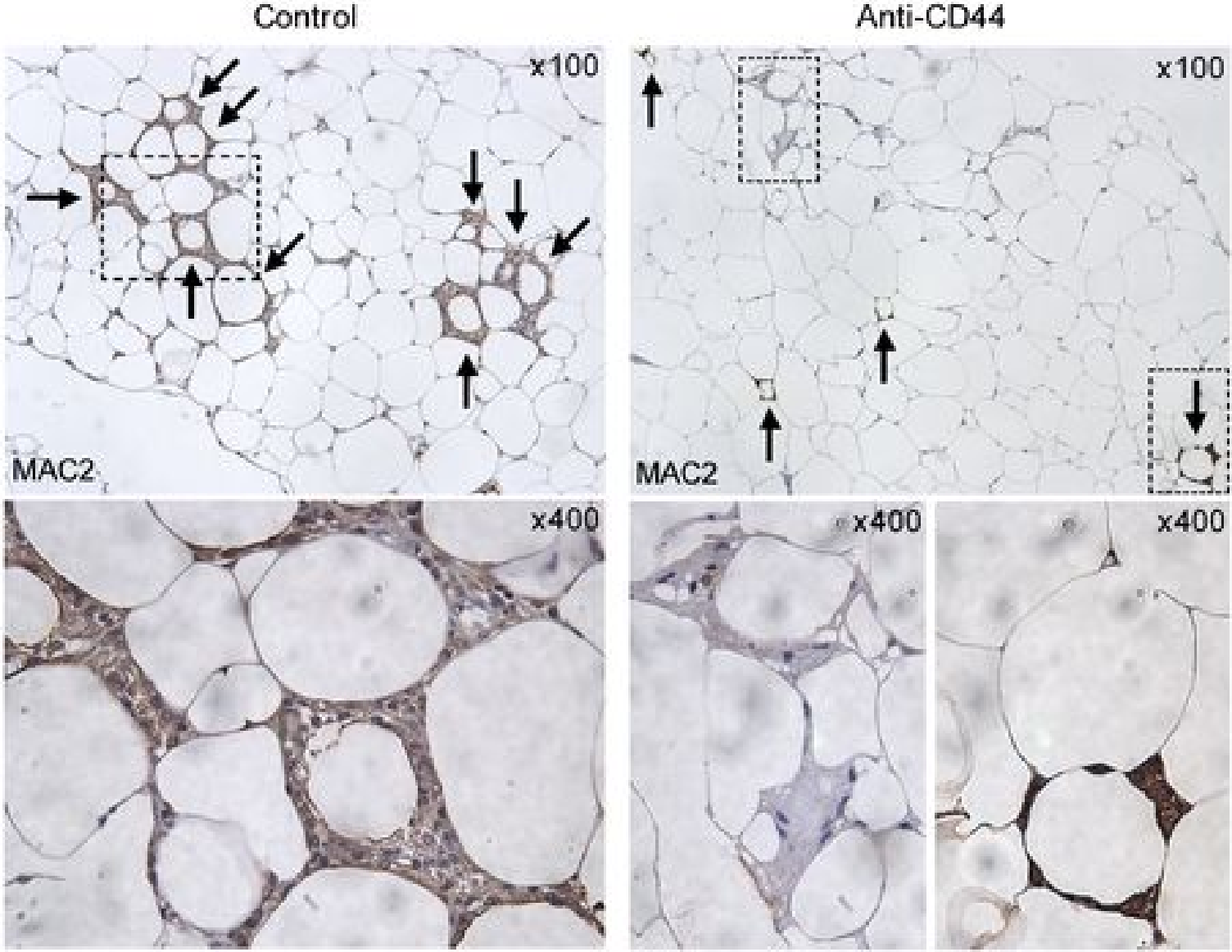
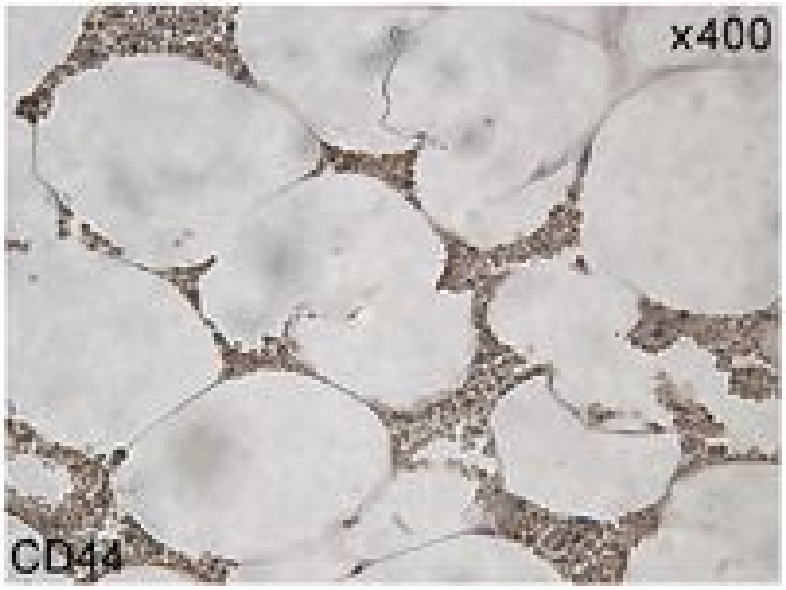
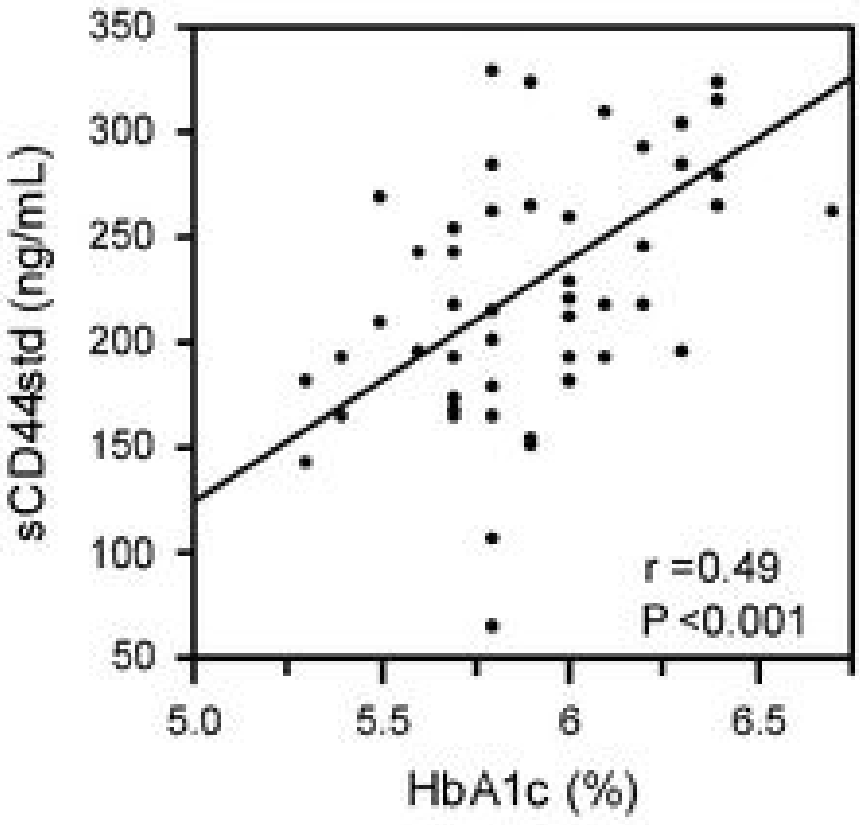

Control
Anti-CD44
x100
x100
MAC2
MAC2
x400
x400
x400

Fig. 5

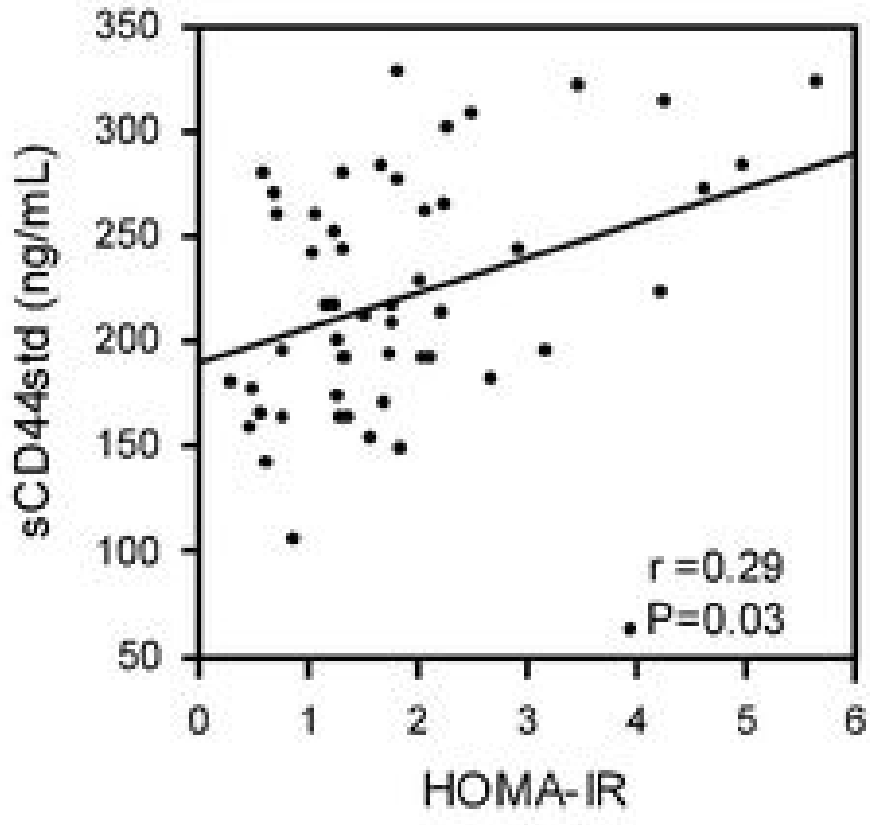
A


x400
CD44

B


sCD44std (ng/mL)
350
300
250
200
150
100
50
5.0
5.5
6
6.5
HbA1c (%)
r =0.49
P <0.001

C


sCD44std (ng/mL)
350
300
250
200
150
100
50
0
1
2
3
4
5
6
HOMA-IR
r =0.29
P=0.03

Supporting information (SI)

SI Methods

Figures:

Fig. S1: Study Design.

Fig. S2: Distribution of Gene Ontology (GO) molecular function annotations for the 127 top-ranked genes in our T2D eGWAS (Bonferroni threshold, $P < 2.0 \times 10^{-6}$).

Fig. S3: Expression-based genome-wide association study (eGWAS) for T2D using a Fisher's exact test.

Fig. S4: Liptak-Stouffer's weighted *Z*-method for T2D case-control microarray experiments.

Fig. S5: The distribution of fold-change for *CD44* across all the experiments in T2D eGWAS.

Fig. S6: Coordinate dysregulation rate between *CD44* and *SPP1*.

Tables:

Table S1: List of microarray experiments in our T2D eGWAS.

Table S2: List of microarray platforms in the 130 experiments.

Table S3: List of top 127 genes in our T2D eGWAS (Bonferroni threshold, $P < 2.0 \times 10^{-6}$).

SI Methods

Expression-based genome-wide association study (eGWAS).

All T2D-related genome-wide microarray experiments used for this meta-analysis were collected from three public data sources; The NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/**)** was searched using the keywords: "*diabetes*" OR "*diabetic*" OR "*NIDDM*" OR "*non-insulin-dependent*" (from inception until July 2009). From the list of all identified GEO Series (GSE), we selected gene expression microarray studies that met the following criteria: *1*) investigating about "type 2 diabetes" (Studies related to type 1 diabetes, maturity-onset diabetes of the young, drug-induced diabetes, gestational diabetes, diabetic complications and other specific diseases causing diabetes were excluded from our data analysis); *2*) using T2D-relevant tissues, such as adipose, liver, muscle, pancreatic islets, kidney and hypothalamus; *3*) using samples from human and rodent models. We also selected microarray studies involving T2D from the supplied list in the Diabetes Genome Anatomy Project (DGAP; http://www.diabetesgenome.org/) and the Nuclear Receptor Signaling Atlas (NURSA; http://www.nursa.org/) on April 2008. We then reassembled microarray samples in the selected studies into independent data set comparisons (experiments): a curated collection of pathophysiologically-comparable microarray samples, always comparing diabetes samples to pathophysiologically matched control samples. This study selection and sample assembly were performed manually by two independent investigators by reviewing the submitter-supplied record and the abstract and full-text of original research articles. Discrepancies in eligibility were discussed between reviewers until agreement was achieved. We subsequently downloaded processed (normalized by the original

methods selected by the original submitters) microarray data corresponding to the identified experiments from the three public repositories. Data were stored in a Microsoft Excel 2007 spreadsheet (Microsoft Corporation, Redmond, WA). Execution of this procedure yielded 130 independent experiments with a total of 1,175 samples (591 diabetic cases and 584 non-diabetic controls) (**Table S1 and S2**).

To estimate differences between groups of samples from diabetic subjects and groups representing control, the downloaded microarray data were re-analyzed using Significance Analysis of Microarrays software (SAM Excel Add-in; http://www-stat.stanford.edu/~tibs/SAM/) with "two class unpaired" design (or with "paired" or "one class" option if necessary) when each group had three or more samples (1). For each gene in every microarray experiment, we calculated a D-score (d_i), which denotes the standardized change in gene expression:

$$d_i = \frac{\bar{x}_{i-t2d} - \bar{x}_{i-control}}{S_i + S_0}$$

where $\bar{x}_{i-t2d}$ is the mean expression level of *gene i* in group T2D, $\bar{x}_{i-control}$ is the mean expression level of *gene i* in group control, S_i is the standard deviation for the numerator calculation, and S_0 is a small positive constant, and also calculated fold change as the ratio between the signal averages of control and experimental samples. When each group had only one or two samples, we calculated only a fold-change value for each gene in these microarray experiments.

We considered genes to be significantly dysregulated with either an absolute value of the D-score ≥ 2 or a fold change ≥ 2 between controls and cases (ref; "SAM": Users guide and technical documents; http://www-stat.stanford.edu/~tibs/SAM/sam.pdf) (2).

We then converted all probe identifiers across the various microarray platforms for mouse, rat and human to the latest human Entrez Gene identifiers using our previously published AILUN system (http://ailun.stanford.edu/; its Cross-species Mapping uses the NCBI Homologene resource) (3). Gene expression profiles were assigned in our eGWAS database (Excel spreadsheet format) according to the standardized (human) Entrez Gene ID. There were 24,898 genes in the database in total.

For each of the 24,898 genes, we counted the observed number of microarray experiments in which each gene was significantly dysregulated, using the Excel 'pivot table' function. We then aligned the number of positive/negative experiments for every one of genes, summed the number of positive/negative experiments for all other genes and calculated chi-square formula using these data in the one row of Excel spreadsheet, and ran the chi-square test to calculate χ^2 value and its corresponding P value for all 24,898 genes (rows). As an alternative methodology, we also conducted a Fisher's exact test using the same data of contingency tables for all the genes, using the R statistical package (http://www.r-project.org/)(4). We subsequently ranked all the genes according to their P values ($-\log_{10}(P)$). We also conducted a weighted Z-method (5) using the same experiments in our eGWAS database, and confirmed that *CD44* is still the top gene in this alternative method based on a different concept.

Animal experiments.

Mice for breeding, C57BL/6J wild-type (diabetes-prone) and *CD44*-deficient mice backcrossed to C57BL/6J for at least 10 generations (B6.Cg-*Cd44*^{tm1Hbg}/J), were obtained from The Jackson Laboratory. To examine the role of CD44 for the development of insulin resistance, male *CD44* $^{-/-}$ and littermate wild type *CD44* $^{+/+}$ mice

(8 weeks of age) were fed diets containing either 12% kcal fat (normal-fat diet; NFD) (CE-2; CLEA Japan, Inc.) or 60% kcal fat (high-fat diet; HFD) (D12492; Research Diets Inc.) for 12 weeks. To assess CD44 and SPP1 expression in obese adipose tissue, C57BL/6J male mice (8 weeks of age) were fed either a NFD or a HFD for 16 weeks. To investigate the therapeutic effect of Anti-CD44 antibody on HFD-induced diabetes, C57BL/6J male mice (8 weeks of age) were fed a HFD for 18 weeks. Weight gain was monitored by weighing mice weekly. Mice had free access to autoclaved water. Mice were housed in a barrier facility under specific pathogen-free conditions. The Animal Care and Use Committee of Kitasato University (Tokyo, Japan) approved all animal experiments.

Metabolic measurements were conducted on male *CD44* $^{-/-}$ and littermate wild type *CD44* $^{+/+}$ mice after feeding a NFD or HFD for 12 weeks. Fasting blood glucose levels were measured after a 14-hour overnight fast. Glucose tolerance tests (GTT) were performed by giving glucose (2 g/kg of body weight) intraperitoneally after a 14-hour overnight fast. Venous blood was obtained for measurement of blood glucose and serum insulin levels at 0, 15, 30, 60, 90 and 120 minutes after the injection. Insulin tolerance tests (ITT) were performed by giving insulin (1.0 unit/kg; Actrapid, Novo Nordisk, Bagsvaerd, Denmark) intraperitoneally after a 4-hour fast. Venous blood was obtained for measurement of blood glucose at 0, 30, 45 minutes after the injection. Blood glucose concentration was determined with a glucose meter (Medisafe-Mini; Terumo, Tokyo, Japan). Serum insulin levels were measured with an ultrasensitive mouse insulin ELISA kit (Morinaga Institute of Biological Science, Kanagawa, Japan).

CD44 and *SPP1* mRNA expression in visceral adipose tissue was measured by quantitative real-time RT-PCR. At 24 weeks of age, epididymal white adipose tissue (EWAT) was removed from wild-type C57BL/6J mice fed either a NFD or HFD, immediately frozen by liquid nitrogen and stored in -80 degree Celsius freezer. Total RNA of EWAT was isolated using the Trizol RNA isolation method (Invitrogen) and purified with the RNeasy Mini Kit spin columns (QIAGEN) according to the manufacturer's instructions. Quantity and quality of isolated RNA was determined by spectrophotometric measurements at optical density (OD) 260 and OD280. The integrity of isolated RNA was checked by agarose gel electrophoresis. Two microgram of RNA was reverse-transcribed to cDNA using a 1st Strand cDNA Synthesis Kit for RT-PCR (AMV) (Roche Diagnostics). PCR reactions were performed with the LightCycler® FastStart DNA master SYBR Green I system (Roche Diagnostics). Each sample was analyzed in triplicate and normalized to the values for glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) mRNA expression. Mouse primer sequences used were as follows: CD44, CCA GGC TTT CAA CAG TAC CTT ACC (forward), CTG AGG CAT TGA AGC AAT ATG TGT C (reverse); SPP1, TGA TAG CTT GGC TTA TGG ACT GA (forward), CCA CTG AAC TGA GAA ATG AGC AG (reverse); GAPDH, TGA ACG GGA AGC TCA CTG G (forward), TCC ACC ACC CTG TTG CTG TA (reverse).

Anti-CD44 antibody treatment was performed on C57BL/6J male mice fed a HFD for 18 weeks. Mice were injected intraperitoneally with purified rat anti-mouse CD44 (IM7; 553131, BD Pharmingen) or purified rat IgG2b, κ isotype control (A95-1; 559478, BD Pharmingen) for 8 days (100 μg at day 0 and 50 μg at day 1-7). Morning blood glucose was measured at day 0, 1, 3, 5 and 7 during the treatment.

Human studies.

Venous peripheral blood samples were collected from human subjects who went through a 75g oral glucose tolerance test after an overnight fast (n=55: age (yr); 60.3 ± 15, gender (M/F); 36/19, BMI (kg/m^2); 23.2 ± 4.3). HbA1c (glycosylated hemoglobin) was measured in Japan Diabetes Society (JDS)-HbA1c units using an ion-exchange high-performance liquid chromatography method. HbA1c was converted to National Glycohemoglobin Standardization Program (NGSP) levels by the formula HbA1c (%) (NGSP) = HbA1c (JDS) (%) + 0.4%, considering the relational expression of HbA1c (JDS) (%) measured by the previous Japanese standard substance and measurement methods and HbA1c (NGSP) (6), and these converted values were used throughout the study. We then calculated homeostasis model assessment as an index of insulin resistance (HOMA-IR = fasting plasma insulin (μU/ml) × fasting plasma glucose (mg/dl) / 405) as previously described (7). Serum sCD44std (standard soluble CD44) and SPP1 concentrations were determined using a quantitative enzyme-linked immunosorbent assay (ELISA) technique (sCD44std ELISA; Bender MedSystems, Vienna, Austria, Human Osteopontin Quantikine ELISA; R&D Systems, Minneapolis, MN).

Informed consent was obtained from all of the subjects enrolled in this study and the protocol was approved by the ethics committee of the University of Tokyo.

Immunohistochemistry.

For histological analysis of CD44 expression in adipose tissue, EWAT was removed from mouse models, and omental adipose tissue obtained from consented donors undergoing elective gastric bypass surgery (lot number. OM020304B) was purchased from Zen-Bio, Inc. (Research Triangle Park, NC). Formalin-fixed paraffin-

embedded sections (5 μm thick) were stained for CD44 using the DAKO CSA II signal amplification system (DAKO). After a peroxidase and protein block, the slides were incubated overnight (4 degrees Celsius) with mouse monoclonal antibody against CD44 at 1:50 dilution (DF1485/sc-7297; Santa Cruz Biotechnology, Inc.), followed by reactions with anti-mouse immunoglobulins-HRP, an amplification reagent, and anti-fluorescein-HRP. 3,3'-Diaminobenzidine (DAB) was used as a chromogen, and the sections were counterstained with hematoxylin.

Analysis of inflammatory cell (macrophage) content in EWAT was performed on tissue pads isolated from model mice. Formalin-fixed paraffin-embedded sections were incubated overnight (4 degrees Celsius) with primary antibody: Purified Anti Mouse MAC-2 (macrophage marker) Monoclonal Antibody (CL8942AP, 1:100; Cedarlane laboratories, Ontario, Canada), and stained using Histofine® Simple Stain Mouse MAX-PO (R) secondary antibody (Nichirei Biosciences, Inc., Tokyo, Japan) with a DAB solution and counterstained with hematoxylin.

In anti-CD44 antibody-treated and control mice, adipose inflammation was quantified as the density of crown-like structures (CLSs). The total number of CLSs was counted in 5 random fields (magnification ×100) of each mouse in a blinded manner, and the average number of CLSs was calculated in all animals in each group, by creating a digitized image with a BIOQUANT Image Analysis System (BIOQUANT Image Analysis Corp.).

REFERENCES

1. Tusher VG, Tibshirani R, & Chu G (2001) Significance analysis of microarrays

applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98(9):5116-5121.

2. Orlova MO*, et al.* (2006) Constitutive differences in gene expression profiles parallel genetic patterns of susceptibility to tuberculosis in mice. *Infect Immun* 74(6):3668-3672.
3. Chen R, Li L, & Butte AJ (2007) AILUN: reannotating gene expression data automatically. *Nat Methods* 4(11):879.
4. R Development Core Team (2004) R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria), 1.9.1.
5. Whitlock MC (2005) Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J Evol Biol* 18(5):1368-1373.
6. Seino Y*, et al.* (2010) Report of the Committee on the Classification and Diagnostic Criteria of Diabetes Mellitus. *J Diabetes Invest* 1(5):212-228.
7. Matthews DR*, et al.* (1985) Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* 28(7):412-419.

SI FIGURE LEGENDS

Fig. S1: Study Design.

Expression-based GWAS (eGWAS) for type 2 diabetes (T2D) was carried out in 1,175 microarray samples collected from public databases. *P* values ($-\log_{10}(P)$) were calculated by comparing dysregulation distribution of genes between T2D and control microarrays. Our T2D candidate gene extracted from eGWAS was verified by functional studies in mouse models and human subjects.

Fig. S2: Distribution of Gene Ontology (GO) molecular function annotations for the 127 top-ranked genes in our T2D eGWAS (Bonferroni threshold, $P < 2.0 \times 10^{-6}$).

Fig. S3: Expression-based genome-wide association study (eGWAS) for T2D using a Fisher's exact test.

Plot of $-\log_{10}$ (*P* value) (y-axis) by chromosomal position (x-axis). *P* values for each gene were calculated using a Fisher's exact test. *CD44* is the top gene in both the Fisher's exact test and the χ^2 analysis (**Fig. 1**). The red line indicates the Bonferroni threshold ($P = 2.0 \times 10^{-6}$). The green dots indicate the several well known T2D-susceptibility genes significant in the χ^2 analysis (**Fig. 1**). All the well-known genes from **Fig. 1** except *KCNQ1* were still significant in the Fisher's exact test ($P < 2.0 \times 10^{-6}$).

Fig. S4: Liptak-Stouffer's weighted Z-method for T2D case-control microarray experiments.

The frequency distribution of $-\log_{10}$ (combined *P* value). The combined *P* values for each gene were calculated using a weighted Z-method; *P* values were computed using a one-tailed t-test for each gene in each of the 110 experiments with four or more samples,

and then the *P* values were converted to Z-scores. Subsequently, the combined Z-scores (Zw) across all the experiments were calculated for each gene, using a weighted Z-method, by weighting each experiment by its sample size (degrees of freedom; d.f.). Then the combined *P*-values for each gene were obtained by converting the weighted Z-scores (Zw) into two-tailed *P* values. The $-\log_{10}$ (combined *P* values) were rounded into the nearest integer, and then the frequency distribution was determined. We confirmed that *CD44* is still the top gene in this alternative method (Zw = 17.48, $P = 2.0 \times 10^{-68}$).

Fig. S5: The distribution of fold change for *CD44* across all the experiments in T2D eGWAS.

Plot of $\log_2$ (fold change) for *CD44* (y-axis; error bar = 95% confidence intervals) in microarray experiments (x-axis; abbreviations: A, adipose tissue; i, pancreatic islets; L, liver; M, muscle; K, kidney; H, hypothalamus). The area of the box is proportional to the sample size of each experiment. The red line indicates the significance threshold (fold change = 2). *CD44* mRNA was the most highly differentially expressed gene across experiments, and was also the top-most highly upregulated mRNA in diabetes. *CD44* mRNA was more highly differentially expressed in adipose tissue than other tissues.

Fig. S6: Coordinate dysregulation rate between *CD44* and *SPP1*.

The Coordinate Dysregulation Rate (CDR) is calculated as the probability of *SPP1* dysregulation, given the occurrence of *CD44* dysregulation:

$$CDR = P(SPP1 \mid CD44) = \frac{n(SPP1 \cap CD44)}{n(CD44)}$$

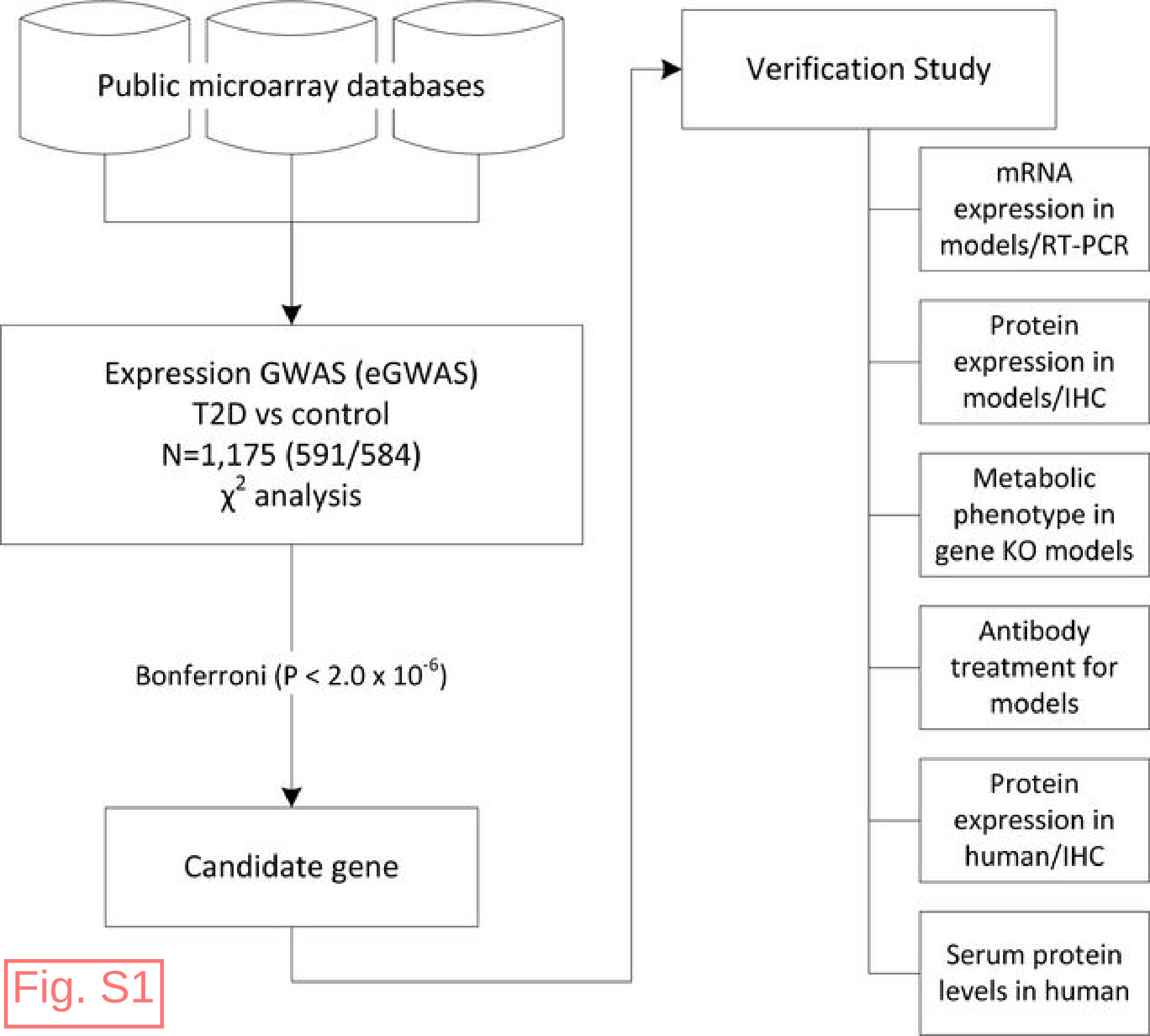

Public microarray databases
Expression GWAS (eGWAS)
T2D vs control
N=1,175 (591/584)
χ^2 analysis
Bonferroni ($P < 2.0 \times 10^{-6}$)
Candidate gene
Verification Study
mRNA expression in models/RT-PCR
Protein expression in models/IHC
Metabolic phenotype in gene KO models
Antibody treatment for models
Protein expression in human/IHC
Serum protein levels in human

Fig. S1

Fig. S2

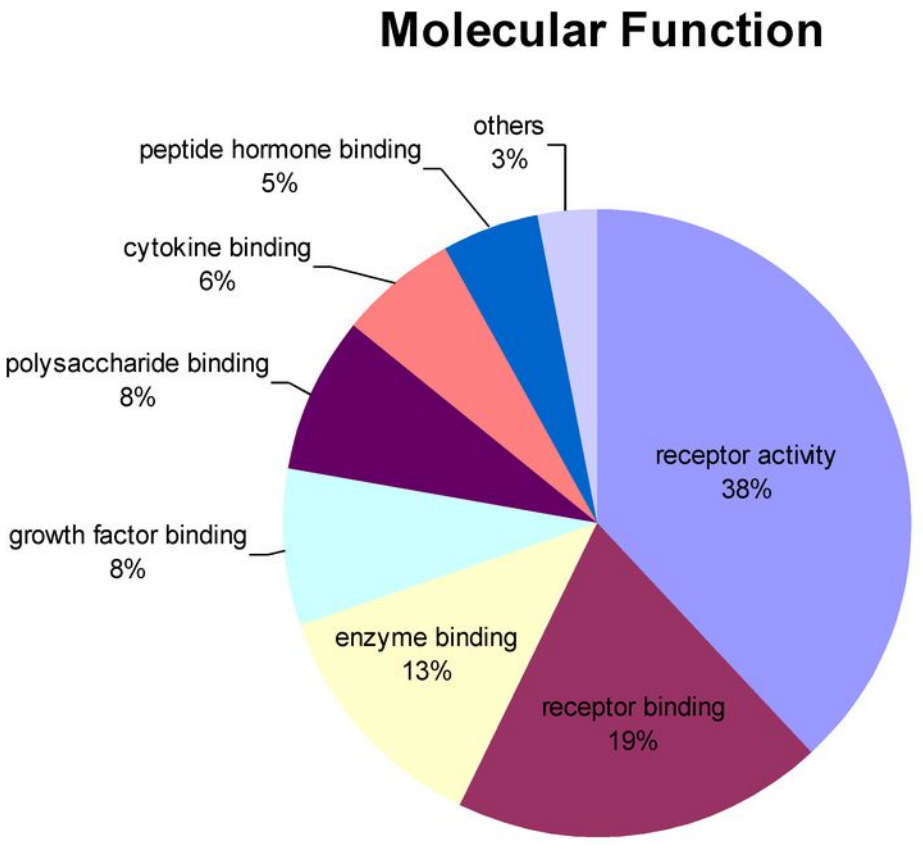

Molecular Function
others
3%
peptide hormone binding
5%
cytokine binding
6%
polysaccharide binding
8%
growth factor binding
8%
receptor activity
38%
enzyme binding
13%
receptor binding
19%

Fig. S3

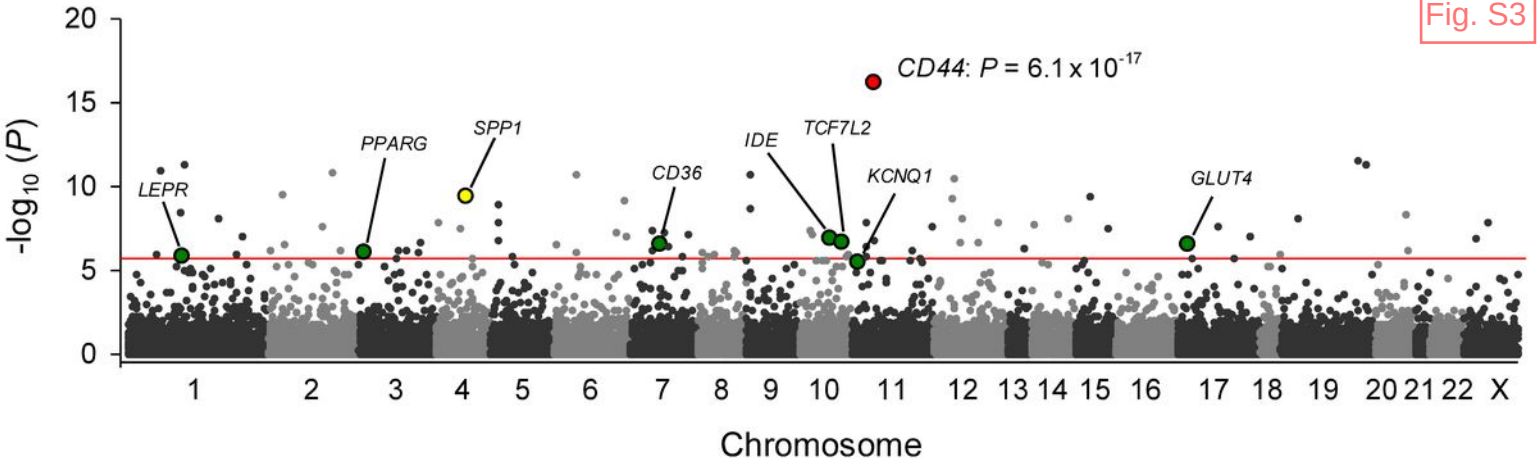

20
15
10
5
0
-log10 (P)
CD44: P = 6.1 x 10-17
LEPR
PPARG
SPP1
CD36
IDE
TCF7L2
KCNQ1
GLUT4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Chromosome

Fig. S4

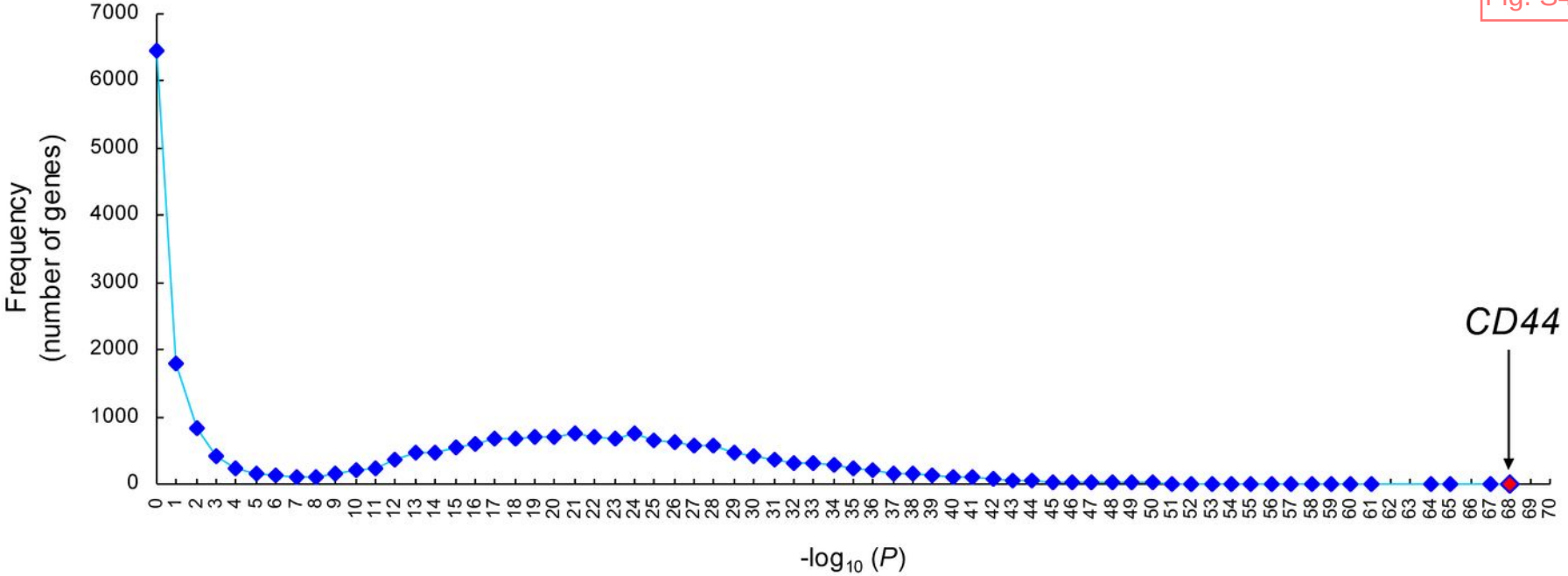

Frequency
(number of genes)
7000
6000
5000
4000
3000
2000
1000
0
CD44
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70
-log10 (P)

Fig. S5

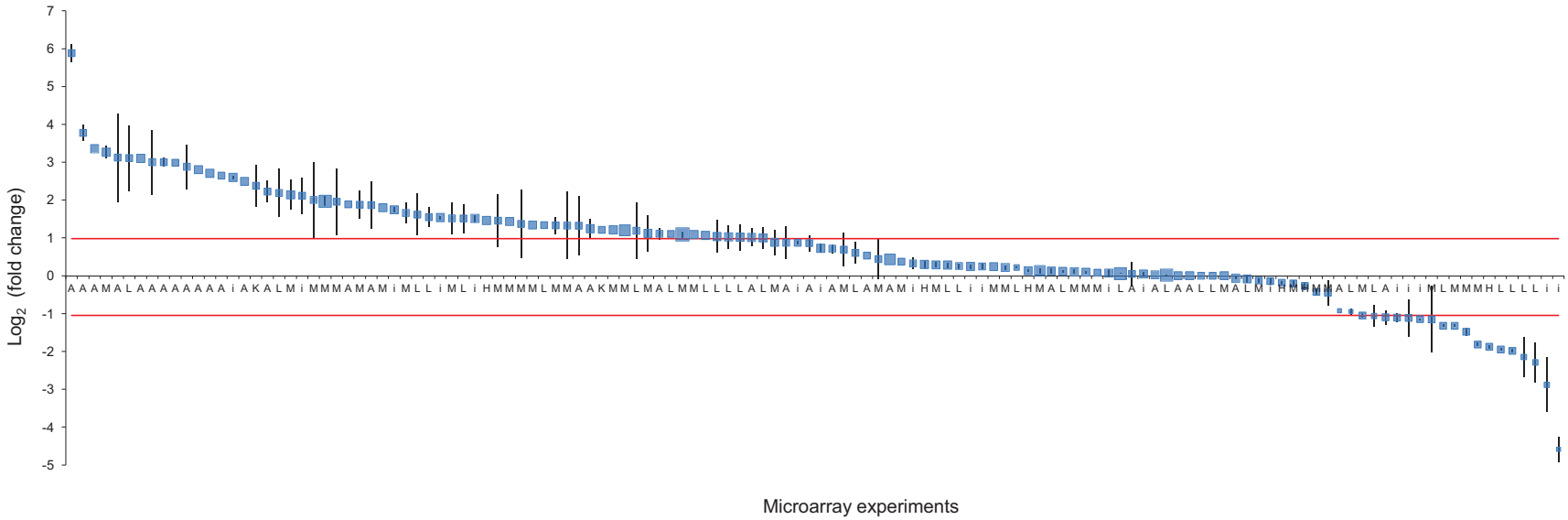


Fig. S6

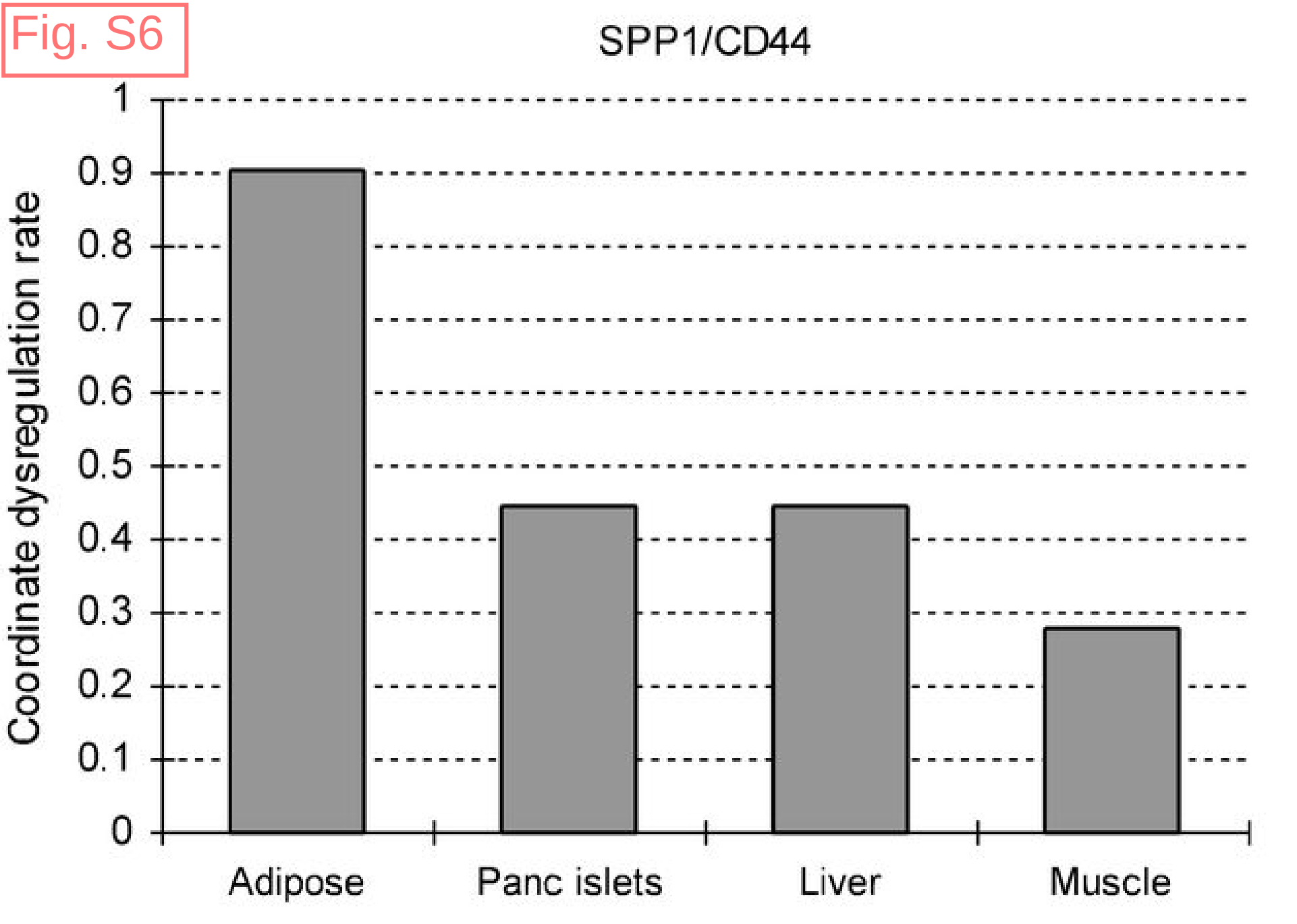

SPP1/CD44
Coordinate dysregulation rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Adipose
Panc islets
Liver
Muscle

Table S1. List of microarray experiments in our T2D eGWAS

Data source	Sample size	Tissue	Species	Experimental model (diabetic)	Control (non-diabetic, anti-diabetic)	Laboratory
GEO	10	Adipose	Human	Diabetes	NGT	Mathur SK
GEO	20	Adipose	Human	Before bariatric surgery	After bariatric surgery	Clement K
GEO	12	Adipose	Human	Stroma vascular fraction	Mature adipocytes	Clement K
GEO	16	Adipose	Human	Adipocytes Tx with activated macrophage	Adipocytes Tx with control medium	Clement K
GEO	10	Adipose	Mouse	BTBR ob/ob diabetes 4-wk-old	BTBR lean NGT 4-wk-old	Attie AD
GEO	10	Adipose	Mouse	BTBR ob/ob diabetes 10-wk-old	BTBR lean NGT 10-wk-old	Attie AD
GEO	2	Adipose	Mouse	BTBR ob/ob diabetes 14-wk-old	BTBR lean NGT 14-wk-old	Attie AD
GEO	10	Adipose	Mouse	B6 ob/ob diabetes 4-wk-old	B6 lean NGT 4-wk-old	Attie AD
GEO	10	Adipose	Mouse	B6 ob/ob diabetes 10-wk-old	B6 lean NGT 10-wk-old	Attie AD
GEO	2	Adipose	Mouse	B6 ob/ob diabetes 14-wk-old	B6 lean NGT 14-wk-old	Attie AD
DGAP	6	Adipose	Mouse	129S6 high fat diet	129S6 control diet	Kahn CR
DGAP	8	Adipose	Mouse	IRS1 KO	WT	Kahn CR
DGAP	8	Adipose	Mouse	IRS3 KO	WT	Kahn CR
DGAP	9	Adipose	Mouse	B6 x 129/Sv large adipocytes	IR KO large adipocytes	Kahn CR
DGAP	9	Adipose	Mouse	B6 x 129/Sv small adipocytes	IR KO small adipocytes	Kahn CR
GEO	2	Adipose	Mouse	B6	B6 Tx with Adrb3 agonist---day 1	Lu Y
GEO	2	Adipose	Mouse	B6	B6 Tx with Adrb3 agonist---day 3	Lu Y
GEO	2	Adipose	Mouse	B6	B6 Tx with Adrb3 agonist---day 6	Lu Y

GEO	11	Adipose	Mouse	Adipocytes Tx with (high) glucose + insulin	Adipocytes Tx with (high) glucose + insulin + Thiazolidinediones	Subramaniam S
GEO	10	Adipose	Mouse	Adipocytes Tx with glucose + insulin	Adipocytes Tx with glucose + insulin + Thiazolidinediones	Keijer J
GEO	10	Adipose	Mouse	Adipocytes Tx with (high) glucose + insulin	Adipocytes Tx with (high) glucose + insulin + Thiazolidinediones	Keijer J
DGAP	4	Adipose	Mouse	Adipocytes Tx with (high) glucose + insulin---day 2	Adipocytes basal	Czech MP
DGAP	4	Adipose	Mouse	Adipocytes Tx with (high) glucose + insulin---day 4	Adipocytes basal	Czech MP
DGAP	4	Adipose	Mouse	Adipocytes Tx with (high) glucose + insulin---day 6 (1)	Adipocytes basal	Czech MP
DGAP	2	Adipose	Mouse	Adipocytes Tx with (high) glucose + insulin---day 6 (2)	Adipocytes basal	Czech MP
DGAP	6	Adipose	Mouse	Adipocytes Tx with (high) glucose + insulin---day 7	Adipocytes basal	Czech MP
GEO	8	Adipose	Mouse	Db/db	Db/db Tx with niacin-bound chromium	Sen CK
GEO	7	Adipose	Mouse	GK KO	WT	Dipple KM
GEO	10	Adipose	Mouse	BTBR ob/ob severe diabetes 4-wk-old	B6 ob/ob mild diabetes 4-wk-old	Attie AD
GEO	10	Adipose	Mouse	BTBR ob/ob severe diabetes 10-wk-old	B6 ob/ob mild diabetes 10-wk-old	Attie AD
GEO	2	Adipose	Mouse	BTBR ob/ob severe diabetes 14-wk-old	B6 ob/ob mild diabetes 14-wk-old	Attie AD
GEO	9	Adipose	Rat	Goto-Kakizaki (GK)	Brown-Norway (BN)	Lindgren CM
GEO	8	Adipose	Rat	ZDF diabets 12-wk-old	ZDF non-diabetes 6-wk-old	Jung MH
GEO	9	Liver	Human	Severe diabetes	NGT	Patti ME
GEO	10	Liver	Human	Diabetes	NGT	Patti ME

GEO	9	Liver	Human	IGT	NGT	Patti ME
GEO	4	Liver	Human	HepG2	HepG2 expressing isocitrate lyase (aceA) and malate synthase (aceB)	Liao JC
GEO	10	Liver	Mouse	BTBR ob/ob diabetes 4-wk-old	BTBR lean NGT 4-wk-old	Attie AD
GEO	10	Liver	Mouse	BTBR ob/ob diabetes 10-wk-old	BTBR lean NGT 10-wk-old	Attie AD
GEO	10	Liver	Mouse	B6 ob/ob diabetes 4-wk-old	B6 lean NGT 4-wk-old	Attie AD
GEO	10	Liver	Mouse	B6 ob/ob diabetes 10-wk-old	B6 lean NGT 10-wk-old	Attie AD
DGAP	8	Liver	Mouse	129S6 high fat diet	129S6 control diet	Kahn CR
DGAP	6	Liver	Mouse	Ob/ob	WT	Kahn CR
GEO	8	Liver	Mouse	B6J high fat diet	B6J control diet	Stephanopoulos G
GEO	8	Liver	Mouse	B6J high fat diet (48-hour fast)	B6J control diet	Stephanopoulos G
DGAP	4	Liver	Mouse	Hepatocytes PGC-1 alpha overexpression	Hepatocytes control	Spiegelman B
DGAP	4	Liver	Mouse	Hepatocytes PGC-1 beta overexpression	Hepatocytes control	Spiegelman B
DGAP	5	Liver	Mouse	B6 (fed)	PGC-1 alpha KO (fed)	Spiegelman B
DGAP	5	Liver	Mouse	B6 (fasted)	PGC-1 alpha KO (fasted)	Spiegelman B
NURSA	4	Liver	Mouse	Ob/ob	SHP KO	Moore DD
NURSA	4	Liver	Mouse	B6×129/Sv	SHP KO	Moore DD
GEO	10	Liver	Mouse	B6J high fat diet	B6J high fat diet JNK1 underexpression	Surowy TK
GEO	2	Liver	Mouse	B6J hepatocytes	B6J hepatocytes BAF60a overexpression	Lin J
GEO	8	Liver	Mouse	B6	AIF KO	Penninger JM

GEO	10	Liver	Mouse	BTBR ob/ob severe diabetes 4-wk-old	B6 ob/ob mild diabetes 4-wk-old	Attie AD
GEO	10	Liver	Mouse	BTBR ob/ob severe diabetes 10-wk-old	B6 ob/ob mild diabetes 10-wk-old	Attie AD
GEO	4	Liver	Mouse	BTBR ob/ob severe diabetes 14-wk-old	B6 ob/ob mild diabetes 14-wk-old	Attie AD
GEO	37	Liver	Mouse	(B6×BTBR) F2-ob/ob severe diabetes GPL339	(B6×BTBR) F2-ob/ob non-diabetes GPL339	Attie AD
GEO	37	Liver	Mouse	(B6×BTBR) F2-ob/ob severe diabetes GPL340	(B6×BTBR) F2-ob/ob non-diabetes GPL340	Attie AD
GEO	4	Liver	Rat	Wistar high fat diet (lard)	Wistar control diet	Bollheimer LC
GEO	4	Liver	Rat	Wistar high fat diet (olive oil)	Wistar control diet	Bollheimer LC
GEO	8	Liver	Rat	Goto-Kakizaki (GK)	Brown-Norway (BN)	Lindgren CM
GEO	8	Liver	Rat	ZDF diabets 12-wk-old	ZDF non-diabetes 6-wk-old	Jung MH
GEO	20	Muscle	Human	Diabetes	NGT	Gaster M
DGAP	35	Muscle	Human	Diabetes	NGT	Altshuler D
DGAP	10	Muscle	Human	Diabetes	NGT family history (-)	Patti ME, Butte AJ
DGAP	9	Muscle	Human	Diabetes	NGT family history (+)	Patti ME, Butte AJ
DGAP	25	Muscle	Human	IGT	NGT	Altshuler D
GEO	10	Muscle	Human	IGT GPL80	NGT GPL80	Permana PA
GEO	10	Muscle	Human	IGT GPL98, 99, 100 and 101	NGT GPL98, 99, 100 and 101	Permana PA
DGAP	10	Muscle	Human	IGT family history (+)	NGT family history (-)	Patti ME
GEO	12	Muscle	Human	Continuous intravenous infusion of glucose/insulin---2 hours	Before infusion	Mootha VK
GEO	12	Muscle	Human	Continuous intravenous infusion of glucose/insulin---3 hours	Before infusion	Mootha VK
GEO	12	Muscle	Human	Continuous intravenous infusion of	Before infusion	Vidal H

				glucose/insulin---3 hours		
GEO	24	Muscle	Human	Continuous intravenous infusion of glucose/insulin---4 hours	Before infusion	Jenkinson CP
GEO	10	Muscle	Mouse	BTBR ob/ob diabetes 4-wk-old; gastroc	BTBR lean NGT 4-wk-old; gastroc	Attie AD
GEO	10	Muscle	Mouse	BTBR ob/ob diabetes 10-wk-old; gastroc	BTBR lean NGT 10-wk-old; gastroc	Attie AD
GEO	10	Muscle	Mouse	B6 ob/ob diabetes 4-wk-old; gastroc	B6 lean NGT 4-wk-old; gastroc	Attie AD
GEO	10	Muscle	Mouse	B6 ob/ob diabetes 10-wk-old; gastroc	B6 lean NGT 10-wk-old; gastroc	Attie AD
GEO	10	Muscle	Mouse	BTBR ob/ob diabetes 4-wk-old; soleus	BTBR lean NGT 4-wk-old; soleus	Attie AD
GEO	10	Muscle	Mouse	BTBR ob/ob diabetes 10-wk-old; soleus	BTBR lean NGT 10-wk-old; soleus	Attie AD
GEO	10	Muscle	Mouse	B6 ob/ob diabetes 4-wk-old; soleus	B6 lean NGT 4-wk-old; soleus	Attie AD
GEO	10	Muscle	Mouse	B6 ob/ob diabetes 10-wk-old; soleus	B6 lean NGT 10-wk-old; soleus	Attie AD
DGAP	8	Muscle	Mouse	129S6 high fat diet	129S6 control diet	Kahn CR
GEO	12	Muscle	Mouse	B6J high fat diet---day 3	B6J control diet	Smit E
GEO	12	Muscle	Mouse	B6J high fat diet---day 28	B6J control diet	Smit E
DGAP	4	Muscle	Mouse	Myocytes PGC-1 alpha overexpression	Myocytes control	Spiegelman B
DGAP	4	Muscle	Mouse	Myocytes PGC-1 beta overexpression	Myocytes control	Spiegelman B
DGAP	5	Muscle	Mouse	B6 (fed)	PGC-1 alpha KO (fed)	Spiegelman B
DGAP	6	Muscle	Mouse	B6 (fasted)	PGC-1 alpha KO (fasted)	Spiegelman B
DGAP	12	Muscle	Mouse	MIRKO	WT	Kahn CR
GEO	4	Muscle	Mouse	PPARα overexpression	WT	Kelly DP
GEO	8	Muscle	Mouse	B6	AIF KO	Penninger JM

DGAP	6	Muscle	Mouse	Continuous intravenous infusion of glucose/insulin---15 min	Before infusion	Kahn CR
DGAP	6	Muscle	Mouse	Continuous intravenous infusion of glucose/insulin---1 hour	Before infusion	Kahn CR
DGAP	6	Muscle	Mouse	Continuous intravenous infusion of glucose/insulin---3 hours	Before infusion	Kahn CR
GEO	10	Muscle	Mouse	BTBR ob/ob severe diabetes 4-wk-old; gastroc	B6 ob/ob mild diabetes 4-wk-old; gastroc	Attie AD
GEO	10	Muscle	Mouse	BTBR ob/ob severe diabetes 10-wk-old; gastroc	B6 ob/ob mild diabetes 10-wk-old; gastroc	Attie AD
GEO	10	Muscle	Mouse	BTBR ob/ob severe diabetes 4-wk-old; soleus	B6 ob/ob mild diabetes 4-wk-old; soleus	Attie AD
GEO	10	Muscle	Mouse	BTBR ob/ob severe diabetes 10-wk-old; soleus	B6 ob/ob mild diabetes 10-wk-old; soleus	Attie AD
GEO	4	Muscle	Mouse	BTBR ob/ob severe diabetes 14-wk-old; soleus	B6 ob/ob mild diabetes 14-wk-old; soleus	Attie AD
GEO	8	Muscle	Rat	High capacity runner sedentary	High capacity runner exercise-trained	Wisløff U
GEO	8	Muscle	Rat	Low capacity runner sedentary	Low capacity runner exercise-trained	Wisløff U
GEO	8	Muscle	Rat	ZDF diabets 12-wk-old	ZDF non-diabetes 6-wk-old	Jung MH
DGAP	2	Pancreatic islets	Human	Asian diabetes	Asian NGT	Kahn CR
DGAP	8	Pancreatic islets	Human	Caucasian diabetes	Caucasian NGT	Kahn CR
GEO	6	Pancreatic islets	Human	Diabetes	NGT	Bruijin JA
GEO	10	Pancreatic islets	Mouse	BTBR ob/ob diabetes 4-wk-old	BTBR lean NGT 4-wk-old	Attie AD
GEO	10	Pancreatic islets	Mouse	BTBR ob/ob diabetes 10-wk-old	BTBR lean NGT 10-wk-old	Attie AD
GEO	10	Pancreatic islets	Mouse	B6 ob/ob diabetes 4-wk-old	B6 lean NGT 4-wk-old	Attie AD
GEO	10	Pancreatic islets	Mouse	B6 ob/ob diabetes 10-wk-old	B6 lean NGT 10-wk-old	Attie AD
GEO	8	Pancreatic islets	Mouse	B6J high fat diet	AKR/J high fat diet	Ahima R

GEO	10	Pancreatic islets	Mouse	B6J islets Tx with high glucose	B6J islets Tx with low glucose	Lloyd DJ
GEO	9	Pancreatic islets	Mouse	BLKSJ islets Tx with high glucose	BLKSJ islets Tx with low glucose	Lloyd DJ
GEO	9	Pancreatic islets	Mouse	B6J islets Tx with high glucose	BLKSJ islets Tx with high glucose	Lloyd DJ
GEO	10	Pancreatic islets	Mouse	B6J islets Tx with low glucose	BLKSJ islets Tx with low glucose	Lloyd DJ
GEO	10	Pancreatic islets	Mouse	BTBR ob/ob severe diabetes 4-wk-old	B6 ob/ob mild diabetes 4-wk-old	Attie AD
GEO	10	Pancreatic islets	Mouse	BTBR ob/ob severe diabetes 10-wk-old	B6 ob/ob mild diabetes 10-wk-old	Attie AD
GEO	4	Pancreatic islets	Mouse	BTBR ob/ob severe diabetes 14-wk-old	B6 ob/ob mild diabetes 14-wk-old	Attie AD
GEO	8	Pancreatic islets	Rat	Islets PPARγ overexpression	Islets control	Rutter GA
GEO	8	Pancreatic islets	Rat	Islets PPARγ activation	Islets control	Rutter GA
GEO	8	Pancreatic islets	Rat	Islets PPARγ overexpression and activation	Islets control	Rutter GA
GEO	6	Kidney	Mouse	db/db 8-wk-old	db/m 8-wk-old	Simonson MS
GEO	6	Kidney	Mouse	db/db 16-wk-old	db/m 16-wk-old	Simonson MS
GEO	10	Hypothalamus	Mouse	BTBR ob/ob diabetes 4-wk-old	BTBR lean NGT 4-wk-old	Attie AD
GEO	10	Hypothalamus	Mouse	BTBR ob/ob diabetes 10-wk-old	BTBR lean NGT 10-wk-old	Attie AD
GEO	10	Hypothalamus	Mouse	B6 ob/ob diabetes 4-wk-old	B6 lean NGT 4-wk-old	Attie AD
GEO	10	Hypothalamus	Mouse	B6 ob/ob diabetes 10-wk-old	B6 lean NGT 10-wk-old	Attie AD
GEO	10	Hypothalamus	Mouse	BTBR ob/ob severe diabetes 4-wk-old	B6 ob/ob mild diabetes 4-wk-old	Attie AD
GEO	10	Hypothalamus	Mouse	BTBR ob/ob severe diabetes 10-wk-old	B6 ob/ob mild diabetes 10-wk-old	Attie AD

Table S2. List of microarray platforms in the 130 experiments

Manufacturer	Species	Title
Affymetrix	Human	[HG_U95Av2] Affymetrix Human Genome U95 Version 2 Array
Affymetrix	Human	[HG-U133A] Affymetrix Human Genome U133A Array
Affymetrix	Human	[HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array
Affymetrix	Human	[HG-U133B] Affymetrix Human Genome U133B Array
Affymetrix	Human	[Hu35KsubA] Affymetrix Human 35K SubA Array
Affymetrix	Human	[Hu35KsubB] Affymetrix Human 35K SubB Array
Affymetrix	Human	[Hu35KsubC] Affymetrix Human 35K SubC Array
Affymetrix	Human	[Hu35KsubD] Affymetrix Human 35K SubD Array
Affymetrix	Human	[Hu6800] Affymetrix Human Full Length HuGeneFL Array
Affymetrix	Mouse	[MG_U74Av2] Affymetrix Murine Genome U74 Version 2 Array
Affymetrix	Mouse	[MG_U74Bv2] Affymetrix Murine Genome U74 Version 2 Array
Affymetrix	Mouse	[MG_U74Cv2] Affymetrix Murine Genome U74 Version 2 Array
Affymetrix	Mouse	[MOE430A] Affymetrix Mouse Expression 430A Array
Affymetrix	Mouse	[MOE430B] Affymetrix Mouse Expression 430B Array
Affymetrix	Mouse	[Mouse430_2] Affymetrix Mouse Genome 430 2.0 Array
Affymetrix	Mouse	[Mouse430A_2] Affymetrix Mouse Genome 430A 2.0 Array
Affymetrix	Mouse	[Mu11KsubA] Affymetrix Murine 11K SubA Array
Affymetrix	Mouse	[Mu11KsubB] Affymetrix Murine 11K SubB Array
Affymetrix	Mouse	Affymetrix GeneChip Mouse Genome 430 2.0 Array [CDF: Mm_ENTREZG_9]
Affymetrix	Rat	[RAE230A] Affymetrix Rat Expression 230A Array
Affymetrix	Rat	[Rat230_2] Affymetrix Rat Genome 230 2.0 Array
Agilent Technologies	Mouse	Agilent-012694 Whole Mouse Genome G4122A (Feature Number version)
Applied Biosystems	Human	ABI Human Genome Survey Microarray Version 2
Illumina	Rat	Illumina ratRef-12 v1.0 expression beadchip
IMP Vienna microarray facility	Mouse	Murine 22K cDNA array version
Massachusetts Institute of Technology	Mouse	MIT-Laboratory for Bioinformatics and Metabolic Engineering, 17K Mouse Arrays

National Institute of Health	Rat	Rat 5K cDNA microarray
Rikilt-Institute of Food Safety	Mouse	RIKILT mouse oligo 10K array
Rosetta Inpharmatics / Merck Pharmaceuticals	Mouse	Rosetta/Merck Mouse 44k 1.0 microarray
Stanford Functional Genomics Facility, Stanford School of Medicine	Human	SHED, SHEA, SHGI, SHCV
University Health Network Microarray Centre	Human	University Health Network Human 19K array version 8 (19Kv8)

Table S3. List of top 127 genes in our T2D eGWAS (Bonferroni threshold, $P < 2.0 \times 10^{-6}$)

Gene name	Description	Entrez gene ID	-logP (χ^2)	-logP (Fisher)
CD44	CD44 molecule (Indian blood group)	960	19.07	16.22
KLK2	kallikrein-related peptidase 2	3817	13.44	11.54
PTGER3	prostaglandin E receptor 3 (subtype EP3)	5733	13.23	11.35
LILRA6	leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 6	79168	13.03	11.25
SFPQ	splicing factor proline/glutamine-rich (polypyrimidine tract binding protein associated)	6421	12.46	10.94
ITGA4	integrin, alpha 4 (antigen CD49D, alpha 4 subunit of VLA-4 receptor)	3676	12.40	10.81
HLA-C	major histocompatibility complex, class I, C	3107	12.28	10.68
ADFP	adipose differentiation-related protein	123	12.22	10.70
PKP2	plakophilin 2	5318	12.08	10.48
HIST2H3A	histone cluster 2, H3a	333932	11.47	9.72
SPP1	secreted phosphoprotein 1	6696	10.90	9.42
CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1	1545	10.85	9.50
MYO5A	myosin VA (heavy chain 12, myoxin)	4644	10.73	9.36
ESR1	estrogen receptor 1	2099	10.60	9.21
ITPR2	inositol 1,4,5-triphosphate receptor, type 2	3709	10.49	9.25
HIST1H2AB	histone cluster 1, H2ab	8335	10.46	8.89
LIFR	leukemia inhibitory factor receptor alpha	3977	10.24	8.97
TTC39B	tetratricopeptide repeat domain 39B	158219	9.99	8.63
NFIA	nuclear factor I/A	4774	9.73	8.44
NFATC2	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2	4773	9.49	8.37
ANGPTL4	angiopoietin-like 4	51129	9.39	8.13
SERPINA1	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1	5265	9.31	8.15
POU6F1	POU class 6 homeobox 1	5463	9.31	8.15
FDPS	farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase)	2224	9.09	8.05
WHSC1	Wolf-Hirschhorn syndrome candidate 1	7468	9.05	7.86
EDA	ectodysplasin A	1896	9.03	7.90
RNASE2	ribonuclease, RNase A family, 2 (liver, eosinophil-derived neurotoxin)	6036	8.90	7.75
SOX6	SRY (sex determining region Y)-box 6	55553	8.78	7.80
PRLR	prolactin receptor	5618	8.76	7.80
CIT	citron (rho-interacting, serine/threonine kinase 21)	11113	8.76	7.80
KRT19	keratin 19	3880	8.75	7.58
APLP2	amyloid beta (A4) precursor-like protein 2	334	8.71	7.63
PRPF40A	PRP40 pre-mRNA processing factor 40 homolog A (S. cerevisiae)	55660	8.65	7.57
CXCL2	chemokine (C-X-C motif) ligand 2	2920	8.55	7.52
BCL2A1	BCL2-related protein A1	597	8.50	7.53
EGFR	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)	1956	8.34	7.43
NRP1	neuropilin 1	8829	8.34	7.43

PDK4	pyruvate dehydrogenase kinase, isozyme 4	5166	8.22	7.29
PTPRK	protein tyrosine phosphatase, receptor type, K	5796	8.17	7.29
CXCL12	chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)	6387	8.07	7.20
BRAF	v-raf murine sarcoma viral oncogene homolog B1	673	8.04	7.15
SYNJ2	synaptojanin 2	8871	7.89	7.05
IDE	insulin-degrading enzyme	3416	7.88	6.91
MAOA	monoamine oxidase A	4128	7.82	6.92
ACOX1	acyl-Coenzyme A oxidase 1, palmitoyl	51	7.76	6.97
PTPRC	protein tyrosine phosphatase, receptor type, C	5788	7.76	6.97
FYB	FYN binding protein (FYB-120/130)	2533	7.75	6.74
FBXO3	F-box protein 3	26273	7.58	6.78
KCNAB1	potassium voltage-gated channel, shaker-related subfamily, beta member 1	7881	7.50	6.70
TCF7L2	transcription factor 7-like 2 (T-cell specific, HMG-box)	6934	7.45	6.62
GLUT4	solute carrier family 2 (facilitated glucose transporter), member 4	6517	7.44	6.56
GPD1	glycerol-3-phosphate dehydrogenase 1 (soluble)	2819	7.44	6.68
LYZ	lysozyme (renal amyloidosis)	4069	7.44	6.68
SPON1	spondin 1, extracellular matrix protein	10418	7.41	6.44
SFRS7	splicing factor, arginine/serine-rich 7, 35kDa	6432	7.38	6.50
BPHL	biphenyl hydrolase-like (serine hydrolase)	670	7.32	6.60
CD36	CD36 molecule (thrombospondin receptor)	948	7.26	6.50
KLF12	Kruppel-like factor 12	11278	7.25	6.35
CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	1576	7.25	6.42
IFNA13	interferon, alpha 13	3447	7.22	6.21
KALRN	kalirin, RhoGEF kinase	8997	7.14	6.20
ABI2	abl interactor 2	10152	7.03	6.23
CREBZF	CREB/ATF bZIP transcription factor	58487	7.03	6.23
MTUS1	mitochondrial tumor suppressor 1	57509	7.01	6.04
PCK1	phosphoenolpyruvate carboxykinase 1 (soluble)	5105	7.01	6.22
HLA-B	major histocompatibility complex, class I, B	3106	6.94	6.07
MYT1L	myelin transcription factor 1-like	23040	6.94	6.21
CD47	CD47 molecule	961	6.91	6.17
PPARG	peroxisome proliferator-activated receptor gamma	5468	6.89	6.10
TSC22D2	TSC22 domain family, member 2	9819	6.89	6.12
C8orf4	chromosome 8 open reading frame 4	56892	6.88	5.97
MTSS1	metastasis suppressor 1	9788	6.85	6.10
GRB10	growth factor receptor-bound protein 10	2887	6.83	6.14
ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2	5168	6.83	6.14
EPB41	erythrocyte membrane protein band 4.1 (elliptocytosis 1, RH-linked)	2035	6.70	6.01
MBP	myelin basic protein	4155	6.68	5.93
RABGAP1L	RAB GTPase activating protein 1-like	9910	6.68	5.93
FGFR2	fibroblast growth factor receptor 2	2263	6.66	5.89

PDE3B	phosphodiesterase 3B, cGMP-inhibited	5140	6.65	5.81
CTBP2	C-terminal binding protein 2	1488	6.63	5.98
NRF1	nuclear respiratory factor 1	4899	6.59	5.83
LEPR	leptin receptor	3953	6.56	5.79
SNTB1	syntrophin, beta 1 (dystrophin-associated protein A1, 59kDa, basic component 1)	6641	6.50	5.85
HOMER1	homer homolog 1 (Drosophila)	9456	6.49	5.78
CLU	clusterin	1191	6.48	5.83
EPHA3	EPH receptor A3	2042	6.45	5.74
THRSP	thyroid hormone responsive (SPOT14 homolog, rat)	7069	6.43	5.64
KIAA0999	KIAA0999 protein	23387	6.40	5.50
COL1A1	collagen, type I, alpha 1	1277	6.32	5.74
NCAM1	neural cell adhesion molecule 1	4684	6.32	5.74
MYH10	myosin, heavy chain 10, non-muscle	4628	6.28	5.66
ELOVL6	ELOVL family member 6, elongation of long chain fatty acids (FEN1/Elo2, SUR4/Elo3-like, yeast)	79071	6.28	5.66
CALM2	calmodulin 2 (phosphorylase kinase, delta)	805	6.25	5.44
C10orf58	chromosome 10 open reading frame 58	84293	6.25	5.58
PSMF1	proteasome (prosome, macropain) inhibitor subunit 1 (PI31)	9491	6.21	5.38
SPOCK2	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 2	9806	6.20	5.62
ANK3	ankyrin 3, node of Ranvier (ankyrin G)	288	6.19	5.62
PTPRJ	protein tyrosine phosphatase, receptor type, J	5795	6.19	5.62
SLC1A1	solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	6505	6.19	5.62
ZFP106	zinc finger protein 106 homolog (mouse)	64397	6.19	5.62
MPEG1	macrophage expressed 1	219972	6.18	5.56
GARNL1	GTPase activating Rap/RanGAP domain-like 1	253959	6.17	5.44
ZDHHC2	zinc finger, DHHC-type containing 2	51201	6.17	5.42
MAP4K4	mitogen-activated protein kinase kinase kinase kinase 4	9448	6.07	5.50
IGFBP1	insulin-like growth factor binding protein 1	3484	6.07	5.49
KCNQ1	potassium voltage-gated channel, KQT-like subfamily, member 1	3784	6.05	5.43
COL5A1	collagen, type V, alpha 1	1289	6.03	5.39
BUB1	budding uninhibited by benzimidazoles 1 homolog (yeast)	699	6.03	5.40
HOXA4	homeobox A4	3201	6.03	5.40
NRXN1	neurexin 1	9378	6.00	5.31
DDHD1	DDHD domain containing 1	80821	5.98	5.40
CHL1	cell adhesion molecule with homology to L1CAM (close homolog of L1)	10752	5.95	5.37
AQP4	aquaporin 4	361	5.91	5.25
CPT2	carnitine palmitoyltransferase 2	1376	5.91	5.25
HLA-DQA1	major histocompatibility complex, class II, DQ alpha 1	3117	5.91	5.25
CCR5	CCR5 chemokine (C-C motif) receptor 5	1234	5.90	5.15
FMOD	fibromodulin	2331	5.87	5.32
MEF2C	myocyte enhancer factor 2C	4208	5.87	5.32

RAD51	RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae)	5888	5.86	5.32
VTI1A	vesicle transport through interaction with t-SNAREs homolog 1A (yeast)	143187	5.82	5.23
SEMA3B	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3B	7869	5.82	5.22
SCD	stearoyl-CoA desaturase (delta-9-desaturase)	6319	5.78	5.28
DNMT3A	DNA (cytosine-5-)-methyltransferase 3 alpha	1788	5.73	5.19
DTNA	dystrobrevin, alpha	1837	5.73	5.19
ENPEP	glutamyl aminopeptidase (aminopeptidase A)	2028	5.73	5.19
TMED5	transmembrane emp24 protein transport domain containing 5	50999	5.72	5.07
SBK2	SH3-binding domain kinase family, member 2	646643	5.71	4.98