Genetic drivers of pancreatic islet function

Mark P. Keller¹#, Daniel M. Gatti²#, Kathryn L. Schueler¹, Mary E. Rabaglia¹, Donnie S. Stapleton¹, Petr Simecek², Matthew Vincent², Sadie Allen⁴, Aimee Teo Broman³, Rhonda Bacher³, Christina Kendziorski³, Karl W. Broman³, Brian S. Yandell⁵, Gary A. Churchill²*, and Alan D. Attie¹*

¹University of Wisconsin-Madison, Biochemistry Department, Madison, WI
²The Jackson Laboratory, Bar Harbor, ME
³University of Wisconsin-Madison, Department of Biostatistics and Medical Informatics, Madison, WI
⁴Maine School of Science and Mathematics, Limestone, ME
⁵University of Wisconsin-Madison, Department of Horticulture, Madison, WI

#These authors contributed equally.
Running title: Genetic regulation of islet function

Key words: mediation analysis; eQTL; module-QTL; T2D; β-cell; genome-wide association (GWA)

*Co-corresponding authors

Alan D. Attie
Department of Biochemistry-UW Madison
433 Babcock Drive
Madison, WI 53706-1544
Phone: (608) 262-1372
Email: adattie@wisc.edu

Gary A. Churchill
The Jackson Laboratory
600 Main Street
Bar Harbor, ME 06409
Phone: (207) 288-6189
Email: gary.churchill@jax.org
Abstract

The majority of gene loci that have been associated with type 2 diabetes play a role in pancreatic islet function. To evaluate the role of islet gene expression in the etiology of diabetes, we sensitized a genetically diverse mouse population with a Western diet high in fat (45%-kcal) and sucrose (34%) and carried out genome-wide association mapping of diabetes-related phenotypes. We quantified mRNA abundance in the islets and identified 18,820 expression quantitative trait loci. We applied mediation analysis to identify candidate causal driver genes at loci that affect the abundance of numerous transcripts. These include two genes previously associated with monogenic diabetes (PDX1 and HNF4A), as well as three genes with nominal association with diabetes-related traits in humans (FAM83E, IL6ST, and SAT2). We grouped transcripts into gene modules and mapped regulatory loci for modules enriched with transcripts specific for α-cells, and another specific for δ-cells. However, no single module enriched for β-cell-specific transcripts, suggesting heterogeneity of gene expression patterns within the β-cell population. A module enriched in transcripts associated with branched chain amino acid metabolism was the most strongly correlated with physiological traits that reflect insulin resistance. Although the mice in this study were not overtly diabetic, the analysis of pancreatic islet gene expression under dietary-induced stress enabled us to identify correlated variation in groups of genes that are functionally linked to diabetes-associated physiological traits. Our analysis suggests an expected degree of concordance between diabetes-associated loci in the mouse with those found in human populations and demonstrates how the mouse can provide evidence to support nominal associations found in human genome-wide association mapping.
Introduction

Type 2 diabetes (T2D) is a highly heritable disease \( (h^2 \approx 0.5) \) (SANGHERA AND BLACKETT 2012). More than 100 gene loci associated with diabetes or diabetes-related phenotypes have been identified through genome-wide association (GWA) studies (MORRIS et al. 2012; NG et al. 2014; REPLICAION et al. 2014; FLANNICK AND FLOREZ 2016; FUCHSBERGER et al. 2016; JASON et al. 2017). However, the effect size of each locus is small; their odds ratios are typically ~1.05 – 1.10, and most of the heritability of T2D in human populations remains to be elucidated. In addition, much remains to be discovered about the regulatory mechanisms responsible for the wide range in susceptibility to diabetes.

The risk of T2D is quite low in the absence of obesity. Prior to the start of the obesity epidemic, ~60 years ago, the incidence of T2D was <1%. Today, the incidence is >9%. Thus, the gene loci responsible for the susceptibility to T2D act primarily in the context of obesity. Obesity usually leads to insulin resistance, resulting in an increased demand for insulin production to maintain normal glucose levels. Although the etiology of T2D involves interactions among multiple organ systems, one concept that has emerged from the GWA studies is that, for the most part, the causal genetic factors leading to T2D exert their effect by limiting the capacity of pancreatic \( \beta \)-cells to secrete sufficient insulin to maintain normal glucose levels. A substantial proportion of the candidate genes that have emerged from genetic studies in humans and model organisms affect \( \beta \)-cell function or \( \beta \)-cell mass (BILLINGS AND FLOREZ 2010; MOHLKE AND BEHNKE 2015; PRASAD AND GROOP 2015). In the case of monogenic diabetes syndromes, essentially all the causal genes are expressed in \( \beta \)-cells (FAJANS et al. 2001; SHIH AND STOFFEL 2001; TANEERA et al. 2014).

For purposes of genetic analysis, treating T2D as a binary disease is clearly inadequate. Thus, human and model organism studies focus on quantitative diabetes-related traits, including plasma glucose and insulin levels. However, normal blood glucose levels could equally reflect a healthy, or compensatory state that is near the breaking-point. One way to confront this complexity is to examine
the compensatory mechanisms in genetically diverse individuals that display a wide range of compensatory responses to an environmental stressor.

We hypothesized that the biochemical reactions and cellular signaling pathways that constitute the range of stress-induced responses across individuals would be observed as correlated changes in the patterns of gene expression in key organ systems. By summarizing expression patterns within groups of mRNAs as meta-traits, we can achieve a large dimension reduction, enabling a clearer understanding of the molecular functions involved in the disease process. To translate findings from mouse to human, an understanding of the processes involved in the disease is perhaps more relevant than the identification of causal variants.

In the context of a genetic study, mRNA abundance can be mapped in much the same way as a physiological trait. The relationship between genotype and mRNA abundance involves a unidirectional line of causality (SCHADT et al. 2005; MILLSTEIN et al. 2009; NETO et al. 2010; NETO et al. 2013). Anchoring mRNA abundance and other phenotypes to genetic variation provides a powerful means to reveal causal drivers – genes that harbor genetic variants that influence disease-associated phenotypes. Often multiple mRNA abundance traits map to the same locus and are influenced by common genetic drivers (ALBERT AND KRUGLYAK 2015; YAO et al. 2017). When these co-mapping mRNAs encode proteins that are associated with common physiological functions this can shed light on potential biological functions of the driver gene(s). These connections evoke testable hypotheses whereby variation in the expression of a driver gene, rather than genetic variant, can be established as a more proximal cause for a disease-related phenotype. The association between driver genes and their downstream effects can unveil novel pathways as well as the tissue sites of their action (FRANZEN et al. 2016).

T2D is a disorder of relative insulin deficiency. Pancreatic β-cells are challenged by an increased demand for insulin resulting from insulin resistance. A deficiency of β-cell mass or β-cell function usually does not result in diabetes, however, in the context of insulin resistance, can lead to an insulin
shortfall, and diabetes. Because pancreatic islets are not accessible for detailed experimental study in humans, it is not feasible to directly interrogate β-cell function in the context of a properly-powered human GWA study. The repertoire of available mouse strains, harboring the same degree of genetic variability in human populations, displays a wide range in β-cell function. This range can be interrogated in non-diabetic mice where diet treatment is used to challenge the β-cells to increase their insulin secretion.

We conducted a study to map diabetes-related traits in Diversity Outbred (DO) mice. The DO is an outbred mouse population derived from 8 inbred strains, including 3 wild-derived strains. The DO captures a broad range of genetic variation comparable to that found in human populations (SVENSON et al. 2012), and presents a similarly broad range of individual susceptibility to T2D. We metabolically challenged 500 DO mice with a high-fat/high-sucrose Western-style diet, measured diabetes-related physiological phenotypes (e.g., plasma glucose and insulin), and isolated their pancreatic islets for molecular characterization using RNA-seq to reveal variation in patterns of gene expression in islets. Previously, this diet was used as a metabolic challenge to evoke broad phenotypic responses in 43 inbred mouse strains, and to genetically map several metabolic phenotypes, including obesity, gut microbial composition, glucose homeostasis, dyslipidemia, hepatic steatosis, atherosclerosis and energy balance (SVENSON et al. 2007; PARKS et al. 2013; SPIEZIO et al. 2014; SINASAC et al. 2016). A recent study by Threadgill and colleagues compared the metabolic effects of the Western-style diet to a control diet, and three additional diets (Mediterranean, Japanese, and Maasai/ketogenic) in four mouse strains; A/J, C57BL/6J, FVB/NJ and NOD/ShiLtJ (BARRINGTON et al. 2018). The Western-style diet, compared to a control diet, had detrimental health effects, such as increased body fat, LDL-cholesterol and liver triglycerides in all strains, but the effect size varied across stains. In response to the other diets, improvements in these metabolic effects was also strain-dependent. These strain-dependent effects underscore the importance of studying the metabolic impact of diet on diverse genetic backgrounds, making the DO panel a particularly relevant population for this study.
Previous studies that used human islets have identified cis-regulatory maps of gene regulation, and overlaid these maps with marks of chromatin accessibility, and association with diabetes-related traits from human GWA studies (VAN DE BUNT et al. 2015; VARSHNEY et al. 2017). Herein, we present a comprehensive picture of the genetically driven variation of >21,000 transcripts expressed in mouse islets. We show how these transcripts are co-regulated and associated with local and distal expression quantitative trait loci (eQTL). We demonstrate that the genetic drivers behind this variability in the DO mice correspond to genetic effects at human GWA loci. Thus, in addition to the conservation of the physiological and biochemical pathways involved in diabetes, genetically variable driver genes appear to be conserved across mammals. Our data provide a resource that can be integrated with other molecular and genetic data, both human and model organism, to identify causal genes and explore the mechanisms involved in initial stages of type II diabetes pathogenesis.
**Material and Methods**

**Animal husbandry and measurement of physiological phenotypes**

Diversity Outbred (DO) mice were obtained from the Jackson Labs (stock no. 009376) at 4 weeks of age and maintained within the Biochemistry Department animal vivarium at the University of Wisconsin. Waves of 100 DO mice, half for each sex, were obtained 3 times per year, until 500 DO mice were surveyed. DO generations 18, 19 and 21 were included in the cohort. Upon arrival, all DO mice were maintained on a diet high in fat and sucrose (44.6%-kcal fat, 34% carbohydrate, 17.3% protein) from Envigo Teklad (cat no. TD.08811). We maintained all DO mice on the HF/HS diet to provide an environmental stressor, sensitizing the mice to the development of diabetes. Our study does not provide information about specific dietary effects, as a control diet was not included in the study design. To measure food intake, all DO mice were singly housed. Although there is no biological replication of genetically identical animals in an outcross population, there is replication of genotypes at specific loci. This local genetic replication enables one to link phenotype with genotype, as in human GWAS, or QTL mapping studies in DO mice.

Body weight was measured bi-weekly, and fasting plasma samples were collected for insulin, glucose and triglyceride (TG) measurements. At ~18 weeks of age, an oral glucose tolerance test (oGTT) was performed on 4-hr fasted mice to evaluate dynamic changes in plasma insulin and glucose. Glucose (2 gm/kg) was administered via oral gavage. Blood was collected from a retro-orbital bleed before glucose administration, and at 5, 15, 30, 60, and 120 minutes. Area under the curve (AUC) was determined from these time points for glucose and insulin. Glucose was measured by the glucose oxidase method using a commercially available kit (TR15221, Thermo Scientific). Insulin was measured by radioimmunoassay (RIA; SRI-13K, Millipore). Pancreas weight was not recorded, as it was inflated *in situ* with collagenase as the first step to islet isolation.

HOMA-IR and HOMA-B, homeostatic measures of whole-body insulin sensitivity and pancreatic islet function, respectively, were determined using fasting plasma values of glucose and insulin at the time
the oGTT was administered (time = 0 values). HOMA-IR is equal to (glucose x insulin)/405; HOMA-B = (360 x insulin)/(glucose – 63). Units for plasma glucose and insulin are mg/dl and mU/L, respectively.

**Islet RNA profiling**

Whole-islet RNA was isolated and evaluated for quality. The average RNA-integrity number (RIN) for all DO islet samples was 9.24 ± 0.04. RNA samples from 96 animals from each of the first 4 waves of mice (384 in total) were submitted to The Jackson Laboratory high throughput sequencing core facility. RNAs were sheared using the E220 Focused-ultrasonicator (Covaris). The Jackson Laboratory core generated whole genome islet mRNA libraries using the KAPA Hyper Prep Kit for Illumina Sequencing (KAPA Biosystems), targeting an insert size of 300bp using magnetic bead-based size selection. Libraries from waves 1, 2, and 3 were constructed at the Jackson Laboratory and randomized into pools of 24 samples with TruSeq RNA indices (6 bp). Libraries from wave 4 were constructed at the New York Genome Center and were pooled in sets of 24 using 8bp dual index TruSeq RNA indices. Each pooled RNA sample was sequenced across 4 randomly assigned lanes on a HiSeq2500 (Illumina), 1x100bp at the New York Genome Center. To control for potential batch effects, we incorporated wave as a covariate in all subsequent analyses.

Reads were aligned to eight strain-specific transcriptomes of the DO founders (MUNGER et al. 2014). We used an expectation maximization algorithm (EMASE, https://github.com/churchill-lab/emase) to obtain estimated total read counts for each gene as a sum across alleles and isoforms (RAGHUPATHY et al. 2018). We normalized read counts in each sample using upper-quantile normalization.

**Mouse genotyping and haplotype reconstruction**

Genotyping was performed on tail biopsies as described in Svenson et al. (2012), using the Mouse Universal Genotyping Array (GigaMUGA) (143,259 markers (MORGAN et al. 2015)) at Neogen (Lincoln, NE). Genotypes were converted to founder strain-haplotype reconstructions using R/DOQTL software (GATTI et al. 2014). We interpolated the GigaMUGA markers onto an evenly spaced grid with 0.02 cM
spacing and added markers to fill in regions with sparse physical representation, resulting in 69,005 pseudo-markers. We also reconstructed individual chromosome haplotypes from the RNA-seq data using a hidden Markov model (GBRS, https://github.com/churchill-lab/gbrs). We identified three samples with inconsistent genotypes between the GigaMUGA and RNA-seq-based haplotype reconstructions. Three additional samples were determined to have poor quality RNA-seq data. These six samples were excluded, leaving a total of 378 samples for subsequent analyses.

**Genome scans to identify QTL for physiological and gene expression traits**

Genetic mapping analysis was carried out with the `qtl2` R package (http://kbroman.org/qtl2) using the founder haplotype regression method with Leave One Chromosome Out (LOCO) option for kinship correction (GATTI et al. 2014) (http://kbroman.org/qtl2). Phenotype data were log transformed. We performed genome scans for each phenotype with sex and experimental cohort (wave) as covariates. We estimated significance thresholds by permuting the phenotype values 1000 times while holding the genotypes fixed, mapping the permuted trait and retaining the maximum LOD score from each permutation (CHURCHILL AND DOERGE 1994). Complete details of our analyses are provided in R Markdown files that can be used to reproduce our work in RStudio (Files S1 – 3). A guide to the contents, and data type analyzed in each Markdown file (e.g., physiological traits, eQTL hotspots, and module eigengenes) is provided in File S4.

Normalized RNA-Seq data were transformed to normal scores (van der Waerden’s method, CONOVER 1999) and we carried out QTL mapping as described above. We called an eQTL a “local-eQTL” if the marker with the maximum LOD score was within ± 4 Mb of the transcriptional start site and called the remaining eQTL “distal-eQTL.” We identified eQTL hotspots by counting the number of eQTL with LOD > 7.2 that occurred in a sliding 4 Mb window with a 1 Mb overlap between windows. We retained all genes with LOD > 6 within ± 2 Mb (4 Mb width) of the center of each hotspot and calculated the first principal component of the genes in each hotspot.
**Mediation analysis to identify causal drivers**

In order to identify candidate driver genes for distal-eQTL we applied mediation analysis (MacKinnon et al. 2007). We denote the genotype at the QTL locus as Q, the gene expression of the distal eQTL transcript as Y, and the gene expression of the candidate driver gene as M. We hypothesize a causal pathway model \( Q \rightarrow M \rightarrow Y \) in which the genetic variation on the distal eQTL transcript is fully or partially mediated by variation in the expression of the driver gene. To evaluate this model, we applied the causal steps method and verified each of 4 conditions:

i) The distal transcript Y should have a significant LOD score at the locus Q.

ii) The candidate mediator M should have a significant LOD score at Q. Typically the LOD score of the mediator will be greater than the distal-eQTL. In addition, we require that the transcript M should derive from a gene at the locus Q, i.e., that the mediator has a local eQTL, and that the allele effects pattern matches that of the distal eQTL.

iii) Adding the mediator M to the regression of Y on Q should cause a drop in the distal eQTL LOD score. To determine if the drop in LOD score is significant, we tested every transcript genome-wide as a candidate mediator and computed a Z-score. We identified candidate mediators with a Z-score less than -6.

iv) Adding the distal transcript Y as a covariate in the regression of M on Q, we require that the conditional LOD score remains significant at a nominal (not genome-wide adjusted) level of 0.01.

We evaluated genes with local eQTL at hotspot locus as candidate mediators, these genes satisfy condition (ii). We evaluated the LOD drop for the first principal component of the hotspot genes and for each individual gene with a distant eQTL at the hotspot. The logic of mediation analysis allows us to rule out candidate mediators, but not to prove them. There are some potential pitfalls to genetic mediation analysis (Didelez and Sheehan 2007). A transcript with a strong local eQTL (typically with LOD > 100) can satisfy conditions i) through iv) because it can act as surrogate for the local genotype.
If the distal eQTL is mediated by a mechanism other than a change in gene expression of a local transcript, mediation on RNA abundance will fail to detect this.

**Weighted gene co-expression network analysis (WGCNA)**

Co-expression gene modules were computed using Weighted Correlation Network Analysis (WGCNA) (Zhang and Horvath 2005), which performs network construction and module detection based on correlation between traits. For the WGCNA analysis, we included all expression traits (21,771 transcripts) that were measured in 378 DO islet samples. We used a signed WGCNA network with minimum module size of 30, and a soft thresholding power of 12. A first principle component (PC1) was calculated for each module, which we refer to as the “module eigengene” (ME), and used for module – trait correlation, and module QTL analyses. The ME for each module was computed from the islet RNA-seq measurements for all DO mice, with no distinction made for sex. Thus, MEs are based purely on the correlation structure for all transcripts within a module for all mice. To identify which modules showed a sex difference in the expression pattern of the transcripts, we performed a Student’s t-test using the ME value between males and females (Table S5). Module enrichment analysis was done using allez (Newton et al. 2007) to identify enriched categories in the Gene Ontology (GO) or Kyoto encyclopedia of genes and genomes (KEGG). This analysis identifies differential enrichment of a functional category among traits in a module, compared to traits not in the module. Of the ~22,000 transcripts used for module calculation, 13,412 (~62%) were assigned to a module. The average number of transcripts per module was 327 ± 51 transcripts and ranged from 42 (lightsteelblue1) to 1494 (turquoise).

**Identification of human loci syntenic to mouse QTL, and integration with diabetes-associated GWAS**

The R/Bioconductor package ChIPseeker (v1.13.1) (Yu et al. 2015) was used to first annotate the QTL locations to the nearest annotated gene in the UCSC R/Bioconductor package.
We used GWAS central (http://www.gwascentral.org/) to download all SNPs that are associated with T1D and T2D above a threshold of \(-\log_{10} p\)-value \(\geq 4\), resulting in 666 and 509 unique SNPs, respectively (BECK et al. 2014); these SNPs are contained within Table S8. We then determined the number of T1D or T2D-associated SNPs occurring within a 1 Mbp genomic window, yielding the histograms shown for each chromosome in Fig. 9. The number of T1D-associated SNPs at the HLA-locus on chromosome 6 (~30 through 33 Mbp) was 344. For display purposes of the T1D GWAS histogram shown in Fig. 9, the maximum value for the 1 Mbp bins was set to 20 at the HLA-locus. All analyses used to integrate the mouse QTL with human GWA studies are reproduced in an R Markdown document, accompanying data files, and guide (Files S5, 6, and 7, respectively).

We identified 67 module QTL, and 30 physiological QTL for a total of 97; 49 of these QTL are syntenic to human loci that are within 1 Mbp of the Type I and Type II diabetes related GWAS SNPs that we obtained from GWAS Central. We performed a test of significance by randomly selecting 97 human genes from a restricted set of genes located on chromosomes 1 – 22 and X. The human genes were restricted to those from biomaRt that have syntenic regions in hg19 and were mappable to mm10. Additionally, we only included genes whose expression mean was greater than the 5th percentile of mean expression in our mouse islet RNA-seq data. For these 97 randomly selected genes, we then...
counted how many were within 1 Mbp of the diabetes related SNPs from GWAS Central. We repeated this sampling 1,000 times to derive an empirical \( p \)-value of enrichment.

**Data and Software Availability**

Raw sequence reads have been deposited with SRA (SRP125176). Phenotypes, genotypes and quantified gene expression have been deposited with Dryad (doi:10.5061/dryad.pj105. Data files: Attie Islet eQTL data). The data are also accessible through an interactive web-based analysis tool that will allow users to replicate the analyses reported here, including genetic mapping and mediation analysis (http://churchill-lab.jax.org/qtl/islet/DO378). Software for this interactive tool is maintained at https://github.com/churchill-lab/qtlviewer, and https://github.com/churchill-lab/qtlapi.
Results

Wide range of diabetes-related phenotypes in DO mice maintained on a Western-style diet

To better understand the impact of Western-style diet on physiological traits related to diabetes, we generated a cohort of ~500 Diversity Outbred (DO) mice that were metabolically challenged with a diet high in fat (45%-kcal) and sucrose (34%); HF/HS diet. We measured body weight, fasting plasma glucose, insulin, and triglycerides (TG) at regular intervals (Fig. 1 A – D). There was a large spread in phenotype values, due in part to the genetic diversity of DO mice. For example, body weight at 14 weeks of age ranged from ~15 to >50 gms (Fig. 1A), and fasting plasma insulin showed a ~100-fold range (Fig. 1C), consistent with significant differences in peripheral insulin resistance among the mice. Only 2 of the ~500 DO mice evaluated became diabetic (blood glucose > 300 mg/dl) (Fig. 1B), suggesting that the HF/HS diet evoked a compensatory increase in insulin production to offset diet-induced insulin resistance. All mice were individually housed throughout the course of our study, allowing us to measure daily food consumption for individual mice. As with the other phenotypes, there was a broad range in the amount of food consumed, 1.9 – 5.5 gm/day, and was on average, higher in males than females (Fig. 1L), reflecting differences in body weight between the sexes. At all times, male DO mice demonstrated higher values for body weight ($p < 2.2e-16$), plasma glucose ($p < 2.2e-16$), insulin ($p < 2.0e-12$) and TG ($p < 6.5e-10$), compared to female DO mice.

To interrogate islet function, we performed an oral glucose tolerance test (oGTT) at ~18 weeks of age. Prior to the oGTT, all mice were fasted for 4 hrs, and a blood sample was collected, from which we computed HOMA-IR and HOMA-B, homeostatic measures that are related to peripheral insulin resistance and β-cell function, respectively (MATTHEWS et al. 1985). HOMA-IR (Fig. 1E) showed a >100-fold range and was higher in males than females ($p = 1.0e-11$), whereas HOMA-B was not different between the sexes (Fig. 1F). The glucose (Fig. 1G) and insulin (Fig. 1H) responses during the oGTT ($\text{AUC}_{\text{glucose}}$ and $\text{AUC}_{\text{insulin}}$, respectively) were significantly higher in males than females ($p = 2.0e-11$) and showed a large dynamic range.
At ~22 weeks of age, all DO mice were euthanized and their islets isolated by hand-picking. The average number of islets per DO mouse was 483, ranging from 42 to 1096 (Fig. 1I). The average amount of insulin per islet was 85 ng/islet and ranged from ~9 to 290 ng/islet (Fig. 1J). The whole-pancreas insulin content (WPIC), a surrogate measure of β-cell mass per pancreas was 42.7 ± 1.4 µg/pancreas (Fig. 1K). The number of islets per pancreas ($p = 0.03$) and WPIC ($p = 0.01$) were significantly elevated in male mice, whereas the insulin content per islet was not significantly different between the sexes ($p \sim 0.2$). All mice were individually housed throughout the course of our study, allowing us to measure daily food consumption for individual mice. As with the other phenotypes, there was a broad range in the amount of food consumed, 1.9 – 5.5 gm/day, and was on average, higher in males than females (Fig. 1L). However, this difference is largely driven by the lower body weight of female mice (Fig. 1A). Food consumption as a function of weight is nearly identical for male and female DO mice (Fig. S1). Some of the physiological traits are highly inter-correlated (Fig. S2). The correlation patterns are largely concordant between the sexes except for HOMA-B, which is more strongly correlated in males, and AUC$_{\text{glucose}}$, which has stronger correlations with other traits in female mice.

**Diabetes-related phenotypes map to multiple QTL loci**

We next asked if the physiological phenotypes map to QTL. We identified a total of 34 physiological QTL (Fig. 2) that were significant, or suggestive. Among the physiological traits, AUC$_{\text{insulin}}$ and HOMA-B showed the strongest QTL, and co-mapped to a locus on Chr 11 at ~85 Mbp with LOD scores of 11.3 and 10.3, respectively. HOMA-B has a second QTL on Chr 18 at ~47 Mbp. Body weight at 6 and 10 weeks of age map to Chr 11 at ~12 Mbp and to Chr 17 at ~35 Mbp. Plasma insulin at 14 weeks co-maps with the body weight QTL to Chr 17, consistent with a model whereby variation in body weight results in changes in peripheral insulin resistance, leading to increased plasma insulin to maintain euglycemia. However, plasma insulin at 14 weeks also mapped to Chr 1, Chr 5 and Chr 13, indicating that genetic factors acting through mechanisms independent of body weight influence circulating insulin levels. Other physiological QTL included whole pancreas insulin content (WPIC), a surrogate for
pancreatic β-cell mass, and insulin per islet on Chr 4 at ~60 Mbp. HOMA-IR, HOMA-B and plasma insulin at 10 weeks mapped to a locus on Chr 12. In summary, these results demonstrate that the complex nature of diabetes-related physiological phenotypes results from genetic regulation at multiple loci.

**Genetic architecture of the transcriptome reveals hotspots in pancreatic islets**

We sequenced whole-islet mRNA from 378 HF/HS-fed DO mice (188 females and 190 males) at an average depth of ~36M reads/sample and obtained quantitative estimates of abundance for 21,771 islet transcripts. RNA abundance was used to perform whole-genome scans, resulting in the identification of 18,820 significant eQTL (LOD > 7.2, corresponding to permutation-based genome-wide p < 0.05; (CHURCHILL AND DOERGE 1994)) (Fig. 3A). We classified eQTL as either local (peak LOD within 4 Mbp of gene locus; blue dots), or distal (black dots). At this LOD threshold, 68% of the eQTL were local, suggesting that genetic variation proximal to the gene locus plays a key role in regulating transcript abundance among DO mice. LOD scores for local-eQTL tended to be higher than for distal-eQTL, consistent with what is normally observed in eQTL studies, and thus the proportion of local-eQTL increases with higher LOD thresholds (WEST et al. 2007; MUNGER et al. 2014).

As we have previously shown (TIAN et al. 2015), islet distal-eQTL can form clusters at genomic hotspots, where many expression traits co-map to the same locus (BREITLING et al. 2008). We used a 4 Mbp sliding window, with 1 Mbp steps, and defined windows with 100 or more distant eQTL that co-map to be a hotspot. In DO islets, we identified five distal-eQTL hotspots, on Chrs 2, 5, 7, 11 and 13 (Fig. 3B; Table S1). Figure 3C illustrates the genomic profile for the number of co-mapping local-eQTL, identifying several loci with ~100 co-mapping local-eQTL (e.g., Chrs, 7, 17, 19). These local-eQTL hotspots are all in regions of high gene density.

**Mediation analysis predicts candidate drivers of eQTL hotspots**
One possible explanation for co-mapping of many expression traits is the shared influence of polymorphic genetic factors present at the hotspot locus that directly or indirectly influence the distal-eQTL genes. The effect of genetic variation on expression of the distal-eQTL genes must be mediated through one or more trans-acting factors encoded at the hotspot locus. If the mediation occurs through transcriptional variation in regulator genes, we can identify candidate regulators by applying mediation analysis (Chick et al. 2016). A key step in mediation analysis looks for a substantial drop in the LOD score at the hotspot after conditioning on the expression of a candidate regulator gene. A candidate gene is the driver of the hotspot. We evaluated the conditions for mediation using the first principal component (PC1) at each of the five hotspots. In addition, we evaluated each of the distal-eQTL genes to identify individual genes that share the same mediator as the PC1.

The PC1 for the Chr 2 hotspot at ~164.0 Mbp (PC1\textsubscript{Chr2}) had a single significant QTL with LOD score of 92.4 (Fig. 4A). The WSB and NZO alleles were associated with low and high expression of PC1\textsubscript{Chr2}, respectively (Fig. 4B). When conditioned on the expression of each of the genes on Chr 2, the LOD profile for PC1\textsubscript{Chr2} was largely unchanged for most genes. However, when conditioned on \textit{Hnf4a}, located on Chr 2 at ~163.5 Mbp, the LOD for PC1\textsubscript{Chr2} dropped significantly (Fig. 4C), indicating that \textit{Hnf4a} is a candidate mediator. The LOD profile (Fig. 4D) and allele dependence (Fig. 4E) for the local-eQTL for \textit{Hnf4a} closely matches that for PC1\textsubscript{Chr2} (Fig. 4A-B), consistent with the results from conditioning PC1\textsubscript{Chr2} on \textit{Hnf4a}. We evaluated each of the distal eQTL genes, conditioning on \textit{Hnf4a} and found that 88 of the 147 genes in the hotspot had a LOD drop of ≥ 1.5 (Table S1). \textit{Hnf4a} is a transcription factor. Using PSCAN (Zambelli et al. 2009), we asked if the promoters of transcripts co-mapping to the \textit{Hnf4a} gene locus are enriched with a motif associated with \textit{Hnf4a} binding. Known motifs for 635 transcription factors were surveyed within -950 to +50 bp of the transcription start site (TSS) for the 206 transcripts that map to the Chr 2 hotspot. Remarkably, the most significantly enriched ($p = 3.6e-9$) motif was for \textit{Hnf4a} (Fig. S3). Our results suggest that \textit{Hnf4a} is a driver that mediates the distal-eQTL for the Chr 2 hotspot expression traits.
We performed mediation analysis for the PC1s at the other four distal-eQTL hotspots. At the Chr 13 hotspot, we identified *Il6st* as a candidate driver for 82 of 104 genes that map to this hotspot (Fig. S4; Table S1). When the PC1\textsubscript{Chr13} was conditioned on the expression of *Il6st*, the LOD profile for the hotspot was reduced essentially to zero, suggesting that *Il6st* explains all of the genetic variation in PC1\textsubscript{Chr13}. Mediation analysis of the hotspots at Chrs 5, 7 and 11, identified *Pdx-1* (mediates 77 of 182 genes; Fig. S5), *Fam83e* (mediates 96 of 123 genes; Fig. S6), and *Sat2* (mediates 115 of 126 genes; Fig. S7) as potential drivers of their respective hotspot eQTL (Table S1). The Chr 5 hotspot is complex and there may be additional driver genes at this locus.

We found concordance between QTL for physiological traits and four of the five eQTL hotspots. Fasting plasma insulin at 14 weeks of age mapped to the eQTL hotspot on Chr 5 and showed the same allele effects pattern. AUC\textsubscript{glucose} mapped to the Chr 11 eQTL hotspot and HOMA-IR mapped to the eQTL hotspot on Chr 13. Despite the weaker genetic signal, and broad confidence intervals associated with many of the physiological QTL, their allelic effects indicate that these traits are responding to the same genetic drivers at the eQTL hotspots. Importantly, mediation analysis can nominate candidate genes with variants that influence diabetes-related phenotypes through their broad effects on pancreatic islet gene regulation.

**Candidate drivers are associated with diabetes traits in human GWAS**

We next asked if genetic variation at the gene loci for the five candidate hotspot drivers is associated with diabetes-related phenotypes in human GWAS. We used LocusZoom (PRUIM et al. 2010) to generate regional association plots at each of the candidate gene loci for diabetes-associated phenotypes measured in the MAGIC studies, e.g., fasting glucose, insulin, or HbA1c (DUPUIS et al. 2010; SORANZO et al. 2010; MANNING et al. 2012; SCOTT et al. 2012a; SCOTT et al. 2012b). While many of the SNPs associated with these phenotypes are sub-threshold for a genome-wide query, the relatively small number of SNPs interrogated in our analysis (<400) greatly reduces the multiple-testing
penalty for these single-gene searches. We identified SNPs with significant association to one or more diabetes-related phenotypes at three of the five driver gene loci.

Genetic variation at *HNF4A* (Fig. 5A) and *PDX1* (Fig. 5B), also known as MODY1 and MODY4, yielded SNPs associated with BMI-adjusted 2-hr glucose ($p < 10^{-5}$), and fasting glucose ($p < 10^{-8}$), respectively. *SAT2*, which has been linked to polyamine (Hyvonen *et al.* 2013), as well as thialysine (Coleman *et al.* 2004) metabolism, was associated with fasting insulin ($p < 10^{-5}$) (Fig. 5C). *IL6ST* (the β-subunit of the IL6 cytokine receptor) and *FAM83E* (a newly discovered gene that may be involved in MAPK signaling (Cipriano *et al.* 2014)) each were associated with HbA1c, albeit with marginal significance ($p < 10^{-3}$) compared to the other loci (Fig. 5D & E, respectively). In summary, these results demonstrate that by integrating our conditional analyses at the eQTL hotspots to predict causal gene drivers, we performed single-gene queries in human GWAS that support a role for these genes in islet function, and potentially diabetes risk.

**Co-expression modules highlight biological processes related to diabetes**

We employed cluster analysis to identify groups of genes with highly correlated expression patterns in the islet transcriptome (Zhang and Horvath 2005; Langfelder and Horvath 2008). Grouping of transcripts into modules provides a useful summary of the complex correlation structure that is typical of whole transcriptome data. When grouping the transcripts into co-expression modules, we did not utilize information about functional annotations or whether the transcripts were subject to genetic regulation. Among the 21,771 transcripts from our whole-islet RNA sequencing, 62% were assigned to a co-expression gene module. We identified a total of 41 modules with varying numbers of transcripts, ranging from 42 to 1494 (Table S2). The average number of transcripts per module was 327. Modules are depicted as the downward branches in the cluster dendrogram (Fig. 6A), the length of which is proportional to the average gene-gene correlation within each module.
Highly correlated transcripts are often associated with common physiological or biochemical processes (Carlson et al. 2006; Gargalovic et al. 2006; Ghazalpour et al. 2006; Horvath et al. 2006; Keller et al. 2008). We performed gene set enrichment analysis (Newton et al. 2007) on each module to determine if any genes in the modules are associated with shared functional annotations. We will refer to modules by both a color key identifier, and by the most common GO and/or KEGG functional annotations (Fig. S8). Among the 41 modules identified, all but 3 were significantly enriched with one or more gene ontology (GO) terms (Fig. 6B) and/or KEGG pathways (Fig. 6C); Table S3 lists the top enrichment terms for all modules, however, we note that not all genes in modules are associated with the functional annotations.

To determine the potential physiological significance of the islet modules, we asked if the modules were correlated with the diabetes-related phenotypes measured in the DO mice. For each module, we first computed a module eigengene (ME), in a similar fashion to PC1, to describe the pattern of transcript abundance among all the DO mice. The average percent variance explained by the MEs was 38 and ranged from ~70 (mediumpurple3) to ~24 (blue) (Table S4). The percent variance explained by the ME for transcripts that were not assigned to a module (transcripts in grey) was ~2, demonstrating the highly coordinate nature of transcripts within the gene modules. Transcript abundance within 16 modules showed a highly significant difference between the sexes (Table S5; Fig. S9).

Due to the strong influence of sex on many of the physiological phenotypes (Fig. 1) and module transcripts, we computed the Spearman’s nonparametric correlation between modules and traits, after adjusting for sex and DO breeding generation as covariates. This allowed us to determine which modules correlate with specific physiological phenotypes without sex as a confounding variable. Some of the physiological traits display significant correlations among themselves (Fig. S2), and as a result several modules were significantly correlated with more than one physiological trait (Fig. 7). We adjusted the p-value for the Spearman correlations using the Benjamini & Yekutieli method (Reiner et al. 2003) and found that an adjusted p-value of 0.05 corresponded to a correlation of ~0.15. Here we
highlight features of the most relevant modules and for each, ask: 1) Are the modules correlated with physiological phenotypes? 2) Are they enriched for annotations suggesting that (some of) the genes may be functionally related to diabetes? and 3) Do the modules demonstrate genetic regulation?

Modules associated with ribosome biogenesis (darkred), branched chain amino acid (BCAA) catabolism (brown), cell cycle (violet), and hexosaminidase activity (yellow) were most strongly correlated with body weight, plasma insulin, AUC_{insulin}, HOMA-B, and HOMA-IR (Fig. 7). The δ-cell module (yellowgreen) module, also correlated with these traits, is discussed below. The correlations are strongest between these modules and the later time points of physiological traits that were measured at multiple times (e.g., insulin at 14 weeks). The brown, violet, and yellow modules all share a significant QTL on Chr 11 at ~70 Mb that is concordant with the Chr 11 hotspot and with a QTL for HOMA-IR (Fig. 8). The brown module is enriched for branched chain amino acid catabolism genes including Bckhda, Bckhdb, and Bckdk. The latter is one of several genes in the module that are mediated by Sat2. The violet module is enriched for the GO term, “chromosome segregation” and the KEGG pathway, “Cell cycle”. It contains 145 transcripts, including many that are known to play a role in the regulation of the cell cycle; the cyclins A2, B2, and B1, Asf1b, Pbk, Aurkb, Cdk1, Bub1, Bub1b, Mcm3, Mcm5, Plk1, Top2a, and Ube2c, among others. Transcripts in the violet module largely overlap with an islet cell cycle-related module we previously identified when comparing diabetes-resistant (B6) and diabetes-susceptible (BTBR) mouse strains (Keller et al. 2008). The violet module has a second QTL on Chr 17 at ~70 Mb, which is concordant with the cell-cycle module discovered in the B6 x BTBR cross. The darkred module has a distinct QTL on Chr 17 at ~34 Mb, which is concordant with QTL for weight at 6 and 10 weeks of age.

Several modules show enrichment for genes with physiological functions that are relevant to the pathology of T2D but did not co-map to hotspots or correlate with physiological traits. The darkmagenta module is enriched for the GO term “ER chaperone complex” and the KEGG pathway “Protein processing in ER”. It includes transcripts for Pdia4, Calr, Pdia3, Hspa5, P4hb, Hsp90b1, Sec23b,
Sec61a1, Dnajb11, and Pdia6, all of which are critical for ER-homeostasis. The cyan module contained 261 transcripts, and enriched for GO, “extracellular matrix” and KEGG, “ECM-receptor interaction”. The skyblue3 module is enriched for GO term “NIK/NF-kappa B signaling” and KEGG pathway “Apoptosis”. It contains, Nfkbi1, Nfkbi2, Tnf, Traf2, Bid, Casp3, Fas, Il1a, Nfkbi1, Nfkbia, and Trp53. The lightgreen module enriched for “glucocorticoid receptor binding”, and “Pancreatic cancer”; Nr4a1, Nr4a3, Nr4a2, Stat3, Bcl2l1, Jak1, Pik3r1, Rac1, and Tgfb3.

Three modules (plum1, lightsteelblue, mediumpurple3) with the strongest mod-QTL did not enrich for functional annotations. The plum1 module consists of 91 transcripts, all but one of which correspond to genes located on proximal Chr 14, and mapped to Chr 14 at ~7.5 Mbp, LOD > 90 (Fig. 8). Transcripts within plum1 derive from a large region of the chromosome (3 – 7.5 Mb) that is rich in repeated gene families, harbors segmental variations across the DO founder strains, and is a recombination cold-spot (Morgan et al, 2017). Eighty-two of the 91 transcripts in the plum1 module are gene models (e.g., GM10409, GM3020, GM26552, etc.), or predicted genes that do not have associated physiological annotations. The correlated expression of these transcripts reflects an unusually elevated level of linkage disequilibrium (LD) across the region. The lightsteelblue1 module consists of 42 transcripts, 40 of which are clustered in another recombination cold-spot with segmental variations located on Chr 2 at 177 Mbp; lightsteelblue1 mapped to Chr 2 at ~175 Mbp, LOD = 89. The genes in the lightsteelblue module are mostly gene models with little functional annotation. The MEs for either plum1 or lightsteelblue1 were not significantly correlated with any of the physiological traits (Fig. 7). The mediumpurple3 module also maps to spatially clustered genes, most of which are located on distal chromosome 6; mod-QTL at ~150 Mbp, LOD ~ 9. The mediumpurple3 gene clusters are not in recombination cold-spots, and encode functional RNAs, including the nuclear 5S ribosomal RNA, and several small nuclear RNAs.

There were 9 modules with functional enrichment, but no significant or suggestive mod-QTL. These include the largest module, turquoise, with 1494 transcripts, many involved in mRNA processing and
ribosome function, as well 12 of the 14 mitochondrial-encoded transcripts. The darkgrey module consists of 175 transcripts enriched for “α-amylase activity”, and “Pancreatic secretion”, and includes Amy1, Amy2a5, Amy2a4, Amy2a3, Amy2a2. These transcripts are typically expressed in acinar tissue, and most likely reflect residual contamination in isolated islet preparations and thus we would not expect to find it to be controlled by genetic factors.

**Modules enriched in α-cell and δ-cell specific transcripts**

Two modules were enriched with transcripts selectively expressed in α-cells or δ-cells. These could reflect either increased activity, or increased cell numbers relative to the predominant β-cell component of islets. The greenyellow module consists of 387 transcripts, including Gcg, Irx1, Irx2, Arx, Mafb, Trr, Gria3 and Sstr2. These transcripts have been previously associated with selective expression in α-cells (DORRELL et al. 2011; ACKERMANN et al. 2016; DIGRUCIO et al. 2016; XIN et al. 2016; LAWLOR et al. 2017). The yellowgreen module contains 109 transcripts, including many that have been linked to selective expression in δ-cells; Sst, Hhex, Rbp4 and Ghsr (DORRELL et al. 2011; ACKERMANN et al. 2016; DIGRUCIO et al. 2016; XIN et al. 2016; LAWLOR et al. 2017).

In addition to the cell-type-specific nature of the yellowgreen (δ-cells) and greenyellow (α-cells) modules, both are significantly enriched with one or more GO terms and KEGG pathways (Fig. 6 B & C, respectively). The yellowgreen is enriched for “voltage-gated Ca²⁺-channel activity” and includes transcripts for several Ca²⁺-channels; Cacna1e, Cacna1g, Cacna2d3, Cacna2d2, Cacna1h, Cacna1i, and Trpa1. Greenyellow is enriched for “synapse assembly” (Cbln2, Ephb2, Lrm3, Nrxn2, Nrxn3, Oxtr, Ptprd, Wnt5a, Flrt3, Lrrtm1, Slitrk1, Shank2, Nrg1, Lrrtm3, Lrp4, Adgrb2, and Lrrc4b), which may be related to the control of glucagon secretion by the central nervous system.

Next, we asked if a module is similarly enriched for known β-cell genes. In contrast to the α-cell and δ-cell-enriched modules, key β-cells genes were identified in separate modules, including Ins1 and Ins2 (yellow), Nkx6.1 (black), Pdx-1 and Glp1r (grey60), Slc30a8 and Slc2a2 (pink), Mafa (lightcyan), Ucn3
(red), and G6pc2 and lapp (brown). These results suggest that unlike known β-cell associated transcripts, those in α-cells and δ-cells are highly coordinate in their expression pattern among the islets of the DO mice. The β-cell transcripts do not show this same coordinate regulation, suggesting that they may reflect a greater degree of heterogeneity within the β-cell population, as has been previously reported for islets collected from different regions of the pancreas in mice following dietary challenge (ELLENBROEK et al. 2013), as well for islets collected from non-diabetic vs. T2D human subjects (DORRELL et al. 2016; XIN et al. 2016; LAWLOR et al. 2017).

Transcripts within a module are highly correlated, suggesting that this coordinate regulation is maintained across the DO islets that were profiled. Further, the α- and δ-cell modules showed significant QTL, implying that at least in part, some of this coordinate regulation is driven by factors present at these loci; e.g., Ptprz1 for the δ-cell mod-QTL on Chr 6. In contrast to the α- and δ-cell modules, β-cell specific transcripts were scattered among several different modules, suggesting the β-cell specific transcripts are not as tightly coordinated. We speculate that the discordant regulation of these genes among the DO mice reflects different compensatory responses. There was a >100-fold range of plasma insulin values among the DO mice, presumably reflecting differences in insulin production in response to diet-induced insulin resistance. It is possible these differences reflect greater changes in the β-cell transcriptome, rather than α- or δ-cells. Further, it may be possible that islets collected from a single mouse harbor greater variability in β-cell transcripts than non-β-cell transcripts.

The α-cell (greenyellow) and δ-cell (yellowgreen) enriched modules each showed a decline in transcript abundance in male vs. female mice (Fig. S9). These results suggest that the cellular composition of the DO islets, or the proportion of α, δ and β-cells, may be different between the sexes, with islets from males having fewer non-β cell-types than females. A similar sex difference in the proportion α-cells and δ-cells was reported for islets evaluated from female versus male BSB mice (SLAVIN et al. 2010). A complete list of modules demonstrating differences between the sexes is provided in Table S5.
We identified distinct mod-QTL for the modules associated with α-cells (greenyellow) and δ-cells (yellowgreen) (Fig. 8). The δ-cell module mapped to 4 distinct loci; Chrs 4 at ~3.2 Mbp, 6 at ~4.8, 18 at ~5.4 Mbp, and 18 at ~44.2 Mbp. The α-cell module showed suggestive mod-QTL on Chrs 11 at ~16.8 Mbp, and 15 at 60 Mbp. Unlike the α-cell and δ-cell modules, we did not identify any mod-QTL for the acinar-enriched module (darkgrey). These results suggest that islet cell composition may be under distinct genetic regulation, whereas the acinar component of the isolated islets is driven by non-genetic influences; e.g., efficiency of the collagenase digestion of the pancreas prior to islet isolation.

An interactive web interface enables browsing of mouse islet eQTL

We have constructed a web interface that allows the user to quickly determine if a transcript shows genetic regulation in the DO islets (http://churchill-lab.jax.org:21010/). For example, entering the gene symbol Tcf7l2, a locus that is strongly associated with diabetes in human GWA studies, into the search window, results in a genome-wide LOD profile that identifies two significant ($p < 0.05$) eQTL; Chr 13 at ~112.5 Mbp, and Chr 19 at ~55.4 Mbp (Fig. S10). The gene for Tcf7l2 is located on Chr 19 at ~55.7 Mbp. The allele dependence for the local-eQTL (Fig. S11A) is driven by alleles from CAST (low) and B6 (high), whereas the distal-eQTL (Fig. S11B) is associated with NZO (high), suggesting that the genetic dependence of each of the eQTL for Tcf7l2 is distinct.

Mediation shows that the local-eQTL for Tcf7l2 is, as expected, driven by Tcf7l2 expression (Fig. S12A), whereas the distal-eQTL appears to be mediated by the expression of Il6st (Fig. 12B), the driver predicted from our mediation analysis for the Chr 13 eQTL hotspot (Fig. S4). The transcript for Tcf7l2 is included in the royalblue module that enriched for GO “positive regulation of cytokine production” (Fig. 6A) and KEGG “Toll-like receptor signaling pathway” (Fig. 6B), was positively correlated with plasma insulin at 10 weeks of age (Fig. 7) and mapped most strongly to Chr 13 at ~111.6 Mbp (Fig. 8). These tools demonstrate how the user can begin with a gene of interest, determine if it shows distal vs. local genetic regulation, and integrate the findings with the physiological pathways associated with the gene modules, as well as the genetic architecture of the modules and
individual transcripts that we have identified. SNPs at each eQTL that are associated with the expression of Tcf7l2, can be downloaded and surveyed for potential consequences, e.g., missense, frameshift, intronic, splice junction, etc.; **Fig. S13A** (Chr 19) and **Fig. S13B** (Chr 13). In addition to eQTL for individual transcripts, the web site allows the user to survey LOD summaries, allele effect and SNP association plots for physiological phenotypes (e.g., HOMA-B, number of islets per mouse), eQTL hotspots, and module eigengenes. This interactive resource enables a user to discover regulatory loci linked to the expression of any gene expressed in pancreatic islets, and to explore potential candidate drivers and downstream physiological pathways.
Discussion

We present, for the first time, the genetic architecture of gene regulation in pancreatic islets from mice subjected to a Western-style diet high in fat (45%-kcal) and sucrose (34%). We used Diversity Outbred mice, which contain a high level of genotypic and phenotypic diversity; a level comparable to that present in the human population. The study in humans most closely resembling our mouse experiment was reported by Groop and colleagues (TANEERA et al. 2012). They obtained pancreatic islets from 63 human cadaveric donors, including 6 who had type 2 diabetes (T2D), and interrogated 48 candidate genes located near SNPs associated with T2D in human GWA studies. The expression of KCNJ11, WFS1, SLCA2A, JAZF1, and G6PC2 was decreased in islets from the T2D donors. They identified 5 cis-eQTL, however, none showed significant expression differences in islets from T2D vs. normal subjects.

We quantitated the abundance of >21,000 mRNA transcripts and identified nearly 19,000 eQTL (genome-wide $p < 0.05$) (Fig. 3). More than 70% of these are local-eQTL. In a previous islet genetic study (TIAN et al. 2015; TIAN et al. 2016), we observed a preponderance of distal-eQTL, and believe the difference is related to a higher LOD threshold in our current study; local-eQTL usually have higher LOD scores than distal-eQTL. We identified five distal-eQTL hotspots and applied mediation analysis to identify candidate causal drivers. Two of these drivers are genes known to be involved in monogenic forms of diabetes, HNF4A (MODY1) and PDX1 (MODY4) (Fig. 5). Genetic variation in the MODY genes has also been shown to play a role in type 2 diabetes (NDIAYE et al. 2017).

We chose Pdx1 as a candidate driver at the Chr 5 eQTL hotspot for several reasons: 1) The LOD drop upon mediation against Pdx1 ranked 2nd among the other local-eQTL at the locus (Fig. S5C); 2) The allele effect pattern for the Pdx1 local-eQTL (Fig. S5E) was the closest match to that for the Chr5 hotspot PC1 (Fig. S5B) with NOD, and B6/129 as the high and low alleles, respectively; and 3) The other candidate mediators have very strong local-eQTL LOD scores. This is one of the pitfalls of mediation; a strong local eQTL can act as a surrogate for genotype, making it difficult to rule out the
eQTL gene as a candidate mediator. The logic of mediation analysis is designed to rule out candidate mediators, so when there are several candidates, we consider other evidence including, the allele effects and functional plausibility of the candidate.

In addition to the MODY genes, we identified three novel drivers of eQTL hotspots, \textit{Il6st} (Chr 13), \textit{Sat2} (Chr 11), and \textit{Fam83e} (Chr 7), and demonstrate associations with diabetes-related traits in human GWAS when we restrict the association testing to the region defined by synteny with the mouse eQTL (Fig. 5). \textit{Il6st}, also known as glycoprotein 130 (gp130), is the \(\beta\)-subunit of the IL-6 cytokine receptor. A recent report showed that \textit{Il6st} is required for IL-6-mediated activation of STAT3, and stimulation of glucagon secretion (Chow et al. 2014). Further, \(\alpha\)-cell-specific deletion of \textit{Il6st} resulted in protection from Streptozotocin-induced diabetes (Chow et al. 2014), suggesting that in the context of inflammation that accompanies T2D, \textit{Il6st} may promote hyperglycemia by stimulating glucagon secretion from \(\alpha\)-cells. \textit{Il6st} is in the \textit{royalblue} module, along with transcripts that are significantly enriched for the GO term, “positive regulation of cytokine production”. The FAM83 family consists of eight members, several of which are increased in expression or copy number in various cancers, and are downstream of the phosphatidylinositol 3-kinase and epidermal growth factor receptor pathways (Snijders et al. 2017).

Both HOMA-IR, a measure of insulin resistance, and HOMA-B, a measure of insulin production, map to a common locus on Chr 11 at \(\sim 85\) Mpb (Fig. 2). Several modules also mapped to this locus, including those enriched in cell cycle, branched-chain amino acid catabolism, and NIK-NF-Kappa B signaling (Fig. 8). Mediation analysis consistently identified \textit{Sat2} as the driver gene for this locus (Fig. S7). \textit{Sat2} encodes Spermidine/Spermine N1-acetyltransferase 2 and is included in the BCAA module. Indeed, \textit{Sat2} has been shown to function with p65 as a co-activator of NF-KB (Vogel et al. 2006). Despite its 61% homology with \textit{Sat1}, \textit{Sat2} does not appear to have significant activity towards polyamines. Instead, it catalyzes the transfer of an acetyl group to thialysine residues, forming N\(^\text{\textsuperscript{\textth}}\)-thialysine (Coleman et al. 2004).
We measured glucose tolerance in all mice, as well as insulin during the glucose tolerance test. Essentially all of the mice were non-diabetic throughout the course of our study, and yet showed a >100-fold range in plasma insulin, suggesting that the dietary challenge elicited a large variation in the physiological responses required to maintain plasma glucose. Insulin, AUC\textsubscript{insulin}, HOMA-IR, and body weight were all positively correlated with a module enriched in BCAA catabolism (Fig. 7). These same phenotypes were negatively correlated with ribosome biogenesis. Thus, protein metabolism, synthesis and turnover, were the strongest correlates. These results are consistent with a growing body of evidence linking branched chain amino acids with metabolic disease (NEWGARD 2012). In contrast, plasma glucose and triglycerides were not strongly correlated with any of the modules, suggesting that these traits are effectively buffered by compensatory variation in insulin levels.

Module-QTL have been identified for hepatic gene expression networks that were shown to be highly correlated, and to co-map with body weight in mice (GHAZALPOUR et al. 2006; FULLER et al. 2007). Other groups have used the WGCNA approach to identify module-QTL in liver that are associated with HDL-cholesterol (LEDUC et al. 2012), or adiposity and hepatic steatosis (DAVIS et al. 2012), and in heart for cardiac left ventricular mass (SCOTT-BOYER et al. 2014). Our study is the first to apply the WGCNA module-QTL approach to pancreatic islets.

Two of the co-expression gene modules we identified appear to reflect the activity, or relative proportion of α-cells and δ-cells in the islets. In contrast, no single module was enriched for transcripts selectively expressed in β-cells. This could be an indication that there is more heterogeneity in the β-cell population of islets than there is for α- and δ-cells, which may have been triggered by the metabolic challenge imposed by the HF/HS diet, requiring increased insulin production in some, but not all mice. The heterogeneity could be due to sub-types of β-cells (DORRELL et al. 2016), or to plasticity of the cells i.e. their ability to transdifferentiate to other cell types (JONAS et al. 1999; PAPIZAN et al. 2011; TALCHAI et al. 2012; WANG et al. 2014).
We identified mod-QTL for the δ-cell-enriched module (yellowgreen) (Fig. 8) on Chrs 4, 6, and 18. None of the genes most strongly associated with selective expression in δ-cells, and included in the yellowgreen module (Hhex, Sst, Rbp4, and Ghsl), are physically located at these QTL. However, mediation analysis for these genes consistently identified Ptprz1 as a candidate driver for the chromosome 6 locus, and Armc4 as a candidate driver for the chromosome 18 locus, suggesting that these genes control the expression of the δ-cell-identity module. Ptprz1 is included in the yellowgreen module and is expressed >10-times higher in δ-cells than α-cells or β-cells (DiGRUCCIO et al. 2016), consistent with it playing a key role δ-cells. Ptprz1 is a receptor for the heparin-binding glycoprotein pleiotrophin (PTN), and has been associated with several cancers, including glioblastoma (SHI et al. 2017), and pancreatic cancer (XUE et al. 2017). In addition to δ-cell-specific transcripts, six genes that encode Ca\(^{2+}\)-channels were included in the yellowgreen module, including R-type Ca\(^{2+}\) channels (e.g., Cacna1e), which have been previously linked to glucose-stimulated somatostatin secretion (ZHANG et al. 2007).

We identified mod-QTL for the α-cell-specific module (greenyellow) and, as for the δ-cell module, the mod-QTL loci were distinct from the genes that encode the α-cell-specific transcripts. Mediation analysis of the α-cell-specific transcripts did not reveal consistent candidates as driver genes. This α-cell module enriched for synapse formation, usually associated with the central nervous system, and may reflect the strong neural connection with α-cell function or development.

We compared the chromosomal locations of the QTL for the diabetes-related phenotypes (Fig. 9A) and islet module pathways (Fig. 9B), with their corresponding locations in the human genome, and observed a significant enrichment \((p = 0.003)\) for diabetes-associated GWA peaks at threshold of \(\log_{10} p \geq 4\). This observation supports the hypothesis that the genetic drivers of diabetes risk are concordant across mice and humans. In other words, not only are the same pathways involved, but our findings suggest that genetic variation in the same gene loci is the underlying cause for the difference in diabetes incidence and susceptibility.
Our project allows one to nominate novel loci in the human genome based on associations with diabetes-related phenotypes, or physiological pathways in pancreatic islets identified in mouse. For example, the mod-QTL we identified for the δ-cell module on Chr 6 at ~4.8 Mbp, is syntenic to Chr 7 at ~94.7 Mbp in human (Fig. 9B). Our mediation analysis strongly suggests that \textit{Ptprz1} is a driver of this mod-QTL. Using the gene search tool at GWAS central (http://www.gwascentral.org/generegion), one can then determine if any SNPs have been identified that are proximal to \textit{PTPRZ1} and associated to diabetes-related traits. The results show that SNPs associated with HbA1c levels have been reported at this locus (STRACHAN et al. 2007). Similarly, the cell cycle module (violet) has a mod-QTL on Chr 11 at ~69.8 Mbp. In human, this region is syntenic to Chr 17 at ~7.3 Mbp, which is proximal to the gene \textit{TMEM102}, where SNPs have been identified that are associated with T2D (SLADEK et al. 2007).

Finally, one can ask if QTL identified for a diabetes-related phenotype in our mouse study, highlight a novel locus in human. For example, among the diabetes-related physiological traits in our study, HOMA-B and AUC\textsubscript{insulin} demonstrated the strongest genetic signal, whereby both mapped to a locus on Chr 11 at ~84 Mbp (Fig. 9A) with LODs > 8. In human, this region is syntenic to Chr 17 at ~60.1 Mbp. Interestingly, at this locus SNPs have been identified that are associated glucose levels 2 hr after an oral glucose challenge (SAXENA et al. 2010).

We applied a systems genetics approach that used gene expression traits to interrogate the function of pancreatic islets under a diabetogenic stressor and identified groups of co-mapping transcripts as well as co-expression modules. When we performed mediation analysis to the co-mapping transcripts, we identified strong candidate genes as likely drivers. Modules provide an alternative view of the coordinated expression patterns in the islet cells; functional enrichment analysis attaches biological interpretation to these modules and are often correlated with physiological measures of diabetes-associated traits. Genetic mapping of the module eigen-genes reveals multiple QTL peaks with modest LOD scores reflecting heterogeneity in the genetic drivers of the module. Mediation analysis of module QTL was less successful at identifying candidate driver genes, however, mediation of individual genes
within the module can be effective. Our web-based analysis tools can support this kind of data exploration.
ACKNOWLEDGEMENTS

The project was supported by grants from the NIH to A.D.A (DK101573, DK102948), C.K. (GM102756 and U54AI117924), G.A.C and K.W.B (GM076483), and the Clinical and Translational Science Award (CTSA) program, through the NIH National Center for Advancing Translational Sciences (NCATS), grant UL1TR000427. We are grateful for graphical expertise provided by Laura Vanderploeg.
**Figure legends**

**Figure 1: Diabetes-related physiological phenotypes in DO mice.**

Body weight (A), and fasting plasma glucose (B), insulin (C) and triglycerides (TG) (D) are shown for female and male HF/HS diet-fed DO mice at 6, 10 and 14 weeks of age. HOMA-IR (E) and HOMA-B (F), measures of insulin resistance and β-cell function, respectively, were computed from fasting plasma insulin and glucose measurements at 18 weeks of age. Area under the curve (AUC) for glucose (G) and insulin (H) were determined from an oral glucose tolerance test (oGTT) performed at 18 weeks of age. The number of islets per mouse (I), and the insulin content per islet (J) were determined at sacrifice at 22 weeks. Whole pancreas insulin content (WPIC), a surrogate measure of pancreatic β-cell mass, was computed from islet number and insulin content per islet (K). Average food consumption (gm/d) was computed over the course of the 4 months the mice were maintained on the HF/HS diet (L). *, p < 0.05 for male vs. female; **, p < 0.01.

**Figure 2: Genetic architecture of diabetes-related phenotypes.**

QTL for diabetes-related physiological traits, colored by LOD value (red = high, yellow = low). Traits are ordered by unsupervised clustering such that traits with similar LOD profiles are grouped together. Table S6 lists all physiological QTL, their LOD scores and genomic position in mouse and human.

**Figure 3: Genetic architecture of gene regulation in pancreatic islets.**

A) Inferred eQTL with LOD ≥ 7.18 (p < 0.05) identified in islets from 378 DO mice maintained on HF/HS diet. Black, distal-eQTL (5,669); blue, local-eQTL (12,802). Data points correspond to the peak position of the eQTL. Y-axis shows position of gene, while x-axis shows the genomic position of the eQTL. Local-eQTL follow a diagonal pattern, whereas the distal-eQTL form vertical bands. Profile illustrating the number of distal (B) or local (C) eQTL occurring within a 4 Mbp genomic window. Distal-eQTL hotspots with >100 co-mapping eQTL were identified in chromosomes (Chr) 2, 5, 7, 11 and 13 (see, ▼).
Figure 4: Mediation predicts *Hnf4a* as a driver at the Chr 2 eQTL hotspot.

**A**) Genome-wide LOD profile for 1\textsuperscript{st} principal component (PC1) of eQTL hotspot at \~165 Mbp on Chr 2. **B**) Allele dependence of Chr 2 hotspot, showing WSB and NZO are the high and low alleles, respectively. **C**) LOD score for Chr 2 eQTL hotspot after conditioning, one at a time, on the expression of 1,892 genes that were located on Chr 2. Conditioning on *Hnf4a* local-eQTL resulted in the largest drop in the LOD profile for hotspot. **D**) LOD profile for *Hnf4a* local-eQTL with a peak at 163.9 Mbp on Chr 2. **E**) Allele dependence for *Hnf4a* local-eQTL demonstrates the same genetic architecture as the eQTL hotspot; WSB and NZO are the low and high allele respectively.

**Figure 5. Candidate driver gene loci are associated with diabetes traits in human GWAS.**

Association (-log\textsubscript{10} \textit{p}-value, left margin) to diabetes-related phenotypes for SNPs near candidate gene drivers of eQTL hotspots. **A**) 2 hr glucose adjusted for BMI at *HNF4A* gene locus. **B**) Fasting plasma glucose at *PDX1* gene locus. **C**) Fasting plasma insulin at *SAT2* gene locus. Hemoglobin A1c (HbA1C) at the *IL6ST* (**D**) and *FAM83E* (**E**) gene loci. Plots were generated using *LocusZoom* (PRUIM et al. 2010) from publicly available GWAS data. For each plot, the index SNP (purple diamond) is identified by Rs number, along with SNPs that are correlated to the index SNP (color scale shows correlation, \textit{r}\textsuperscript{2}), defining a haplotype block. The total number of SNPs plotted are 359, 324, 289, 315, and 217 for **A** – **E**, respectively. Recombination frequencies (cM/Mb) are plotted as blue trace and is shown along right-margin for each locus. Legend for \textit{r}\textsuperscript{2} values is shown in panel **B** and corresponds to all panels.

**Figure 6: Co-expression gene modules in islets from DO mice.**

Cluster dendrogram illustrates modules (denoted by color and shown as downward branches) of highly correlated transcripts in islets from 378 DO mice that were maintained on HF/HS diet (**A**). Transcripts falling within grey colored areas were not sufficiently correlated to be included in a module; \~62% of 21,771 transcripts were assigned to a module. Gene ontology (GO) (**B**) and KEGG pathway (**C**) enrichments for all modules. Modules are ranked by Z-score, highlighting those with the most
significant enrichment. Top GO/KEGG terms are listed for each module. **Table S2** provides a list of transcript membership for each module. **Table S3** contains all significantly enriched (Z-score > 3) GO/KEGG terms for the modules.

**Figure 7: Correlation between islet gene modules and physiological phenotypes.**

Heat map illustrates the Spearman’s nonparametric correlation between islet MEs and diabetes-related physiological phenotypes, after adjustment for sex and batch. Module names are shown along the bottom axis, and top-enriched GO terms along the top axis. Modules are ordered by unsupervised hierarchical clustering, resulting in them grouping by correlation patterns to the physiological traits. Red, positive correlation; blue negative correlation.

**Figure 8: Genetic architecture of islet gene modules.**

Mod-QTL are illustrated for all modules, colored by LOD value (red = high, yellow = low). Top-enriched GO term for each module is shown; three modules did not significantly enrich for any terms (*plum1, lightsteelblue1, mediumpurple3; NA*). Modules are ordered by unsupervised clustering such that traits with similar LOD profiles are grouped together (e.g., Chr 11). **Table S7** lists all mod-QTL, their LOD scores and genomic position in mouse and human.

**Figure 9. Architectural view of diabetes-related physiological phenotypes and islet-based pathways in mouse and human.**

QTL for diabetes-related phenotypes (A) and physiological pathways (B) are illustrated for the mouse genome, which are connected to their syntenic regions in human. The gene nearest to the peak for the QTL in mouse was used to identify the corresponding locus in human. The thickness of arc originating from the mouse QTL is proportional to the 95% confidence interval for the QTL. Arcs interconnect mouse QTL to syntenic regions in human; numbers listed next to human chromosomes are Mbp position. The color of the arcs in B correspond to the module color; top-enriched physiological pathways for modules are illustrated in legend for B. Chr numbers are shown; the Y-chromosome is
excluded. The number of SNPs from human GWAS (http://www.gwascentral.org/) with nominal or higher associations (-log₁₀ p ≥ 4) to T2D (green) or T1D (purple) that occur within a 1 Mbp interval is shown for each human Chr, resulting in the identification of diabetes GWAS loci. The genomic location is shown for the five candidate gene drivers predicted from our mediation analysis of the eQTL hotspots; *IL6ST, PDX1, SAT2, FAM83E,* and *HNF4A.*
References


Hyvonen, M. T., J. Weissell, A. R. Khomutov, L. Alhonen, J. Vepsalainen et al., 2013 Metabolism of triethylenetetramine and 1,12-diamino-3,6,9-triazadecane by the spermidine/spermine-N(1)-acytlytransferase and thialysine acetyltransferase. Drug Metab Dispos 41: 30-32.


Replication, D. I. G., C. Meta-analysis, C. Asian Genetic Epidemiology Network Type 2 Diabetes, C. South Asian Type 2 Diabetes, C. Mexican American Type 2 Diabetes et al., 2014 Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet 46: 234-244.


van de Bunt, M., J. E. Manning Fox, X. Dai, A. Barrett, C. Grey et al., 2015 Transcript Expression Data from Human Islets Links Regulatory Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream Effectors. PLoS Genet 11: e1005694.


Xin, Y., J. Kim, H. Okamoto, M. Ni, Y. Wei et al., 2016 RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes. Cell Metab 24: 608-615.


