

Systematic Functional Characterization of Candidate Causal Genes for Type 2 Diabetes Risk Variants

Short running title (less than 47 chars): A genetic screen for candidate causal T2D genes

Soren K Thomsen¹, Alessandro Ceroni², Martijn van de Bunt^{1,3}, Carla Burrows¹,
Amy Barrett¹, Raphael Scharfmann⁴, Daniel Ebner², Mark I McCarthy^{1,3,5}, Anna L Gloyn^{1,3,5}

1) Oxford Centre for Diabetes Endocrinology & Metabolism, University of Oxford, UK

2) Target Discovery Institute, University of Oxford, Oxford, UK

3) Wellcome Trust Centre for Human Genetics, University of Oxford, UK

4) INSERM U1016, Institut Cochin, Université Paris Descartes, Paris, France

5) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK

Address for Correspondence

Professor Anna L Gloyn

Oxford Centre for Diabetes Endocrinology & Metabolism

Churchill Hospital, Oxford, OX3 7LE, UK

Tel: +44 1865 857219

Email: Anna.Gloyn@drl.ox.ac.uk

Abstract

Most genetic association signals for type 2 diabetes risk are located in non-coding regions of the genome, hindering translation into molecular mechanisms. Physiological studies have shown a majority of disease-associated variants to exert their effects through pancreatic islet dysfunction. Systematically characterizing the role of regional transcripts in β -cell function could identify the underlying disease-causing genes, but large-scale studies in human cellular models have previously been impractical. We developed a robust and scalable strategy based on arrayed gene silencing in the human β -cell line EndoC- β H1. In a screen of 300 positional candidates selected from 75 type 2 diabetes regions, each gene was assayed for effects on multiple disease-relevant phenotypes, including insulin secretion and cellular proliferation. We identified a total of 45 genes involved in β -cell function, pointing to possible causal mechanisms at 37 disease-associated loci. The results showed a strong enrichment for genes implicated in monogenic diabetes. Selected effects were validated in a follow-up study, including several genes (*ARL15*, *ZMIZ1* and *THADA*) with previously unknown or poorly described roles in β -cell biology. We have demonstrated the feasibility of systematic functional screening in a human β -cell model, and successfully prioritized plausible disease-causing genes at more than half of the regions investigated.

Type 2 diabetes risk is determined by a complex interplay between environmental and genetic factors, with heritability estimates ranging from 20-80% (1). Over the past decade, genome-wide association studies (GWAS) of ever-increasing size have discovered more than a hundred regions of the genome (loci) associated with type 2 diabetes risk (2). Studies in non-diabetic individuals have demonstrated that a large number of these association signals exert their effects on disease susceptibility through pancreatic islet dysfunction (3).

Despite these advances, progress in translating genetic findings into disease biology has been relatively slow. The majority of risk variants are located in non-coding regions of the genome, and pinpointing the underlying causal genes or “effector transcripts” has proved challenging (4). Recent efforts have focused on identifying structural or functional links between association signals and regional genes (5, 6). A complementary strategy uses candidate-gene biology to prioritize genes located near association signals. High-throughput screening (HTS) could facilitate the identification of genes implicated in β -cell function, and thereby highlight potential effector transcripts at type 2 diabetes GWAS loci. To date, such approaches have been limited by inadequacies of available human cellular models and the high cost of insulin immunoassays (~\$2 per data point), the gold standard for measuring insulin. To circumvent these issues, previous studies have relied on rodent β -cell models and either used reporter assays as a proxy for insulin measurements, or focused on cellular proliferation (7-11).

Recently, the first glucose-responsive human β -cell line, EndoC- β H1, was generated (12, 13). The line is derived from fetal pancreatic buds matured *in vivo*, and displays modest but robust induction of insulin secretion in response to glucose and secretagogues. Detailed characterizations have shown the cell line to be an authentic model system for studying stimulus-coupled secretion (14-16).

To accelerate the discovery of causal genes for type 2 diabetes, the present study performed and validated a genetic screen in the EndoC- β H1 cell line. We identified genes at half of the type 2 diabetes-associated loci studied (37/75) where siRNA-mediated silencing resulted in β -cell dysfunction. This demonstrates the feasibility of performing systematic screening for insulin secretion in a human β -cell model, with implications for both high-throughput genetic and chemical compound screening. Our results can be integrated with existing lines of evidence to prioritize effector transcripts at GWAS loci, and highlight potential roles for *ARL15*, *ZMIZ1* and *THADA* in the regulation of insulin secretion.

Research Design and Methods

RNA-seq

The EndoC- β H1 cell line was cultured as previously described and grown to near confluency (12). RNA was then TRIzol-extracted, and sequenced at the Oxford Genomics Centre (Wellcome Trust Centre for Human Genetics, University of Oxford) (see **Fig. S4** for details). The raw sequencing data have been deposited at the European Nucleotide Archive (ENA; <http://www.ebi.ac.uk/ena>) under accession number XXXXXXXXXXXX.

Cellular assays

Cellular phenotypes were adapted for automated screening on a Perkin Elmer Janus liquid handling workstation based on previously described assays (**Fig. S1A**) (17). Briefly, 20,000 cells/well were reverse transfected in 96-well format at final siRNA concentrations of 25 nmol/L pre-incubated with 0.2 μ L RNAiMAX in Opti-MEM. Custom libraries of siRNAs (ON-TARGETplus SMARTpools [Dharmacon] for the primary screen and Silencer Select [Thermo Fisher Scientific] for follow-up validation) were designed based on criteria described in table S2. In each case, non-targeting sequences based on the same chemistries were used as negative controls. Three days after transfection, cells were starved overnight in complete media containing 2.8 mmol/L glucose followed by 1 h starvation in 0 mmol/L media. Static insulin secretion assays were then performed for 1 h in complete media under the indicated conditions, after which cells were counted as described below.

Sample analysis

Following secretion assays, supernatants were analyzed for insulin content using AlphaLISA Human Insulin Immunoassays (Perkin Elmer) on a PHERAstar FS plate-reader (BMG). Supernatants (50-250 nL) and beads pre-diluted in water (500 nL) were dispensed into 384-shallow well microplates with an Echo 550 (Labcyte) acoustic liquid handler, before manual addition of immunoassay buffer to a final volume of 5 μ L. Cell counts were measured using the CyQUANT Direct Cell Proliferation kit (Thermo Fisher) on an EnVision plate-reader (Perkin Elmer). All responses were normalized as indicated (see relevant figure legends), and expressed as a percentage of non-targeting (NT) control for each phenotype. Effect sizes are given as the percentage difference from NT ($Response_{Gene} - Response_{NT}$), and the absolute values hereof ($|Response_{Gene} - Response_{NT}|$).

Statistical analysis

Data analysis was performed using R 3.0.2. To identify significant responses, cell counts and normalized insulin secretion measurements for each gene were compared to NT control using Student's two-sample t-test. The false-discovery rate (FDR) was controlled at 5% by applying the Benjamini-Hochberg procedure to produce adjusted p-values (q-values) for each phenotype. The Z-factor measuring the control response for each phenotype was calculated

$$\text{as } Z' = 1 - \frac{3(\sigma_{INS} + \sigma_{NT})}{|\mu_{INS} - \mu_{NT}|}.$$

Results

We first developed an automated assay for disease-relevant phenotypes in the human β -cell line EndoC- β H1 (**Fig. S1A**). Selected targets were silenced in a parallel format using RNA interference (RNAi). Cells were then assessed for effects on cell number and insulin secretion under four different conditions: low glucose (1 mM), high glucose (20 mM), and high glucose with the sulphonylurea tolbutamide (100 μ M) or with the phosphodiesterase inhibitor IBMX (100 μ M). Low and high glucose conditions were included to provide information on the effect of gene silencing under conditions representing the fasted and fed states *in vivo*. Tolbutamide and IBMX act on the depolarizing and the potentiating pathways of insulin secretion, respectively, and were included to provide additional mechanistic insights through modulation (e.g. synergy or pharmacologic rescue) of any primary defects observed in low or high glucose.

To reduce the cost of sample analysis, we made use of acoustic liquid handling to miniaturize insulin immunoassays. This generalizable method enabled us to maintain high sensitivity for insulin measurements (coefficient of variation [CV] < 3%; **Fig. S2A-B**), while obtaining a ten-fold reduction in the cost of sample analysis (\$0.20 per data-point). Using the insulin gene (*INS*) as a positive control, we confirmed that we were able to robustly detect effects of gene silencing on the phenotypes of our assay (mean $Z' = 0.6$ across conditions; **Fig. S3**).

Based on this combined analysis and assay pipeline, we designed a primary screen to assess the role of positional candidate genes for type 2 diabetes GWAS loci in β -cell function (**Fig. S1B**). For target selection, we considered all protein-coding genes located within 1 Mb of a type 2 diabetes association signal. To exclude genes not expressed in our cellular model, we performed whole-genome RNA sequencing of the EndoC- β H1. Our expression data

strongly correlated with published sequencing data for enriched primary β -cells ($\rho = 0.78$; **Fig. S4**) and showed robust expression of key β -cell genes (12, 18) (**Table S1**). We included only genes expressed in both EndoC- β H1 and primary β -cells (**Table S2**), resulting in inclusion of 300 positional candidates from 75 type 2 diabetes GWAS loci.

We next performed our primary screen in triplicate, and derived standardized scores for each phenotype. Knockdown was visibly confirmed using *PLK1*, an essential gene, which caused extensive cell death across all conditions. In a representative subset of 16 genes we assessed knockdown efficiency at the transcript level and found the median residual expression to be 43% (**Fig. S5**), roughly equivalent to monoallelic loss-of-function. To account for differences in plating efficiency and proliferation, cell counts were used to normalize insulin secretion data on a per-well basis. Two criteria were then applied to identify robust effects (“hits”): (1) an FDR-adjusted q-value < 0.05 ; and (2) an absolute effect size among the top 5% (**Fig. S6**). This identified a total of 67 hits (15 for cell count and 52 for insulin secretion phenotypes) between 45 genes at 37 loci (**Table 1**).

For cell numbers, effect sizes for each gene were estimated based on 12 independently plated replicates (four conditions in triplicate), and therefore likely represent true differences in cellular proliferation and/or viability rather than random plating effects. Aside from *KIF11*, a gene with a known role in cell division, the largest effect sizes compared with NT control (CV = 4% for cell numbers) were observed for *ZMIZ1* (-15.2%; $q = 6.5 \times 10^{-5}$) and *PRDX3* (+16.5%; $q = 9.2 \times 10^{-5}$).

For the insulin secretion data, we first performed an enrichment analysis for genes implicated in maturity-onset diabetes of the young (MODY). MODY describes a collection of monogenic subtypes of diabetes, characterized by insufficient release or production of insulin. As would be expected for a set of *bona fide* regulators of β -cell function, we observed a strong

enrichment of MODY genes among the significant hits (Fisher's exact test, $p = 5.5 \times 10^{-9}$). Aggregating absolute effect sizes for MODY and non-MODY genes revealed this enrichment to be driven by altered insulin secretion, and not through effects on cell numbers (**Fig. 1**).

Further validating our secretion data, we observed strong positive correlations between the normalized responses across conditions (p -values $< 2.2 \times 10^{-16}$; **Fig. 2**), and found 10 of 35 genes to cause significant effects under two or more conditions. This included four known MODY genes and *ZMIZ1*, which, independently of the effect on cell numbers, was one of the strongest hits for reduced insulin secretion ($q < 0.01$ for low and high glucose). Knockdown of the *ABCC8* gene, which encodes a subunit of the ATP-sensitive potassium channel, was found to significantly increase insulin secretion under low glucose and IBMX stimulation. As expected, the depolarization caused by this was masked under high glucose (as cells are already partially depolarized), and fully rescued by tolbutamide (due to pharmacological depolarization of the cells). The pattern of modulation by secretion conditions can thus be used to pinpoint specific biological pathways affected by gene silencing. To explore the relationship between conditions in greater detail, we performed clustering analysis on Z-scores derived from the normalized secretion values. This revealed high glucose and tolbutamide to be most similar in terms of modulating knockdown effects, with low glucose and tolbutamide being most dissimilar (**Fig. S7**).

Finally, we assessed the contribution of off-target effects by performing a small-scale validation experiment using siRNAs designed with an alternate algorithm. The sequences were confirmed to be different from those of the primary screen, and could thus be used to establish the biological relevance of positive hits. We selected eight target genes, representing hits for both positive and negative defects across the four conditions, and confirmed that knockdown efficiency was satisfactory (median residual expression = 19.3 %; **Fig. S8**). Compared with insulin secretion results from the primary screen, we observed an excellent linear correlation

($\rho = 0.85$, $p = 6.7 \times 10^{-10}$, **Fig. S9**) and 88% directional consistency in normalized responses. The validated hits included several genes with limited prior evidence of a role in the regulation of β -cell function, including; *ARL15* and *ZMIZ1*, which were found to significantly reduce insulin secretion across conditions (q-values < 0.05 ; **Fig. 3A-B**), and *THADA*, which modestly elevated insulin secretion across three conditions, though the effect under low glucose was not observed in the primary screen (q = 5.6×10^{-3} ; **Fig. 3C**). Interestingly, gene silencing of the known MODY gene *HNF4A* was confirmed to cause a paradoxical increase in insulin secretion across all four conditions tested (q-values < 0.001 , **Fig. 3D**).

Discussion

High-throughput screens for β -cell dysfunction offer the opportunity to systematically characterize the role of genes in a disease-relevant tissue for type 2 diabetes. Previous efforts have focused on non-human model systems (7-10), reporter-based proxy measurements for insulin (7, 8), and/or phenotypes not directly related to insulin production and secretion (10, 11). Here, we report a genetic screening strategy for the interrogation of multiple disease-relevant phenotypes in the human β -cell line EndoC- β H1. In a primary screen of 300 positional candidates, we successfully identified 15 genes regulating cell number (proliferation and/or viability), and 35 genes regulating insulin secretion. This is, to our knowledge, the first systematic, large-scale effort to identify genes involved in insulin secretion. Importantly, the identified hits can be used to prioritize novel effector transcripts for type 2 diabetes GWAS loci, and may shed further light on mechanisms underlying genes previously implicated in β -cell dysfunction.

The known MODY gene *HNF4A* was unexpectedly observed to cause a consistent increase (> 40%) in insulin secretion across all conditions. *HNF4A* encodes the transcription factor Hepatocyte nuclear factor 4 alpha (HNF4 α), and is mutated in about 10% of all MODY cases (19). *HNF4A* loss-of-function mutations that cause monogenic diabetes later in life have also been associated with increased birth-weight (indicative of increased fetal insulin secretion) and congenital hyperinsulinism in early infancy (20). The underlying reason for this switch from elevated to reduced insulin secretion is unknown, but it has been speculated that gradual β -cell exhaustion or, alternatively, a shift in the modulating co-factors of HNF4 α may underlie this phenomenon (21, 22).

Among the hits with limited prior evidence of a role in β -cell function, we independently validated *ZMIZ1*, *ARL15*, and *THADA*. Overexpression and knockdown of

ZMIZ1, encoding Zinc Finger, MIZ-Type Containing 1 (ZMIZ1), has recently been shown to negatively impact on insulin secretion in primary human islets (6). Moreover, a nearby type 2 diabetes association signal overlaps a cis-eQTL for the gene, supporting its candidacy as the regional effector transcript (6). *ARL15* encodes ADP-Ribosylation Factor-Like 15, a relatively uncharacterized member of the ARF-family of proteins involved in regulation of vesicle trafficking and biogenesis. The gene is highly expressed in β -cells and located downstream of an islet-active enhancer bound by key β -cell transcription factors (18, 23) (**Fig. S10A**). *THADA* encodes the protein Thyroid Adenoma-Associated (THADA), and contains a coding disease-association signal that has also been associated with reduced β -cell function (24) (**Fig. S10C**). Consistent with the directionality of our findings, expression profiling has shown the gene to be more highly expressed in patients with type 2 diabetes compared with controls (25). All three genes thus emerge as strong candidates for future studies.

While successfully enabling unbiased functional characterization, our current screening strategy has a number of limitations. False negatives (i.e. true causal genes not identified as hits) could arise as a result of primary effects of the causal gene on non-beta cell tissues, or through effects on genes expressed at different developmental stages. Likewise, overexpression or greater knockdown efficiency may in some cases be required to expose a disease-relevant phenotype. Among the targets analyzed for silencing efficiency, a variable range of knockdown was observed (34% - 88%), and some genes might remain undetected due to insufficient silencing. Conversely, false-positive effects (i.e. non-beta cell regulators identified as hits) also cannot be excluded, and unexpected findings should be further functionally validated (e.g. *SLC2A4* effect on IBMX-stimulated insulin secretion). Though the EndoC- β H1 cell line has been found to recapitulate many aspects of β -cell function, it remains a possibility that some findings would not translate directly into human physiology. Finally, a subset of the identified hits may represent true β -cell regulators that are independent of any disease risk variants and,

though still of biological importance, not genuine effector transcripts for type 2 diabetes. In addition to the possibility of more than a single effector transcript per locus, this phenomenon likely also contributes to the relatively high proportion of multi-hit loci observed in the primary screen (8/37).

Despite these limitations, our screening strategy successfully replicated well-established biological mechanisms, and identified genes involved in β -cell function at half of the loci investigated. This demonstrates, for the first time, the feasibility of performing scalable screens for insulin secretory defects in human pancreatic β -cells, and opens up the possibility, not only for large-scale genetic manipulations, but also for compound HTS to therapeutically manipulate human β -cells. Insights from this and subsequent functional screens can be integrated with complementary lines of evidence from exome-wide association studies, chromatin conformation capture and cis-eQTL studies to prioritize genes for follow-up studies. Ultimately, this could accelerate the translation of genetic association signals into molecular mechanisms for β -cell dysfunction, insulin insufficiency, and type 2 diabetes.

Acknowledgments

This study was funded in Oxford by the Wellcome Trust (095101/Z/10/Z and 098381). SKT is a Radcliffe Department of Medicine scholar. MvdB is supported by a Novo Nordisk postdoctoral fellowship run in partnership with the University of Oxford. MIM is Wellcome Trust Senior Investigator. ALG is a Wellcome Trust Senior fellow in Basic Biomedical Science.

SKT, AC, DE, MIM, ALG conceived and designed the study. RS provided protocols. SKT, AC, CB and AB performed the experiments. SKT, MvdB, MIM and ALG analyzed and interpreted the data. SKT, MIM and ALG wrote the manuscript. SKT, AC, MvdB, CB, AB, RS, DE, MIM and ALG edited and approved the manuscript. ALG is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. The authors declare that there is no conflict of interest.

References

- [1] O. Ali, Genetics of type 2 diabetes. *World journal of diabetes* 4 (2013) 114-23.
- [2] R.B. Prasad, and L. Groop, Genetics of Type 2 Diabetes-Pitfalls and Possibilities. *Genes* 6 (2015) 87-123.
- [3] A.S. Dimas, et al., Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* 63 (2014) 2158-71.
- [4] S.K. Thomsen, and A.L. Gloyn, The pancreatic beta cell: recent insights from human genetics. *Trends Endocrinol Metab* 25 (2014) 425-34.
- [5] J. Fadista, et al., Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proc Natl Acad Sci U S A* 111 (2014) 13924-9.
- [6] M. van de Bunt, et al., 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 (2015) e1005694-e1005694.
- [7] S.M. Burns, et al., High-Throughput Luminescent Reporter of Insulin Secretion for Discovering Regulators of Pancreatic Beta-Cell Function. *Cell Metabolism* 21 (2015) 126-137.
- [8] G.M. Ku, et al., An siRNA Screen in Pancreatic Beta Cells Reveals a Role for Gpr27 in Insulin Production. *PLoS Genet* 8 (2012).
- [9] W. Wu, et al., Identification of glucose-dependant insulin secretion targets in pancreatic beta cells by combining defined-mechanism compound library screening and siRNA gene silencing. *Journal of Biomolecular Screening* 13 (2008) 128-134.
- [10] J.A. Lee, et al., Open Innovation for Phenotypic Drug Discovery: The PD2 Assay Panel. *Journal of Biomolecular Screening* 16 (2011) 588-602.
- [11] D. Walpita, et al., A Human Islet Cell Culture System for High-Throughput Screening. *Journal of Biomolecular Screening* 17 (2012) 509-518.
- [12] P. Ravassard, et al., A genetically engineered human pancreatic beta cell line exhibiting glucose-inducible insulin secretion. *J Clin Invest* 121 (2011) 3589-97.
- [13] G.C. Weir, and S. Bonner-Weir, Finally! A human pancreatic beta cell line. *J Clin Invest* 121 (2011) 3395-7.
- [14] L.E. Andersson, et al., Characterization of stimulus-secretion coupling in the human pancreatic EndoC-betaH1 beta cell line. *PLoS One* 10 (2015) e0120879.
- [15] K. Krishnan, et al., Calcium signaling in a genetically engineered human pancreatic beta-cell line. *Pancreas* 44 (2015) 773-7.
- [16] E. Gurgul-Convey, et al., Physiological characterization of the human EndoC-betaH1 beta-cell line. *Biochem Biophys Res Commun* 464 (2015) 13-9.
- [17] A. Pal, et al., Loss-of-Function Mutations in the Cell-Cycle Control Gene CDKN2A Impact on Glucose Homeostasis in Humans. *Diabetes* 65 (2016) 527-33.
- [18] A.C. Nica, et al., Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. *Genome Res* 23 (2013) 1554-62.
- [19] B.M. Shields, et al., Maturity-onset diabetes of the young (MODY): how many cases are we missing? *Diabetologia* 53 (2010) 2504-2508.
- [20] E.R. Pearson, et al., Macrosomia and hyperinsulinaemic hypoglycaemia in patients with heterozygous mutations in the HNF4A gene. *Plos Medicine* 4 (2007) 760-769.
- [21] B. Glaser, Type 2 diabetes: Hypoinsulinism, hyperinsulinism, or both? *Plos Medicine* 4 (2007) 619-620.
- [22] S.A. Rahman, et al., Molecular mechanisms of congenital hyperinsulinism. *Journal of Molecular Endocrinology* 54 (2015) R119-R129.
- [23] L. Pasquali, et al., Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet* 46 (2014) 136-43.
- [24] A.M. Simonis-Bik, et al., Gene Variants in the Novel Type 2 Diabetes Loci CDC123/CAMK1D, THADA, ADAMTS9, BCL11A, and MTNR1B Affect Different Aspects of Pancreatic beta-Cell Function. *Diabetes* 59 (2010) 293-301.
- [25] H. Parikh, et al., Prioritizing genes for follow-up from genome wide association studies using information on gene expression in tissues relevant for type 2 diabetes mellitus. *BMC Medical Genomics* 2 (2009) 72.

Gene	Locus	Low glucose	High glucose	IBMX	Tolbutamide	Cell count
<i>ABCC8</i>	<i>KCNJ11</i>	48.2*	24	26.7*	1.8	-1.4
<i>ADAMTS9</i>	<i>ADAMTS9</i>	6.3	-4.8	2.8	-8	12.2*
<i>ADIPOQ</i>	<i>ST64GAL1</i>	87*	23.2	23.1	8.8	-7.3
<i>ARL15</i>	<i>ARL15</i>	-5.5	-25.9*	-2.1	-15.5	-1.5
<i>BCAR1</i>	<i>BCAR1</i>	5.9	25.2	28.5*	9.5	-7
<i>BCL6</i>	<i>LPP</i>	-20.7*	-8.9	-1	-12	5.6
<i>BMP8B</i>	<i>MACF1</i>	7.8	16.5	9.5	26.9*	-1
<i>CCNT2</i>	<i>TMEM163</i>	-32.6*	1.8	4.7	5	-2.8
<i>CDKAL1</i>	<i>CDKAL1</i>	2.4	1.7	-10.6	-23.5*	7.5
<i>DGKQ</i>	<i>MAEA</i>	3.3	17.5	20.4	33.6*	-9.8
<i>DMRTA2</i>	<i>FAF1</i>	32.3*	24.5	13.1	22.7	-0.3
<i>ELAVL4</i>	<i>FAF1</i>	-3.6	9.1	21.2	22.8	-11.4*
<i>ETV5</i>	<i>IGF2BP2</i>	-12.4	-25.6*	-10.9	-12.6	-2.5
<i>FAH</i>	<i>ZFAND6</i>	-23.1*	-20.6*	-16.9	-27.2	-2.3
<i>FBXW7</i>	<i>TMEM154</i>	45.2*	9.6	8.7	11.7	9.9*
<i>GINS4</i>	<i>ANK1</i>	-16.1	-14.2	-9.5	-17	8.7*
<i>GLIS3</i>	<i>GLIS3</i>	-13	-10.6	6.4	-9.3	-10.3*
<i>HEYL</i>	<i>MACF1</i>	29.2	29.2*	0.4	20	-7.2
<i>HMGA2</i>	<i>HMGA2</i>	16.5	4.1	14.2	24.1*	-1.2
<i>HNF1A</i>	<i>HNF1A</i>	21.7	38.3*	23.1	25.5*	5.2
<i>HNF4A</i>	<i>HNF4A</i>	36.7*	66.9*	74.9	92.9*	-8.8
<i>IGF2</i>	<i>DUSP8</i>	-26.3*	-10.4	1.2	0.1	-1.3
<i>INS</i>	<i>DUSP8</i>	-53.5*	-44.8*	-48.7	-38.7*	0.9
<i>KCNK17</i>	<i>KCNK16</i>	-9.4	-17.3	-2.5	-1.2	9.8*
<i>KCTD15</i>	<i>PEPD</i>	21.7	9.4	12.5	0.8	-10.9*
<i>KIF11</i>	<i>HHEX/IDE</i>	45.4*	35*	26.9	55.4*	-40.1*

<i>LINGO1</i>	<i>HMG20A</i>	0	19.1	-12.4	-14.9	7.9*
<i>MFGE8</i>	<i>AP3S2</i>	30*	3.4	2.8	-1.7	-5.7
<i>MIER3</i>	<i>ANKRD55</i>	5.9	36.5*	5.8	18.2	1.9
<i>NDUFS4</i>	<i>ARL15</i>	1.6	-4.9	3.9	-1.7	10.8*
<i>PABPC1L</i>	<i>HNF4A</i>	-9.7	-12.3	-10.4	-28.8*	-1.2
<i>PHF23</i>	<i>SLC16A11</i>	-25.6*	-1.7	-5.9	3.5	-0.9
<i>PLA2R1</i>	<i>RBMS1</i>	8.6	1.6	10.8	1.6	9.1*
<i>PRDX3</i>	<i>GRK5</i>	24	31.7*	23.4	9.6	16.6*
<i>PTHLH</i>	<i>KLHDC5</i>	-2.8	-6.5	-0.5	-25*	-5.9
<i>RND3</i>	<i>RND3</i>	-8.7	-6.1	0	-3	-14.9*
<i>SLC2A4</i>	<i>SLC16A11</i>	14.5	0	27.2*	18.2	-1.6
<i>SOCS7</i>	<i>HNF1B</i>	3.9	-18.5*	11.2	-14.7	-1.6
<i>SPPL3</i>	<i>HNF1A</i>	-11.9	-21.9*	-6.1	-10	-10.8*
<i>STK38L</i>	<i>KLHDC5</i>	15.2	40.9*	4.7	25.2*	-3
<i>THADA</i>	<i>THADA</i>	-1.9	6.5	27.5	24.8*	-10.3
<i>TLE1</i>	<i>TLE1</i>	4.3	-5	-23*	16.2	4.6
<i>TM6SF2</i>	<i>CILP2</i>	-22.6*	-8.3	-0.8	-12	-8.7
<i>UPF2</i>	<i>CDC123</i>	7.5	-12.5	4.7	-24.9*	-3.2
<i>ZMIZ1</i>	<i>ZMIZ1</i>	-29.5*	-21.4*	-19.8	-16.8	-15.2*

Table 1 Effects of significant hits identified in a primary screen for β -cell dysfunction. The table lists effect sizes (% deviation from NT control) for each gene with a least one significant effect across the five phenotypes measured. All insulin secretion measurements were normalized on a per-well basis to cell-counts, and the mean percentage deviations from NT control were then calculated for each condition. For cell counts, values were median-normalized for interplate differences, and the mean percentage deviations from NT control were calculated across conditions. * $q < 0.05$ by Student's t-test (FDR-adjusted).

Figure legends

Figure 1 *Comparing mean absolute effect sizes for MODY and non-MODY genes.* Boxplots of mean absolute effect sizes for MODY genes (*white*) and non-MODY genes (excluding controls, *black*) across the five phenotypes measured. Effect sizes were calculated as described for table 1, and the absolute values were then averaged for the two categories of genes. Among 14 identified MODY genes, eight fulfilled criteria for inclusion in the screen: *HNF4A*, *GCK*, *HNF1A*, *HNF1B*, *PAX4*, *INS*, *ABCC8* and *KCNJ11*. Boxplots show median and interquartile ranges for groups of $n = 8$ and 292 data-points. *** q -value < 0.001 by Student's t-test (FDR-adjusted). Tol = Tolbutamide.

Figure 2 *Comparison of insulin secretion data for high and low glucose.* Normalized insulin secretion responses under high glucose versus low glucose, with selected hits annotated. The blue circle indicates the 95 % confidence contour for NT control, and the orange circle indicates the 95 % confidence contour for insulin (*INS*) positive controls. All measurements were normalized on a per-well basis to cell-counts, and averages for each condition were then subsequently normalized to the mean of NT control. Data points are mean of $n = 3$ and shown as percentage of NT control.

Figure 3 *Insulin secretion data for selected genes in a follow-up validation experiment.* Insulin secretion for (A) *ARL15*, (B) *ZMIZ1*, (C) *THADA*, and (D) *HNF4A* (*white*) versus non-targeting (NT; *black*) negative control under the indicated conditions. Measurements were processed as described for Fig 2, and shown as percentage of NT control. Data points are the mean of $n = 6$ for NT and $n = 3$ for other genes, and error bars are SEM. + q -value < 0.1 , * q -value < 0.05 , ** q -value < 0.01 , *** q -value < 0.001 by Student's t-test (FDR-adjusted). Tol = Tolbutamide.

Figure 1

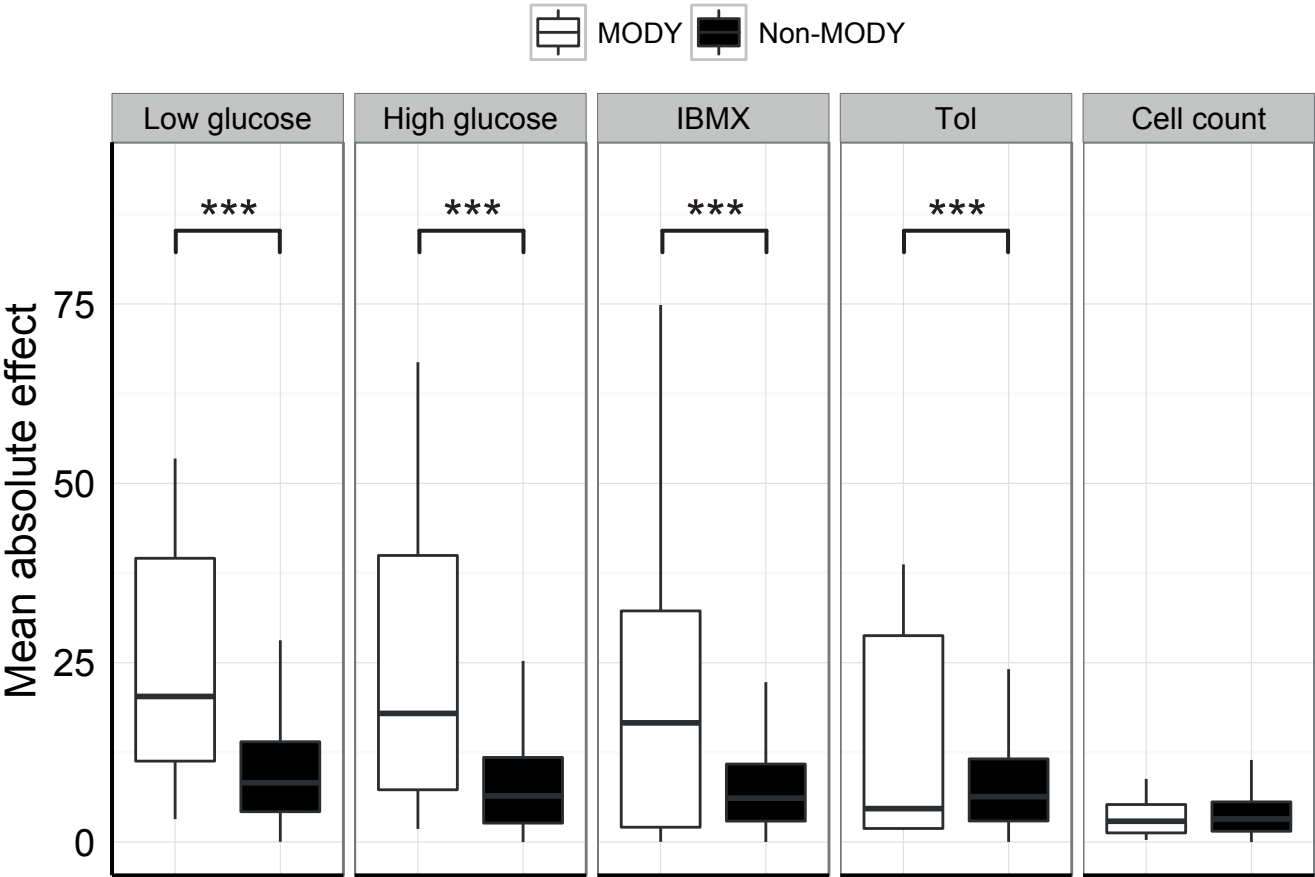


Figure 2

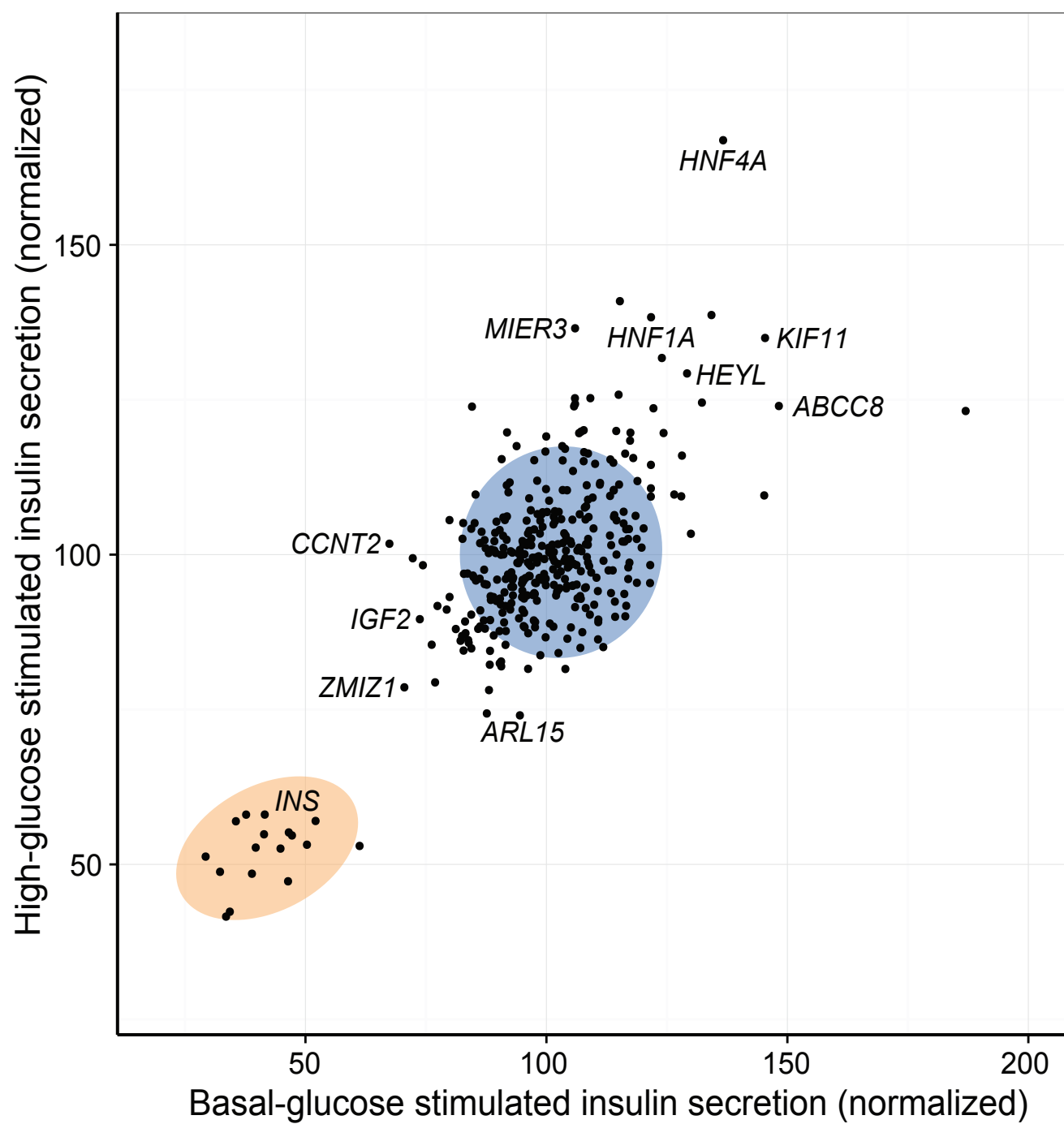
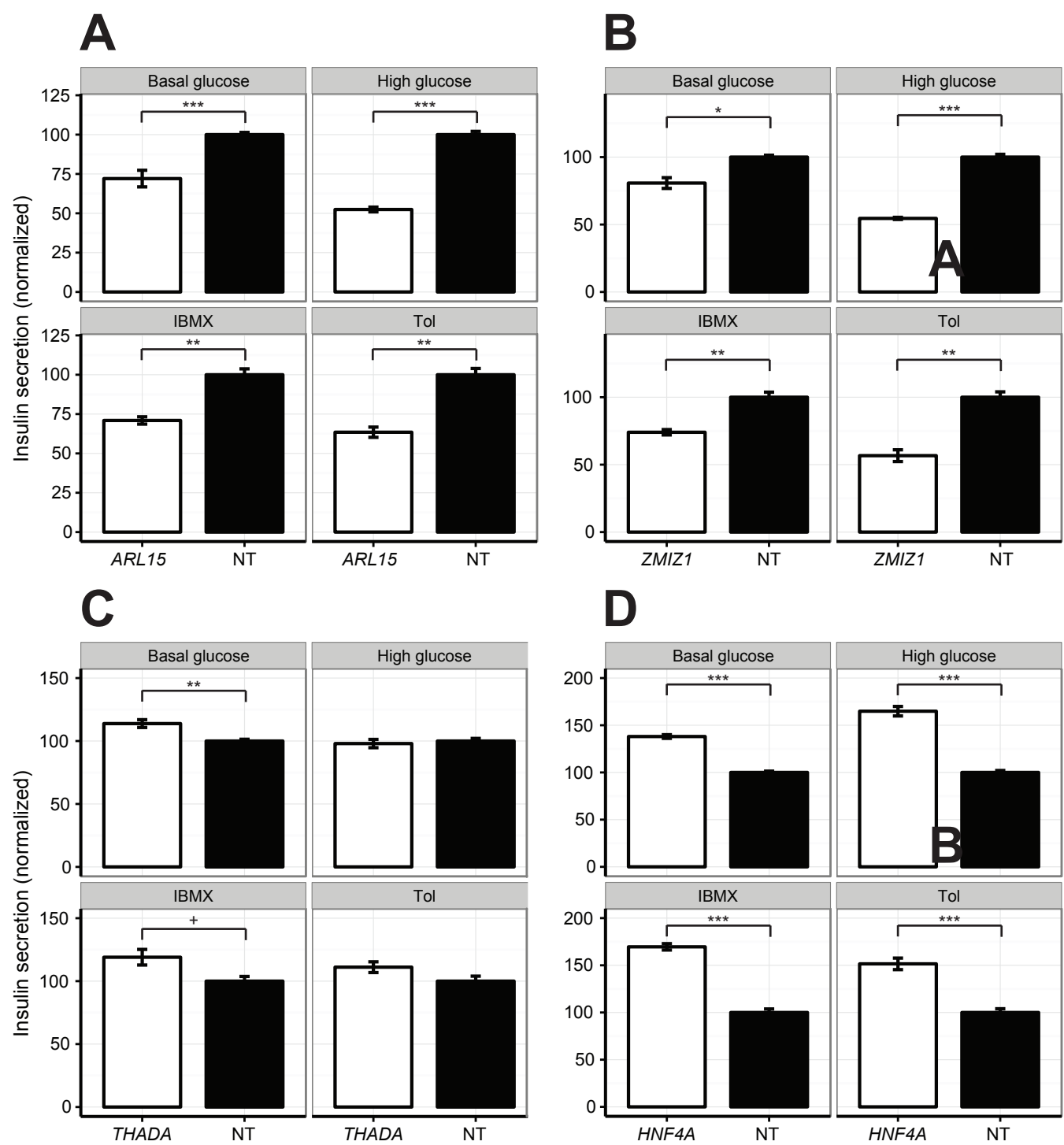


Figure 3



Online Supplemental Data

Ensembl gene ID	Gene Name	EndoC-βH1 [RPKM]	Primary β cells [RPKM]
ENSG00000254647.2	<i>INS</i>	8012.452	166347.111
ENSG00000087086.9	<i>FTL</i>	3090.7454	2066.464
ENSG00000100604.8	<i>CHGA</i>	2853.107	1113.162
ENSG00000099194.5	<i>SCD</i>	1411.631	238.714
ENSG00000118271.5	<i>TTR</i>	1312.8928	1488.996
ENSG00000184009.5	<i>ACTG1</i>	1108.0277	839.681
ENSG00000124172.5	<i>ATP5E</i>	863.42334	254.779
ENSG00000156508.13	<i>EEF1A1</i>	831.17316	637.281
ENSG00000112972.10	<i>HMGCS1</i>	719.7504	22.104
ENSG00000167552.9	<i>TUBA1A</i>	689.1415	511.699

Supplementary table 1 *Top ten most highly expressed protein-coding genes in the EndoC-βH1 cell line.* Expression levels provided for non-mitochondrial genes in EndoC-βH1 and the corresponding expression levels in sorted primary human β-cells (1).

Gene Name	T2D Locus	EndoC-βH1 [RPKM]	Primary β cells [RPKM]
<i>MACF1</i>	<i>MACF1</i>	10.727805	15.879
<i>BMP8A</i>	<i>MACF1</i>	1.4541069	0.64
<i>HEYL</i>	<i>MACF1</i>	1.5134418	0.791
<i>BMP8B</i>	<i>MACF1</i>	12.184336	5.887
<i>MFSD2A</i>	<i>MACF1</i>	1.2909733	0.846
<i>RHBDL2</i>	<i>MACF1</i>	0.6474848	0.224
<i>CDKN2C</i>	<i>FAF1</i>	80.097336	5.212
<i>FAF1</i>	<i>FAF1</i>	15.062617	6.613
<i>DMRTA2</i>	<i>FAF1</i>	0.046803616	0.014
<i>ELAVL4</i>	<i>FAF1</i>	10.931084	19.561
<i>TTC39A</i>	<i>FAF1</i>	14.173553	10.022
<i>PROX1</i>	<i>PROX1</i>	10.218364	12.548
<i>PTPN14</i>	<i>PROX1</i>	0.01770482	0.14
<i>SMYD2</i>	<i>PROX1</i>	7.387975	2.934
<i>PPM1G</i>	<i>GCKR</i>	66.50968	27.658
<i>KHK</i>	<i>GCKR</i>	25.9325	1.467
<i>CENPA</i>	<i>GCKR</i>	9.154673	0.098
<i>PLB1</i>	<i>GCKR</i>	0.04874327	0.456
<i>THADA</i>	<i>THADA</i>	3.465101	2.575
<i>LRPPRC</i>	<i>THADA</i>	49.46855	16.941
<i>DYNC2LI1</i>	<i>THADA</i>	5.1710835	4.216
<i>PLEKHH2</i>	<i>THADA</i>	1.5512799	2.889
<i>BCL11A</i>	<i>BCL11A</i>	0.42303053	0.44
<i>PAPOLG</i>	<i>BCL11A</i>	8.144624	3.191
<i>MGAT5</i>	<i>TMEM163</i>	11.647037	6.124
<i>CCNT2</i>	<i>TMEM163</i>	14.048298	10.553
<i>ACMSD</i>	<i>TMEM163</i>	0.021140289	0.875
<i>MAP3K19</i>	<i>TMEM163</i>	0.002687708	0.056
<i>R3HDM1</i>	<i>TMEM163</i>	20.273561	7.659
<i>RIF1</i>	<i>RBM43/RND3</i>	13.902107	5.398
<i>NMI</i>	<i>RBM43/RND3</i>	4.603304	2.955
<i>RBM43</i>	<i>RBM43/RND3</i>	0.9375855	3.668
<i>NEB</i>	<i>RBM43/RND3</i>	0.28890586	0.038
<i>RND3</i>	<i>RBM43/RND3</i>	5.034543	1.76
<i>ITGB6</i>	<i>RBMS1</i>	0.052311223	0.845
<i>PLA2R1</i>	<i>RBMS1</i>	0.003079499	0.115
<i>LY75</i>	<i>RBMS1</i>	11.025683	3.095
<i>RBMS1</i>	<i>RBMS1</i>	12.231243	2.013
<i>CD302</i>	<i>RBMS1</i>	7.510972	2.322
<i>GRB14</i>	<i>GRB14</i>	8.159008	2.218
<i>COBLL1</i>	<i>GRB14</i>	5.161108	9.288
<i>SCN2A</i>	<i>GRB14</i>	0.29049847	2.111
<i>SCN3A</i>	<i>GRB14</i>	0.4411591	4.009
<i>IRS1</i>	<i>IRS1</i>	4.0152564	2.761
<i>NKIRAS1</i>	<i>UBE2E2</i>	7.2671003	12.391
<i>UBE2E1</i>	<i>UBE2E2</i>	11.214179	14.05
<i>PSMD6</i>	<i>PSMD6</i>	39.243126	32.951
<i>PRICKLE2</i>	<i>PSMD6</i>	0.29158774	4.445
<i>SYNPR</i>	<i>PSMD6</i>	0.34662268	0.02
<i>THOC7</i>	<i>PSMD6</i>	26.802912	11.24
<i>ATXN7</i>	<i>PSMD6</i>	6.243315	8.72
<i>ADAMTS9</i>	<i>ADAMTS9</i>	0.006904925	0.497
<i>ADCY5</i>	<i>ADCY5</i>	0.10015087	0.78
<i>MYLK</i>	<i>ADCY5</i>	0.67700344	0.167
<i>PDIA5</i>	<i>ADCY5</i>	3.5262358	2.34
<i>PTPLB</i>	<i>ADCY5</i>	41.775444	10.987
<i>SEC22A</i>	<i>ADCY5</i>	6.3320513	5.297
<i>DIRC2</i>	<i>ADCY5</i>	6.195492	6.982
<i>IGF2BP2</i>	<i>IGF2BP2</i>	0.044629	1.983
<i>SENP2</i>	<i>IGF2BP2</i>	14.213694	7.035
<i>DGKG</i>	<i>IGF2BP2</i>	1.6199253	2.127
<i>MAP3K13</i>	<i>IGF2BP2</i>	5.584649	8.417
<i>LIPH</i>	<i>IGF2BP2</i>	0.029294837	12.143
<i>ETV5</i>	<i>IGF2BP2</i>	6.104882	6.329
<i>TRA2B</i>	<i>IGF2BP2</i>	48.647205	27.016
<i>ST6GAL1</i>	<i>ST6GAL1</i>	25.443888	6.183
<i>ADIPOQ</i>	<i>ST6GAL1</i>	0.006442239	0.001
<i>AHSG</i>	<i>ST6GAL1</i>	0.07925758	0.012
<i>TBCCD1</i>	<i>ST6GAL1</i>	6.7398887	5.311
<i>CRYGS</i>	<i>ST6GAL1</i>	1.1851472	1.43
<i>SST</i>	<i>LPP</i>	43.108746	3555.411
<i>BCL6</i>	<i>LPP</i>	2.189676	10.494
<i>LPP</i>	<i>LPP</i>	3.889396	5.031
<i>CTBP1</i>	<i>MAEA</i>	19.709608	18.553

RNF212	MAEA	0.007606029	0.677
DGKQ	MAEA	6.5478916	4.204
FAM53A	MAEA	0.5075767	0.237
TMEM175	MAEA	14.789575	15.925
WFS1	WFS1	48.648613	64.697
S100P	WFS1	0.094338536	0.239
KIAA0232	WFS1	13.799467	20.42
ARFIP1	TMEM154	15.90032	12.549
TIGD4	TMEM154	1.2456352	1.317
FBXW7	TMEM154	5.181745	13.471
FHDC1	TMEM154	2.2825546	3.897
TMEM154	TMEM154	0.09626208	0.536
NDUFS4	ARL15	19.788746	11.809
ARL15	ARL15	15.944361	23.853
MAP3K1	ANKRD55	3.182025	3.707
MIER3	ANKRD55	9.665144	10.573
PDE8B	ZBED3	1.3867581	9.642
F2R	ZBED3	1.0097425	1.087
ZBED3	ZBED3	8.214433	10.218
S100Z	ZBED3	5.69416	0.415
RREB1	SSR1/RREB1	8.097911	9.925
BMP6	SSR1/RREB1	0.097318925	0.079
SNRNP48	SSR1/RREB1	17.41273	6.136
CAGE1	SSR1/RREB1	0.026015453	0.022
E2F3	CDKAL1	7.951786	3.897
MBOAT1	CDKAL1	0.6750357	0.225
CDKAL1	CDKAL1	9.097854	4.014
HLA-C	POU5F1/TCF19	20.314272	128.262
HLA-B	POU5F1/TCF19	1.0356976	89.07
PRR3	POU5F1/TCF19	10.210759	8.001
GNL1	POU5F1/TCF19	18.412489	10.526
BTBD9	ZFAND3	3.0280964	5.045
GLO1	ZFAND3	60.216225	38.375
ZFAND3	ZFAND3	32.21161	30.611
SAYS1	ZFAND3	12.916478	13.804
KIF6	KCNK16	1.6660756	0.227
GLP1R	KCNK16	1.8521343	54.753
KCNK5	KCNK16	0.2531769	0.361
DAAM2	KCNK16	0.24726911	0.065
KCNK17	KCNK16	104.62451	52.748
KCNK16	KCNK16	102.129074	169.1
DGKB	DGKB	0.056520004	2.545
CREB5	JAZF1	0.03463798	2.023
TAX1BP1	JAZF1	15.722485	37.226
JAZF1	JAZF1	1.3256978	5.059
HIBADH	JAZF1	7.88914	21.946
GCK	GCK	42.563545	21.444
YKT6	GCK	36.720226	26.003
UBE2D4	GCK	8.489575	8.93
TMED4	GCK	41.99634	44.895
PAX4	GCC1	1.3463995	1.025
SND1	GCC1	54.31868	38.648
RBM28	GCC1	24.012737	7.622
METTL2B	GCC1	16.067684	4.139
COPG2	KLF14	28.098372	6.619
CPA1	KLF14	21.122427	215.01
MKLN1	KLF14	13.815692	10.359
KLF14	KLF14	0.30431783	0.105
NKX6-3	ANK1	46.473755	4.079
SFRP1	ANK1	19.806135	0.178
GINS4	ANK1	4.1673384	0.309
AP3M2	ANK1	14.139829	5.208
TP53INP1	TP53INP1	34.056793	21.408
CCNE2	TP53INP1	11.706554	0.69
NDUFAF6	TP53INP1	3.0254307	1.929
DPY19L4	TP53INP1	10.406852	5.737
SLC30A8	SLC30A8	50.082935	517.613
RAD21	SLC30A8	125.50391	43.451
MED30	SLC30A8	5.853778	7.912
EIF3H	SLC30A8	30.243889	23.136
GLIS3	GLIS3	9.963094	16.547
AK3	GLIS3	41.39607	41.579
SLC1A1	GLIS3	0.668251	2.219
CDC37L1	GLIS3	6.7730727	8.744
CDKN2B	CDKN2A/B	14.910821	2.787

DMRTA1	CDKN2A/B	2.3605123	4.108
CDKN2A	CDKN2A/B	16.370024	3.16
MTAP	CDKN2A/B	7.5099783	2.603
TLE4	TLE4	6.7753386	1.565
TLE1	TLE1	18.546194	10.735
GPSM1	GPSM1	6.6489964	9.399
FAM69B	GPSM1	12.96178	20.863
C8G	GPSM1	0.25058675	0.447
CAMK1D	CDC123/CAMK1D	3.047827	1.783
CDC123	CDC123/CAMK1D	21.241869	13.717
UPF2	CDC123/CAMK1D	5.952305	7.418
SEC61A2	CDC123/CAMK1D	4.153639	3.573
NUDT5	CDC123/CAMK1D	11.880336	6.168
DHTKD1	CDC123/CAMK1D	7.0249395	4.896
HK1	VPS26A	0.039167475	0.905
NEUROG3	VPS26A	1.7002876	0.003
VPS26A	VPS26A	17.768085	10.95
STOX1	VPS26A	4.8269176	0.077
C10orf35	VPS26A	12.514925	7.626
PPIF	ZMIZ1	33.738903	8.455
SFTP1A1	ZMIZ1	0.25245523	12.246
ZMIZ1	ZMIZ1	25.735308	34.923
ZCCHC24	ZMIZ1	0.43032268	0.357
HHEX	HHEX/IDE	0.06326138	1.421
IDE	HHEX/IDE	6.702786	5.403
KIF11	HHEX/IDE	19.037043	0.48
CPEB3	HHEX/IDE	2.9978027	5.982
TCF7L2	TCF7L2	2.292974	5.015
VTI1A	TCF7L2	2.0465448	4.366
GRK5	GRK5	4.9136653	0.907
NANOS1	GRK5	3.1578026	7.825
PRDX3	GRK5	49.93399	32.619
SFXN4	GRK5	6.9072022	5.75
INPP5F	GRK5	2.946096	41.344
PRLHR	GRK5	2.050672	1.39
KCNJ11	KCNJ11	28.18143	40.775
ABCC8	KCNJ11	67.27029	244.683
RPS13	KCNJ11	468.69717	83.633
PLEKHA7	KCNJ11	6.322894	3.168
NCR3LG1	KCNJ11	24.444202	13.894
ARAP1	ARAP1 (CENTD2)	13.448667	14.159
INPPL1	ARAP1 (CENTD2)	44.3585	9.872
STARD10	ARAP1 (CENTD2)	111.17294	36.176
CLPB	ARAP1 (CENTD2)	6.619748	4.168
ATG16L2	ARAP1 (CENTD2)	7.742456	2.648
MTNR1B	MTNR1B	0.015433707	0.642
FAT3	MTNR1B	0.14323217	0.037
SLC36A4	MTNR1B	19.859741	15.109
PTHLH	KLHDC5	0.07151515	0.192
PPFIBP1	KLHDC5	4.0497327	2.349
MRPS35	KLHDC5	49.13422	21.838
ARNTL2	KLHDC5	5.9386883	0.446
STK38L	KLHDC5	9.114885	6.608
REP15	KLHDC5	0.3843115	0.272
KLHL42	KLHDC5	26.8217	30.478
HMGA2	HMGA2	0.013419024	0.009
IRAK3	HMGA2	0.22330433	0.329
HELB	HMGA2	0.51851743	1.773
TMBIM4	HMGA2	3.3154817	1.907
LGR5	TSPAN8/LGR5	0.09165835	0.173
PTPRR	TSPAN8/LGR5	0.03522559	0.137
TPH2	TSPAN8/LGR5	0.2344639	0.175
RAB21	TSPAN8/LGR5	25.835938	22.635
HNFI1A	HNFI1A (TCF1)	11.920588	4.427
CABP1	HNFI1A (TCF1)	0.47474712	0.92
CAMKK2	HNFI1A (TCF1)	5.4568853	2.067
KDM2B	HNFI1A (TCF1)	4.485531	2.082
SPPL3	HNFI1A (TCF1)	21.66794	9.827
MLEC	HNFI1A (TCF1)	52.709465	36.949
CDK2AP1	MPHOSPH9	12.791418	28.311
OGFOD2	MPHOSPH9	3.5529344	4.656
ABCB9	MPHOSPH9	2.5140104	1.566
EIF2B1	MPHOSPH9	23.559034	9.742
TCTN2	MPHOSPH9	9.149691	5.229
ZCCHC8	MPHOSPH9	6.5567923	6.602

<i>RASGRP1</i>	<i>RASGRP1</i>	0.6741565	7.274
<i>SPRED1</i>	<i>RASGRP1</i>	2.1910295	1.428
<i>FAM98B</i>	<i>RASGRP1</i>	12.422814	5.6
<i>VPS13C</i>	<i>C2CD4A</i>	23.849556	23.607
<i>HMG20A</i>	<i>HMG20A</i>	14.298759	12.474
<i>PSTPIP1</i>	<i>HMG20A</i>	0.33292773	0.032
<i>LINGO1</i>	<i>HMG20A</i>	7.0082254	0.117
<i>TBC1D2B</i>	<i>HMG20A</i>	9.462256	7.787
<i>FAH</i>	<i>ZFAND6</i>	0.85486287	1.303
<i>ARNT2</i>	<i>ZFAND6</i>	24.756863	11.261
<i>MTHFS</i>	<i>ZFAND6</i>	4.262472	3.327
<i>ZFAND6</i>	<i>ZFAND6</i>	37.31066	19.454
<i>MESP1</i>	<i>AP3S2</i>	3.0201075	0.914
<i>MESP2</i>	<i>AP3S2</i>	0.10175843	0.035
<i>ZNF710</i>	<i>AP3S2</i>	6.5410094	4.404
<i>MFGE8</i>	<i>AP3S2</i>	19.445135	0.704
<i>VPS33B</i>	<i>PRC1</i>	7.206141	4.765
<i>BLM</i>	<i>PRC1</i>	4.386412	0.217
<i>HDDC3</i>	<i>PRC1</i>	7.861545	5.238
<i>RCCD1</i>	<i>PRC1</i>	14.031963	1.311
<i>FTO</i>	<i>FTO</i>	5.057489	7.094
<i>RBL2</i>	<i>FTO</i>	21.088936	17.104
<i>RPGRIP1L</i>	<i>FTO</i>	2.059937	0.574
<i>AKTIP</i>	<i>FTO</i>	6.018698	6.684
<i>CHD9</i>	<i>FTO</i>	8.3386135	13.784
<i>CFDP1</i>	<i>BCAR1</i>	23.719404	10.403
<i>BCAR1</i>	<i>BCAR1</i>	14.373125	20.272
<i>ZFP1</i>	<i>BCAR1</i>	7.834037	6.043
<i>ADAT1</i>	<i>BCAR1</i>	13.082058	8.633
<i>SRR</i>	<i>SRR</i>	6.8297243	2.776
<i>PAFAH1B1</i>	<i>SRR</i>	59.24745	43.372
<i>MNT</i>	<i>SRR</i>	5.6960053	9.685
<i>RTN4RL1</i>	<i>SRR</i>	2.3101637	10.227
<i>METTL16</i>	<i>SRR</i>	9.847925	12.353
<i>SGSM2</i>	<i>SRR</i>	23.97471	34.762
<i>SLC16A11</i>	<i>SLC16A11</i>	5.575549	1.285
<i>SLC2A4</i>	<i>SLC16A11</i>	0.0136352	0.013
<i>PHF23</i>	<i>SLC16A11</i>	13.260954	10.724
<i>TNFSF13</i>	<i>SLC16A11</i>	2.1491694	6.405
<i>HNF1B</i>	<i>HNF1B (TCF2)</i>	0.030431787	2.781
<i>ACACA</i>	<i>HNF1B (TCF2)</i>	44.620758	10.826
<i>GPR179</i>	<i>HNF1B (TCF2)</i>	1.1049079	0.087
<i>DUSP14</i>	<i>HNF1B (TCF2)</i>	7.0064955	5.688
<i>SOC57</i>	<i>HNF1B (TCF2)</i>	6.7027836	5.675
<i>DDX52</i>	<i>HNF1B (TCF2)</i>	10.564817	5.468
<i>TBC1D3F</i>	<i>HNF1B (TCF2)</i>	0.32611838	0.114
<i>LAMA1</i>	<i>LAMA1</i>	0.92843896	0.15
<i>PTPRM</i>	<i>LAMA1</i>	1.0576878	7.344
<i>ARHGAP28</i>	<i>LAMA1</i>	0.99569255	0.08
<i>MC4R</i>	<i>MC4R</i>	0.09446975	0.001
<i>PMAIP1</i>	<i>MC4R</i>	35.006386	3.308
<i>CCBE1</i>	<i>MC4R</i>	0.057355918	0.151
<i>BCL2</i>	<i>BCL2A</i>	0.53563124	0.175
<i>KDSR</i>	<i>BCL2A</i>	9.550464	10.958
<i>PHLPP1</i>	<i>BCL2A</i>	6.758193	2.535
<i>VPS4B</i>	<i>BCL2A</i>	24.030014	14.894
<i>TM6SF2</i>	<i>CILP2</i>	0.92930496	1.832
<i>YJEFN3</i>	<i>CILP2</i>	1.3774142	1.29
<i>ZNF486</i>	<i>CILP2</i>	0.0070848	1.712
<i>TDRD12</i>	<i>PEPD</i>	0.051692348	0.076
<i>CEBPA</i>	<i>PEPD</i>	14.847472	0.377
<i>CEBPG</i>	<i>PEPD</i>	29.904247	20.203
<i>KCTD15</i>	<i>PEPD</i>	2.694448	0.526
<i>CEP89</i>	<i>PEPD</i>	4.883483	3.482
<i>HNF4A</i>	<i>HNF4A</i>	21.057198	6.233
<i>PABPC1L</i>	<i>HNF4A</i>	19.464327	3.764
<i>OSER1</i>	<i>HNF4A</i>	23.073265	20.905
<i>INS</i>	<i>DUSP8</i>	8012.452	166347.111
<i>IGF2</i>	<i>DUSP8</i>	95.30156	8.054
<i>MRPL23</i>	<i>DUSP8</i>	15.220552	7.366
<i>CTSD</i>	<i>DUSP8</i>	150.01717	134.269
<i>TH</i>	<i>DUSP8</i>	1.5446588	0.735

Supplementary table 2 *List of genes selected for inclusion in the primary screen.* Expression levels in EndoC- β H1 and sorted primary human β -cells are shown for all genes targeted for silencing in the primary screen, ordered by locus association (1). For gene selection, the following criteria were applied: we first considered (1) all protein-coding genes within 1 Mb of a type 2 diabetes association signal that (2) had non-zero expression (RPKM > 0) in both EndoC- β H1 and primary human β -cells. Previous studies have shown cis-eQTLs to form a relatively tight, symmetrical distribution around the target-gene transcription start site, and a 1 Mb cut-off is thus likely to capture most effector transcripts subject to cis regulation (2-5). (3a) Loci with fewer than five genes satisfying criteria 1 and 2 were then represented with all qualified genes in the library for the primary screen. (3b) For loci with five or more of such genes, we implemented a systematic selection strategy to reduce our search space where possible; genes with prior plausibility of causality (based on presence of coding association signals for type 2 diabetes, an islet cis-eQTL overlapping an association signal, and/or an established role in β -cell function based on automated text mining of PubMed abstracts) were all included. We used these criteria to effectively increase the information density (the proportion of expected effector transcripts) for those loci where a specific causal mechanism was favored by existing evidence. For all loci, at least two genes with no prior evidence of causality were also included per locus to extend our ability to pick up novel regulators of β -cell function.

Supplementary figure legends

Supplementary figure 1 *Overview of assay and screening strategy.* **(A)** At day 0, cells in 96-well format were transfected with pools of four siRNAs per gene. 72 h after knockdown, cells were starved overnight before measurements of the indicated phenotypes on day 4. For a subset of genes, mRNA was extracted to assess knockdown efficiency. For all genes, cell counts and insulin secretion were measured under four different conditions (see methods). **(B)** Based on the assay outlined above, we developed a screen of positional candidate genes located within 1Mb of a genome-wide significant signal for type 2 diabetes risk (see supplementary table 2 for detailed criteria). For the hypothetical locus shown, the three protein-coding genes *A*, *B* and *C* would be silenced in parallel. Each gene would then be assayed for multiple disease-relevant phenotypes, and significant effects identified according to a stringent FDR-based threshold (see methods for details). Measurement of extracellular insulin secretion is not amenable to pooled screening using emerging genome editing strategies, but our approach based on arrayed gene silencing provides a direct genotype-phenotype link that enables such screening.

Supplementary figure 2 *Optimization of nano-liter insulin immunoassays.* Using a range of known insulin concentrations, the concentration following dilution by acoustic liquid handling was measured by AlphaLISA insulin immunoassays (*see methods*) **(A)** Measured versus predicted insulin concentration (in pg/mL; both log-transformed). A linear regression (*blue*) and 95% confidence intervals for the linear fit (*grey*) are shown, highlighting an excellent correlation for predicted concentrations between 100 and 10,000 pg/mL. **(B)** The coefficient of variation (CV) for a range of measured concentrations of insulin. A logarithmic regression has been fitted to the data, and the dashed line indicates a CV of 5 %. At low input concentrations, the coefficient of variation rises exponentially, but the optimal dilution range

(1,000-10,000 pg/mL) for insulin samples could be determined for a CV < 3 %. Data points are the mean of n = 4, and error bars are SEM.

Supplementary figure 3 *Validation of screening assay using INS control.* Insulin secretion under the indicated conditions following knockdown of insulin (INS) positive control (white bars) and non-targeting (NT) negative control (black bars). All measurements were normalized on a per-well basis to cell-counts, and then averaged across conditions. As a measure of the robustness of the assay, the Z-factor (Z' , a standard indicator of statistical effect size in high-throughput screens; see methods) was derived for all conditions: Z' (low glucose) = 0.75, Z' (low glucose) = 0.68, Z' (IBMX) = 0.51, Z' (Tol) = 0.44. A $Z' > 0.5$, corresponding to 12 standard deviations between positive and negative controls, is generally considered an excellent assay. Data points are the mean of n = 3, and error bars are SEM. Tol = Tolbutamide.

Supplementary figure 4 *RNA-seq of sorted primary β -cell versus EndoC- β H1.* Expression levels in a published data-set for sorted primary human β -cells versus expression levels for the EndoC- β H1 cell line as measured by RNA-sequencing (1). A Poly-A selected library was prepared using the NEBNext PolyA mRNA Magnetic Isolation Module and NEBNext Ultra Directional RNA Library Prep Kit for Illumina (both New England Biolabs, E7490L and E7420L, respectively) with 12 cycles of PCR and custom 8bp indexes (6). The library was sequenced as a multiplex with other samples over seven lanes of Illumina HiSeq2000 as 100-nucleotide paired-end reads to a total depth of ~50 million read pairs, using the TruSeq PE Cluster Generation Kit v3 and TruSeq SBS Kit v3 (both Illumina, PE-401-3001 and FC-401-3001 respectively). TopHat2 on default settings was used to align sequencing reads to the human genome reference GRCh37.p13, and using GENCODE release 18 as the transcriptome reference (7, 8). Gene level read counts were quantified as reads per kilobase of transcript per million mapped reads (RPKM) using RNA-SeQC with the “strictMode” flag

set (9). All values are log-transformed RPKM, and a linear regression (*blue*) has been fitted to the data.

Supplementary figure 5 *Knockdown efficiency observed in the primary screen.* Boxplot of residual expression levels for a subset of 16 genes included in the primary screen following 96 h knockdown. Expression levels were measured by quantitative PCR (qPCR) using commercially available TaqMan probes, and normalized first to the two housekeeping genes (*TBP* and *PPIA*) and then to NT control. The genes included for assessment of knockdown efficiency were: *INS*, *KIF11*, *CAMK1D*, *AYKT6*, *IDE*, *KCNK17*, *ABCC8*, *GCK*, *ZMIZ1*, *KCNK16*, *TCF7L2*, *CDKAL1*, *DGKB*, *HHEX*, *KCNJ11*, *PAM*.

Supplementary figure 6 *Quantile-Quantile (Q-Q) plots for normalized insulin secretion data.* Sample quantiles for insulin secretion under (A) low glucose, (B) high glucose, (C) IBMX, and (D) Tolbutamide (Tol) versus theoretical quantiles for normally distributed variables. A linear fit (*black*) is shown for each of the four plots to indicate regions of approximate normality. The tails of non-normal deviations indicate enrichments for non-random significant effects, which were identified by an FDR-based threshold (see methods) (table 1).

Supplementary figure 7 *Clustering of insulin secretion data.* Dendrograms and heatmap of Z-scores derived from normalized insulin secretion values with length indicating the Euclidian distance between gene effects (vertical) or conditions (horizontal). Conditions (or genes) clustered together tend to similarly modulate the effect of genes (or behave similarly across conditions). Calculated distances between conditions (low glucose, “A”; high glucose, “B”; IBMX, “C”; Tolbutamide, “D”): A-B = 16.2, A-C = 18.2, A-D = 19.0, B-C = 17.8, B-D = 15.9, C-D = 16.6.

Supplementary figure 8 *Knockdown efficiency observed in the follow-up validation.*

Boxplot of residual expression levels for 8 genes taken forward for validation following 96 h knockdown. Expression levels were measured by quantitative PCR (qPCR) using commercially available TaqMan probes, and normalized first to the two housekeeping genes (*TBP* and *PPIA*) and then to NT control.

Supplementary figure 9 *Technical validation of primary screen results with non-identical siRNAs.*

Effect sizes for eight selected hits in a follow-up experiment versus corresponding effect sizes for the same subset of genes measured in the primary screen. The genes taken forward for validation (*THADA*, *ARL15*, *MIER3*, *ZMIZ1*, *KIF11*, *ABCC8*, *HNF4A* and *TH*) were selected to represent targets with varying degrees of prior evidence supporting causality and effect sizes of different magnitudes across conditions. Measurements were processed as described for Fig. 2. A linear regression (*blue*) and 95% confidence intervals for the linear fit (*grey*) are shown. The follow-up experiment was performed using pools consisting entirely of different sequences of siRNA. The strong correlation observed ($r^2 = 0.72$) suggests that off-target effects had minimal or no effect on the recorded effects for this subset genes.

Supplementary figure 10 *Regional plots with genomic and epigenetic annotations for selected loci.*

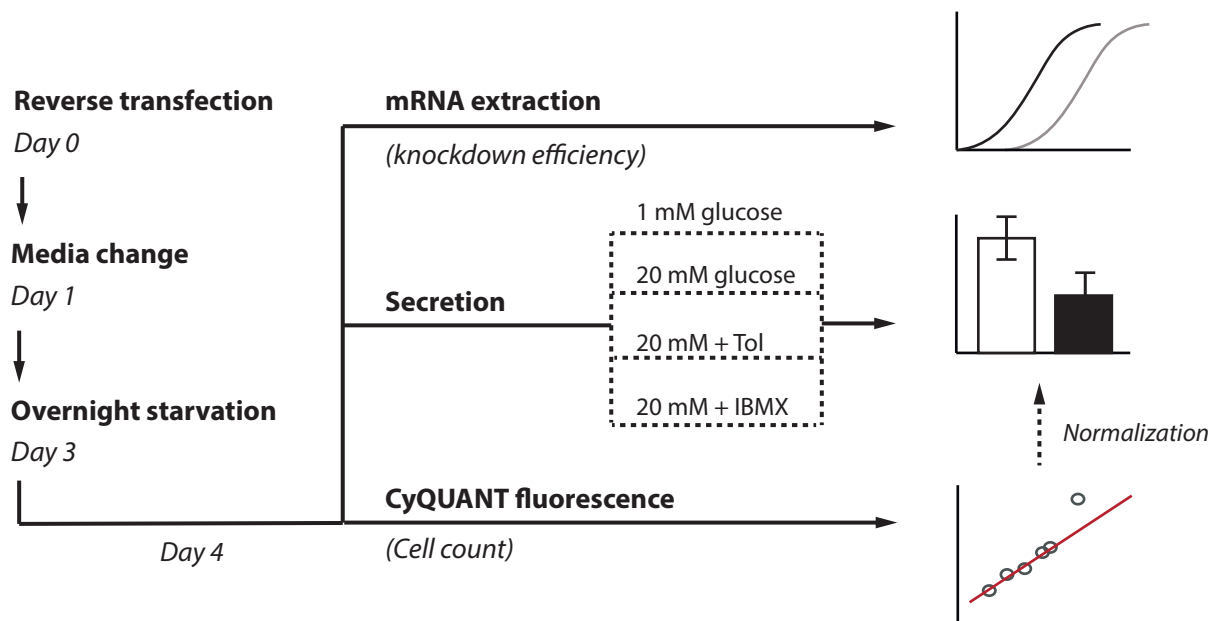
Screenshots from the Human Islet Regulome Browser for the following loci: (A) *ARL15*, (B) *ZMIZ1*, (C) *THADA*, and (D) *HNF4A*, discussed in the main text and Fig. 3 (10). The plots show annotations of chromatin state and transcription factor binding sites in primary human islets. Binding sites for the key β -cell transcription factors PDX1, NKX2.2, FOXA2, NKX6.1, MAFB are shown. Variant association with type 2 diabetes risk for the MAGIC (*blue*) and DIAGRAM (*red*) datasets are indicated in the Manhattan plots (11, 12).

Supplementary references

- [1] A.C. Nica, et al., Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. *Genome Res* 23 (2013) 1554-62.
- [2] B.E. Stranger, et al., Population genomics of human gene expression. *Nat Genet* 39 (2007) 1217-24.
- [3] A.S. Dimas, et al., Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 325 (2009) 1246-50.
- [4] J.B. Veyrieras, et al., High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet* 4 (2008) e1000214.
- [5] Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348 (2015) 648-60.
- [6] S. Lamble, et al., Improved workflows for high throughput library preparation using the transposome-based Nextera system. *BMC Biotechnol* 13 (2013) 104.
- [7] D. Kim, et al., TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14 (2013) R36.
- [8] J. Harrow, et al., GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 22 (2012) 1760-74.
- [9] D.S. DeLuca, et al., RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* 28 (2012) 1530-2.
- [10] L. Pasquali, et al., Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet* 46 (2014) 136-43.
- [11] A.P. Morris, et al., Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 44 (2012) 981-90.
- [12] B.F. Voight, et al., Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 42 (2010) 579-89.

Figure S1

A



B

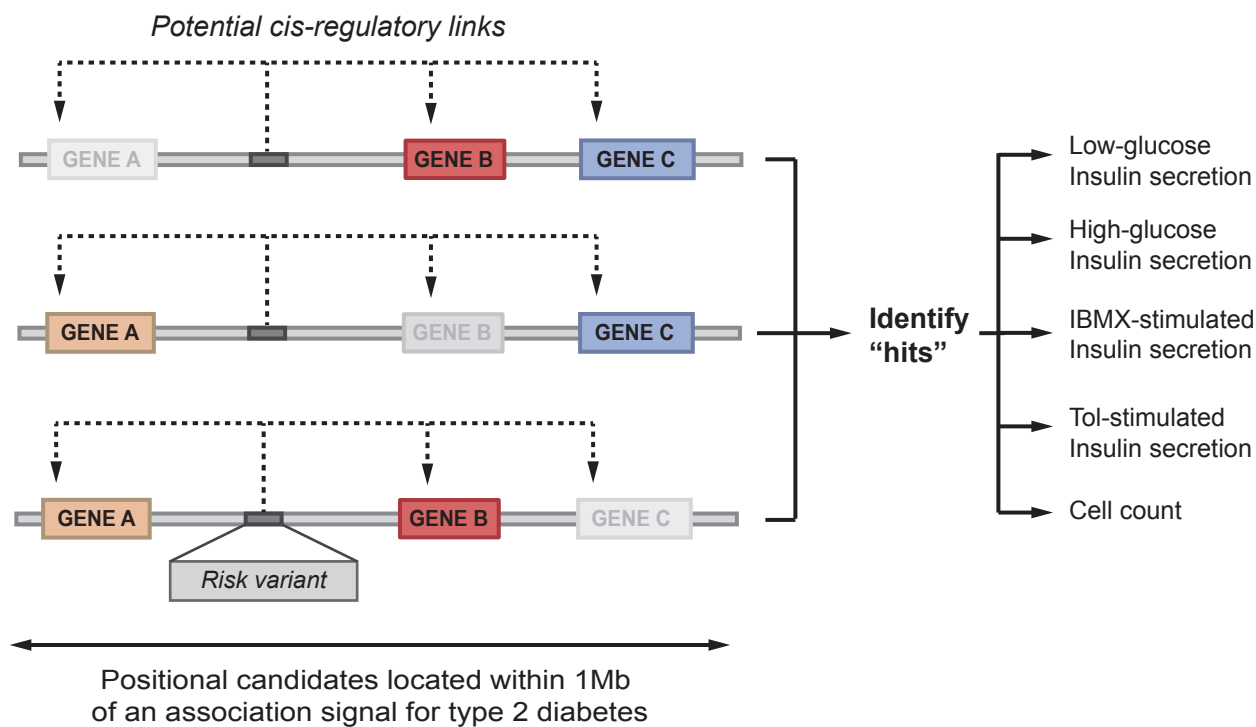
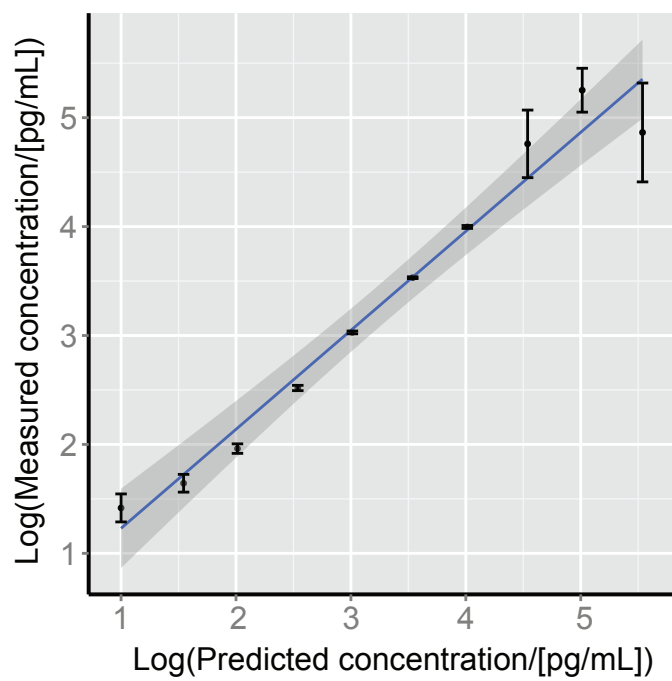


Figure S2

A



B

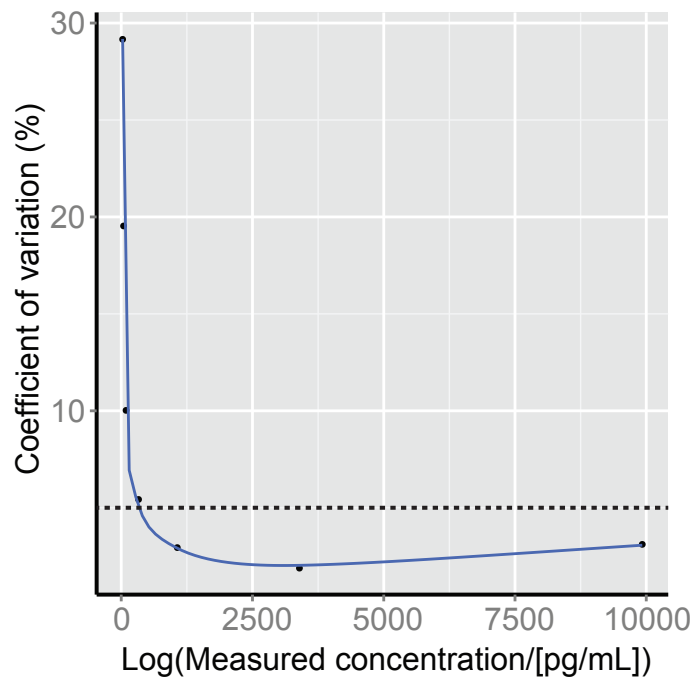


Figure S3

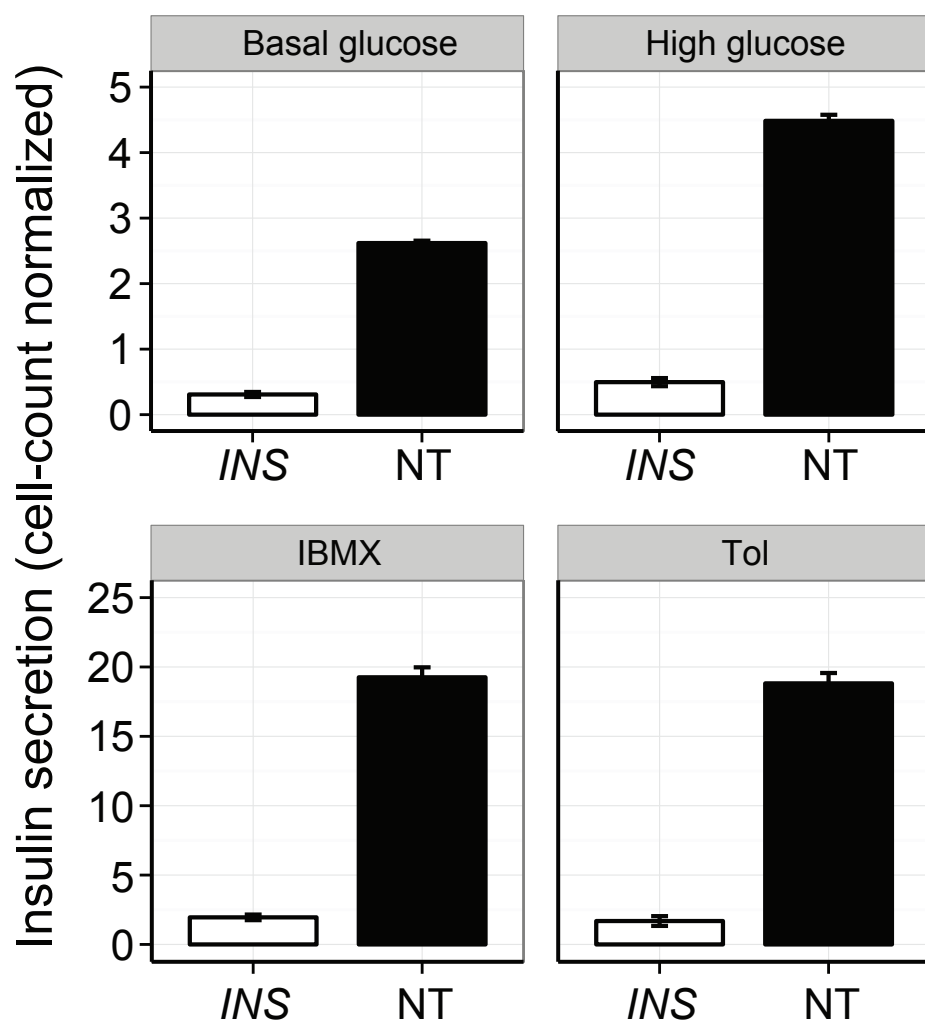


Figure S4

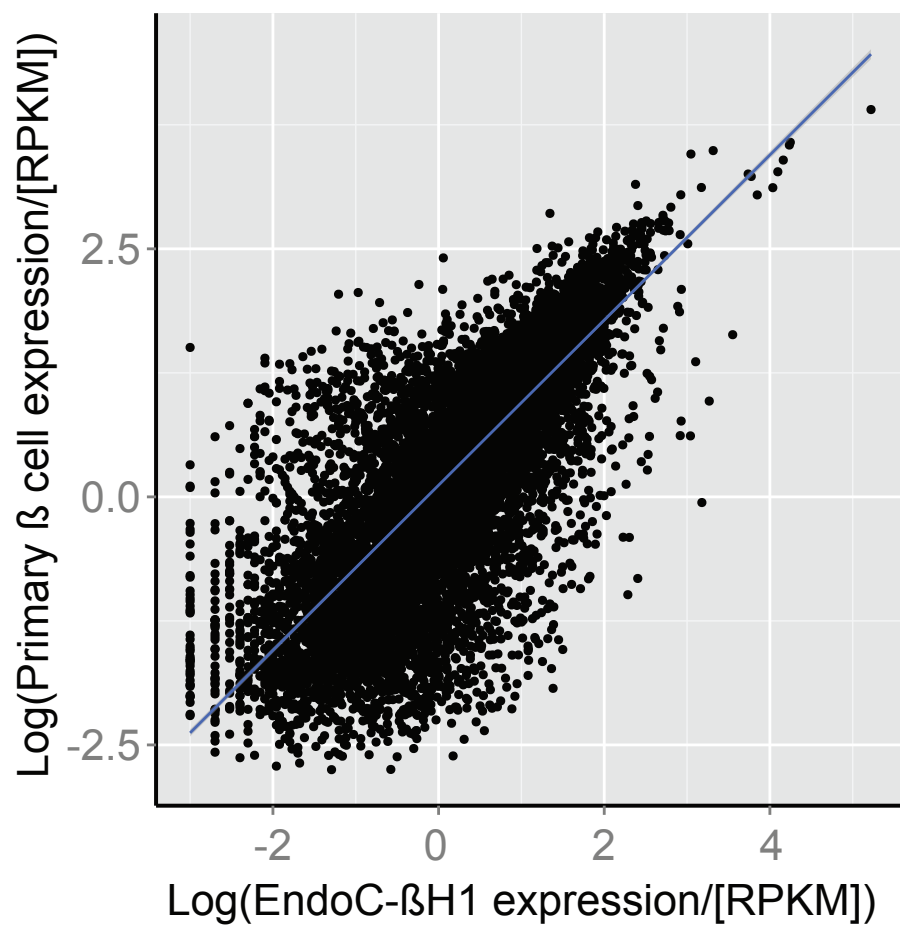


Figure S5

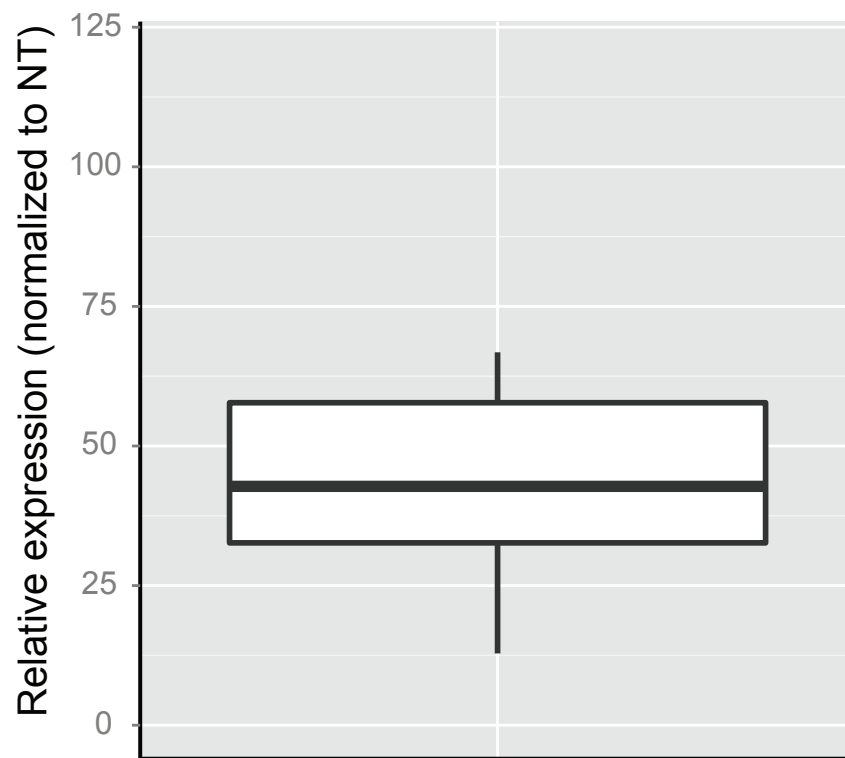


Figure S6

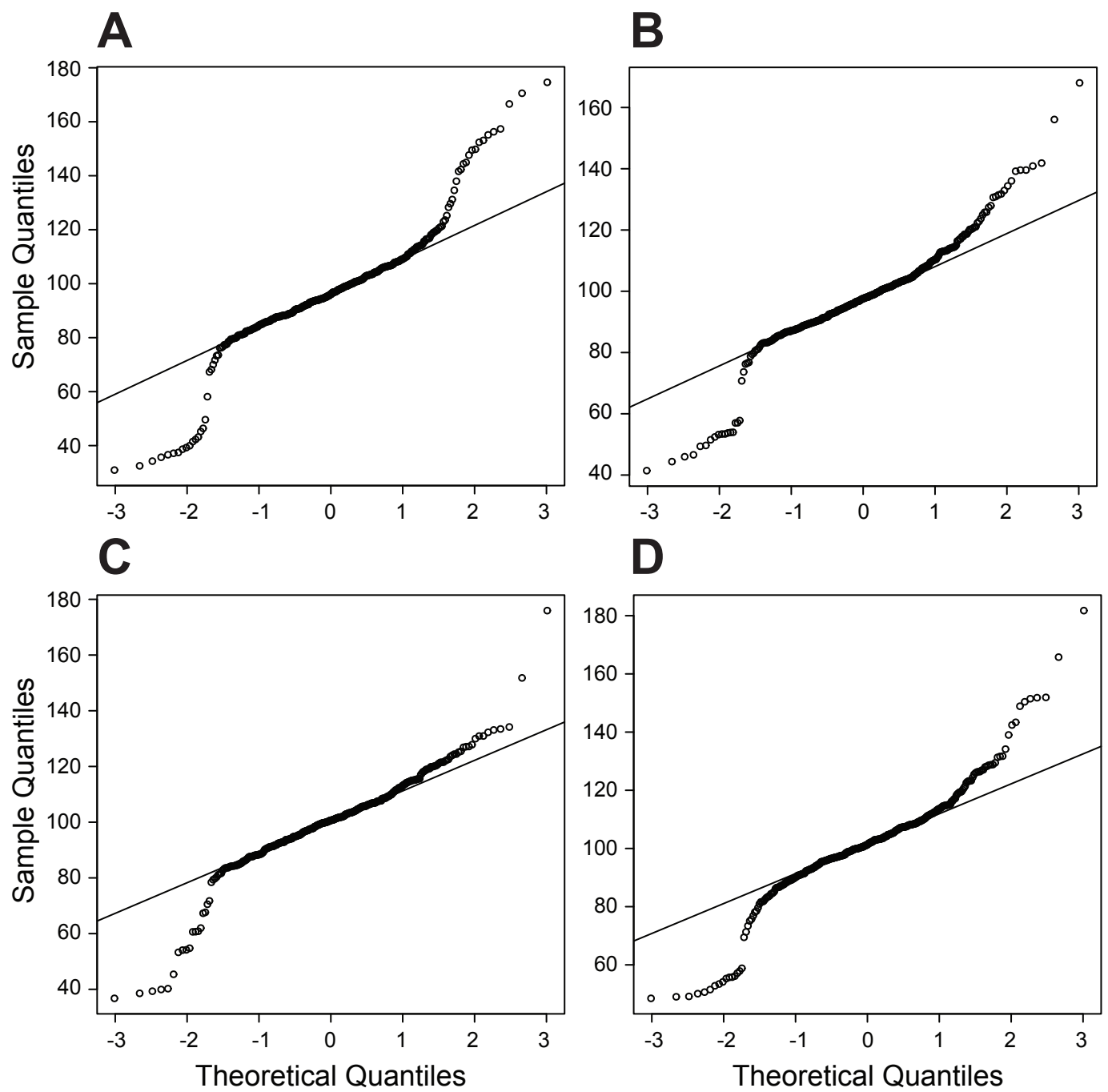


Figure S7

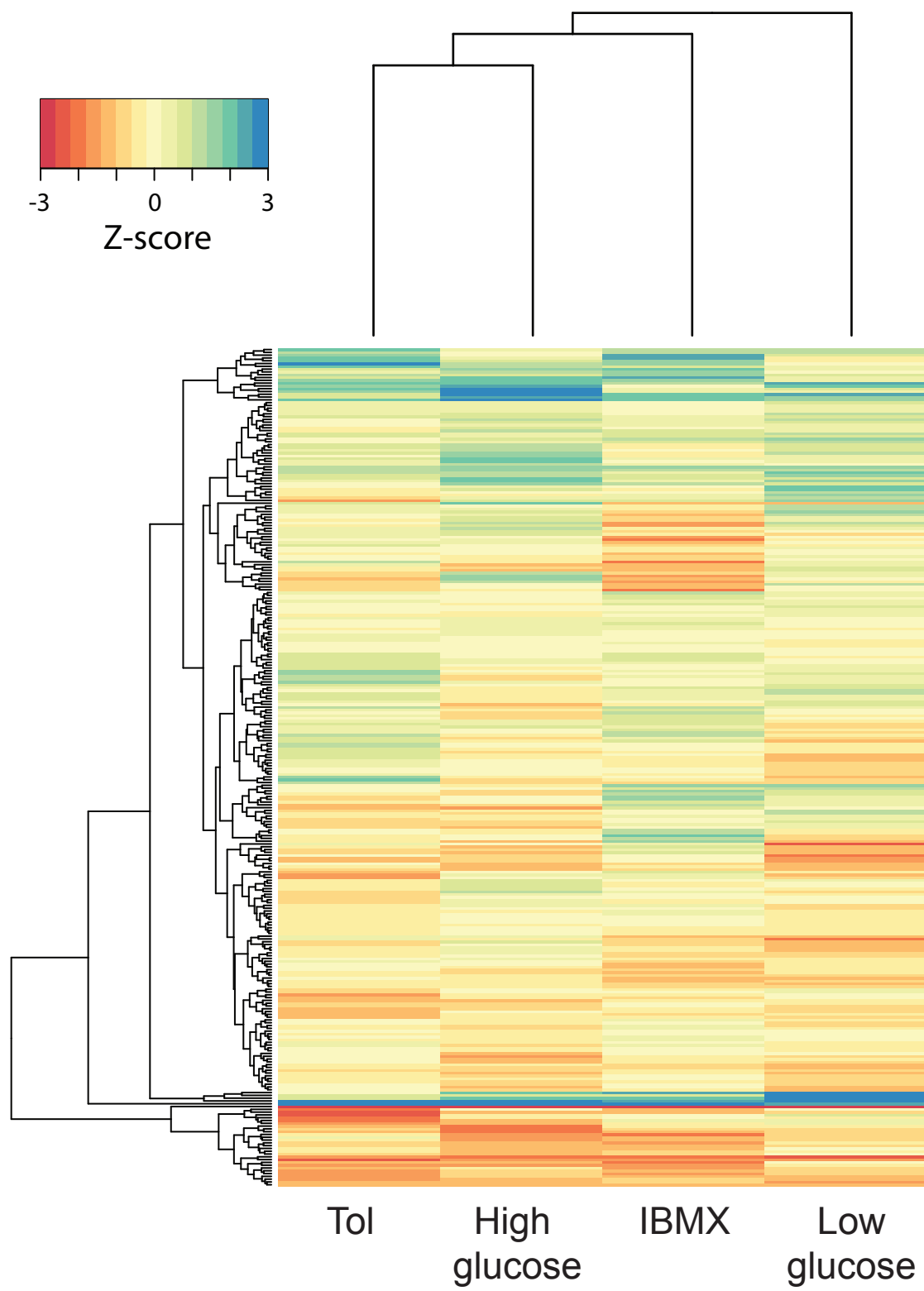


Figure S8

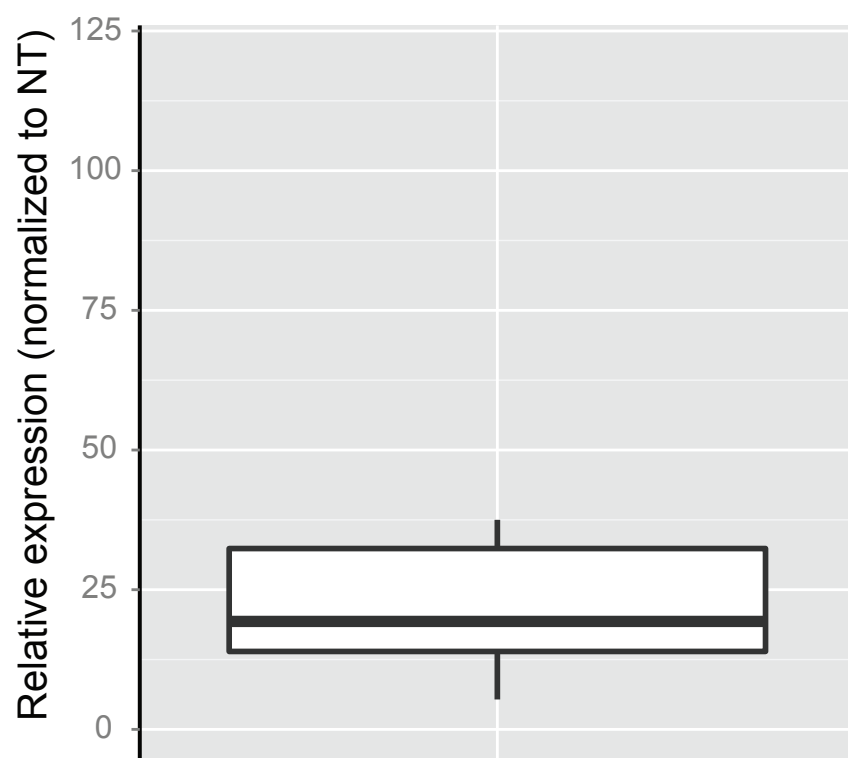


Figure S9

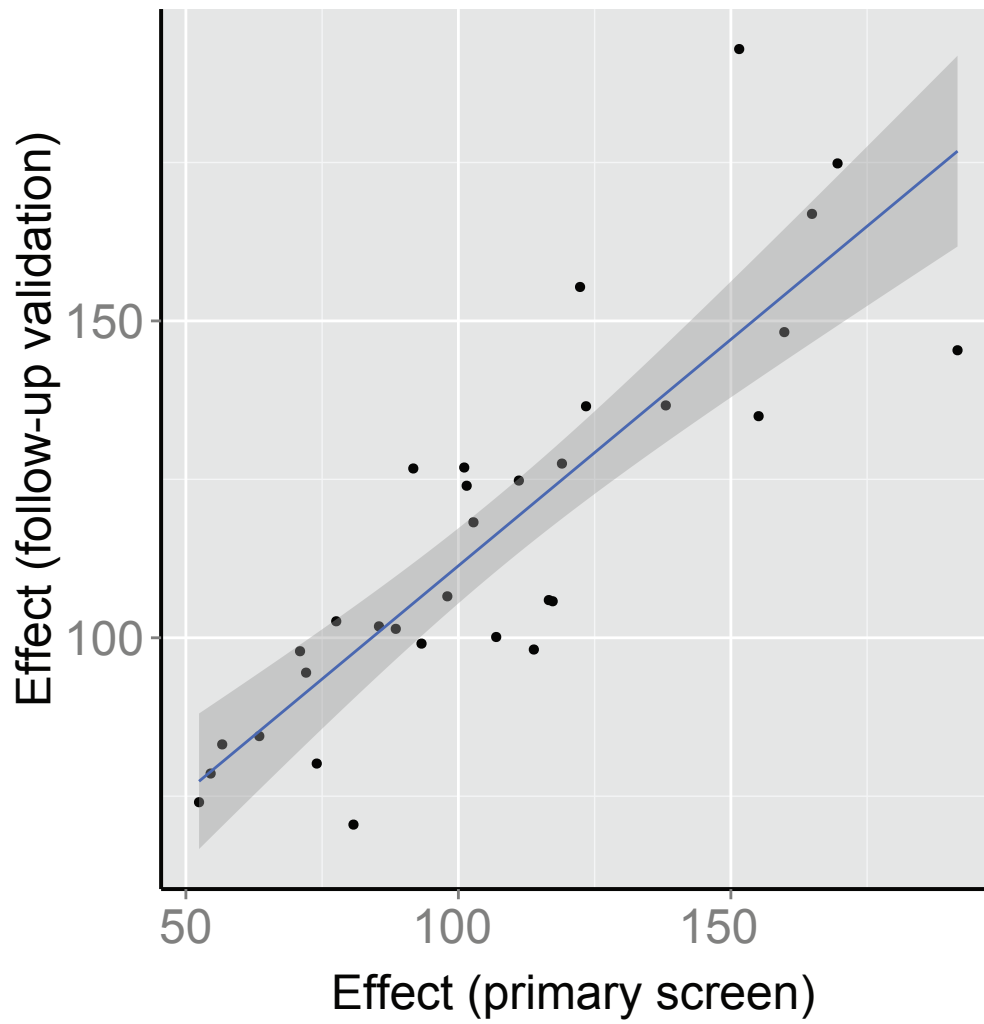


Figure S10

