A genome-wide CRISPR screen identifies regulators of beta cell function involved in type 2 diabetes risk

1   Antje K Grotz[1], Elena Navarro-Guerrero[2], Romina J Bevacqua[3,4], Roberta Baronio[2], Soren K

2   Thomsen[1], Sameena Nawaz[1], Varsha Rajesh[4,5], Agata Wesolowska-Andersen[6], Seung K Kim[3,4],

3   Daniel Ebner[2], Anna L Gloyn[1,4,5,6,7*]

1. Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, OX3 7LE, UK.

2. Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, UK.

3. Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA.

4. Stanford Diabetes Research Centre, Stanford School of Medicine, Stanford University, Stanford, CA, USA

5. Department of Pediatrics, Division of Endocrinology, Stanford School of Medicine, Stanford University, Stanford, CA, USA.

6. Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK.

7. Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, OX3 7LE, UK.

Correspondence: Anna L. Gloyn, Division of Endocrinology, Department of Pediatrics, Stanford School of Medicine, Stanford University, Stanford, CA, USA. agloyn@stanford.edu

Abstract

Identification of the genes and processes mediating genetic association signals for complex disease represents a major challenge. Since many of the genetic signals for type 2 diabetes exert their effects through pancreatic islet-cell dysfunction, we performed a genome-wide pooled CRISPR loss-of-function screen in human pancreatic beta cells. We focused on the regulation of insulin content as a disease-relevant readout of beta cell function. We identified 580 genes influencing this phenotype: integration with genetic and genomic data provided experimental support for 20 candidate type 2 diabetes effector transcripts including the autophagy receptor *CALCOCO2*. Our study highlights how cellular screens can augment existing multi-omic efforts to accelerate biological and translational inference at GWAS loci.

Main

Genome-wide association studies (GWAS) have delivered thousands of robust associations for type 2 diabetes (T2D) and related glycemic traits but most map to non-coding regions and are presumed to have a regulatory function[1]. Incomplete fine-mapping efforts mean that most GWAS loci are not mapped to a single causal variant but rather include multiple variants in a credible set, each of which could potentially influence gene expression in a different cellular context. The typical step after fine-mapping involves attempts to connect the putatively causal variants – and the regulatory elements in which they reside – to the genes they regulate, using methods such as cis-eQTL colocalization, single cell chromatin co-accessibility, and DNA proximity assays (such as HiC)[2-5]. These approaches however have limitations. The two most important are the cell type- and context-dependency of these regulatory connections (which means that assays conducted in inappropriate cell-types or states may reveal variant-to-gene connections that have nothing to do with disease pathogenesis) and molecular pleiotropy (which means that variants of interest may regulate transcription of several genes in cis, obscuring the identity of the transcript responsible for disease mediation). These correlative approaches can generate hypotheses about candidate effectors but typically fall short of providing definitive evidence.

Perturbation studies can provide more compelling evidence of causation, but only if conducted in authentic models and using disease-relevant phenotypic readouts[6]. The strongest such evidence arises from disease-associated coding variants that provide a readout in free-living humans of the consequences of perturbations of gene and protein function, but the low frequency of most such variants limits this approach[2]. The availability of human cellular models and CRISPR based technologies provides an increasingly attractive alternative approach for generating genome-wide profiles of the phenotypic consequences of gene perturbation and thereby inform understanding of disease biology[6]. Central to this aspiration is confidence in the disease relevance of a cell type. For T2D, both physiological and epigenomic evidence highlight the central role of pancreatic islets, and consequently the insulin producing beta cells, in mediating disease risk [7-11]. Substantial differences between mouse and human islets (architecture) and beta cells (ion channel composition, cell cycle

3

61   regulation and glucose transporter expression) argue for the use of human tissue and cell lines[12–17]. We

62   and others have generated large transcriptomic and epigenomic resources in this key human tissue

63   enabling genome-wide integration of genetic and genomic data which has helped to identify candidate

64   effector transcripts at T2D GWAS loci [8,11,18,19]. We now complement these resources with a genome-

65   wide CRISPR loss-of-function (LoF) screen in the well characterized human pancreatic beta cell line

66   EndoC-βH1 to identify genes which regulate insulin content [20–22]. Using this resource, we provide

67   cellular evidence to support the causal role of 20 genes at T2D loci in beta cell function and identify a

68   novel role for the autophagy receptor *CALCOCO2* in human beta cell function.

69   **Results**

70   **Development of a pooled CRISPR assay for human beta cell function**

71   Glucose stimulated insulin secretion is the primary measure of beta cell function but is an unsuitable

72   phenotypic readout for pooled high-throughput screening as secreted insulin cannot be linked directly

73   to its cellular source. Changes in insulin content are also causally linked to T2D but unlike insulin

74   secretion can be quantified using FACS.[23] We therefore developed an assay for this closely-related

75   and disease-relevant phenotype, endogenous intracellular insulin content, to enable fluorescence

76   based cellular sorting using a primary monoclonal insulin antibody and a fluorescently labelled

77   secondary antibody. Using this assay as our cellular phenotypic readout, we designed a pooled

78   CRISPR LoF screening pipeline in the human pancreatic beta cell line EndoC-βH1 based on

79   introducing single gene KOs through a lentiviral genome-wide CRISPR library followed by antibiotic

80   selection to remove untransduced cells (Fig. 1a). Following gene knockout, cells were FAC-sorted

81   into populations containing either low or high levels of intracellular insulin and the stably integrated

82   sgRNAs within these cell populations identified through next-generation sequencing.

83   To ensure the FACS based insulin content readout in EndoC-βH1 was highly specific and sensitive,

84   we chose an antibody and staining protocol based on comparing several complementary strategies

85   (Supplementary Fig S1). Near complete siRNA-based knockdown of *INS* in EndoC-βH1 resulted in

86   two populations with distinct average fluorescence intensities, validating the sensitivity of the INS

87   antibody (Fig 1b,c). The level of separation was in concordance with an independent alphaLISA

4

88   based quantification of insulin content demonstrating around 46% of residual protein after knockdown

89   (Fig 1d). The specificity of the antibody was further confirmed in the non-*INS* expressing human

90   embryonic kidney cell line HEK293T (Fig 1e). Finally, the screening protocol was validated by

91   targeting three genes in a small-scale screen including the positive control genes *INS* and *PAM*, which

92   are known to induce a reduction in insulin content upon deletion, and *NMS*, a negative control which

93   is not expressed in EndoC-βH1[23] (Fig 1f). Compared to an empty vector (EV) control coding for Cas9

94   only, the cells demonstrated, as expected, reduced INS FACS signal and enrichment of *INS* and *PAM*

95   sgRNA within the INS^{LOW} population, whereas sgRNAs targeting *NMS* showed no difference (Fig

96   1g,h). Consistent with its lower effect on insulin content as shown in previous studies, sgRNAs

97   targeting *PAM* demonstrated a highly replicable but lower enrichment compared to sgRNAs targeting

98   *INS*, further confirming the suitability of the pipeline for genome-wide screening.

A FACS-based CRISPR screen identifies regulators of insulin content

100   We performed two independent genome-wide screen replicates in the slow-growing (doubling in 7

101   days) human beta cell line, EndoC-βH1, using the Toronto KnockOut version 3.0 (TKOv3) CRISPR

102   library targeting 18053 protein-coding human genes with four sgRNAs per gene[24]. Per replicate, a

103   minimum of 700 million cells were lentivirally transduced at a low multiplicity of infection (MOI)

104   with a library coverage of around 500 cells per guide, thus requiring more than 4 months to obtain

105   sufficient cells for the screen. Cells were subjected to puromycin selection for seven days followed by

106   fixation, permeabilization, staining and FACS sorting (Fig. 1a). Compared to EV (Cas9) control cells,

107   the cell population transduced with the CRISPR library demonstrated a wider insulin signal

108   distribution, particularly towards lower insulin, indicative of altered insulin content due to CRISPR

109   KO mediated effects (Fig. 2a, Supplementary Fig S2). Two cell populations were collected, those

110   with low insulin content (INS^{LOW}) and those with high insulin content (INS^{HIGH}), followed by DNA

111   extraction, sgRNA amplification and sequencing (Supplementary Fig S3). We used the MAGeCK

112   algorithm to compare sgRNA abundance and identify sgRNAs that were enriched or depleted in

113   INS^{LOW} cells relative to INS^{HIGH} cells[25]. Enriched sgRNAs identify genes which are positive

114   regulators leading to a reduction in insulin content upon gene KO, and depleted sgRNAs identify

115 negative regulators associated with an increase in insulin content upon gene KO. To reduce the

116 number of false positive hits, genes were only classified as screening hits if sgRNAs demonstrated

117 consistent effects across replicates and met the stringent threshold (FDR<0.1) for at least two of the

118 four sgRNAs in both independent screening replicates. A total of 580 genes fulfilled these criteria: we

119 considered these as robust, reproducible screening hits and took them forward for further evaluation

120 (Supplementary Table 1).

121 To evaluate the biological relevance of these screen hits, we asked if our screen was able to identify

122 genes known to be involved in insulin regulation and beta cell function. As expected, sgRNAs

123 targeting *INS* were enriched in the INS^{LOW} population, confirming the sensitivity of the screen (Fig.

124 2b,e,f). The overall enrichment of *INS* was lower than that of other known regulators of insulin

125 secretion as only three out of the four sgRNAs induced an effect (Fig. 2b,f). Other established

126 regulators of beta cell function and genes with known roles in monogenic forms of diabetes or T2D-

127 risk were identified among the screening hits, such as the beta cell transcription factor *NKX2.2*, the

128 glucose transporter *SLC2A2*, the cellular stress regulatory subunit *PPP1R15B* and the insulin-like

129 growth factor *IGF2* (Fig. 2c,e,f) [26–33]. Non-targeting control sgRNAs showed no effect on insulin

130 content (Fig. 2d). In addition to direct regulators of insulin content, the screen also identified multiple

131 general regulators of transcription and translation with indirect effects on the phenotype, such as the

132 translation initiation factor *EIF3B*, the RNA polymerase II subunit *POLR2G*, the transcription

133 coactivator *TAF7* and the ribosomal protein *RPL12* (Fig. 2f), all of them classified as common

134 essential genes in non-beta cell CRISPR screens with longer culturing durations [34–36].

135 **Network analysis of CRISPR screening hits**

136 We next considered whether the identified screening hits were enriched for functional classifications

137 involved in the regulation of insulin content, by undertaking pathway and protein network analysis.

138 Gene ontology (GO) term and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis

139 highlighted insulin and beta cell function associated categories such as 'Type II diabetes mellitus',

140 'Insulin receptor signaling' or 'Calcium mediated signaling', while transcription and translation

141 related categories consistent with modifying total intracellular protein levels were also enriched (Fig.

142   3a)[37,38]. STRING protein network analysis further emphasized the fundamental role of INS within the

143   screen as a central node with several independent connections (Fig. 3b)[39]. The screening hits were

144   significantly enriched for involvement in shared protein networks, providing additional confidence in

145   the sensitivity of the screen to identify interrelated complexes (Fig. 3c, Supplementary Fig S4). Many

146   of the hits mapped to the functional categories of ubiquitin mediated proteolysis and autophagy,

147   mitochondrial ATP synthesis, vesicle trafficking and exocytosis, GPCR signaling, lipid metabolism

148   and the MAPK signaling pathway, all of them containing both previously unknown regulators of

149   insulin content and those with known roles in insulin secretion or beta cell function such as

150   *RFWD2/COP1, HUWE1, FFAR1, TIAM1, SGPP2, MAPK1, IRS2, PCNT* and *PTPN1*[40–48]. The T2D

151   associated gene *TBC1D4* mapped to a GTPase cluster and whilst the effects of this gene on T2D

152   pathogenesis have been considered to be primarily associated with insulin resistance, the effect on

153   insulin content observed in this screen is consistent with previous studies indicating a crucial role

154   within beta cells[49–54]. The GTPase network also contained *ARAP1*, a gene previously linked to T2D

155   susceptibility adding another dimension to the interesting story at this locus where deletion or

156   overexpression in mice has failed to detect an impact on beta cell function pointing towards the

157   nearby genes *STARD10* or *FCHSD2* as modulators of insulin secretion and mediators of disease risk

158   at this GWAS locus[55–57].

159   **Integrative analysis of CRISPR screening hits with effector transcript prioritization**

160   **methods**

161   Next, we sought to use our cellular screen as complementary perturbation evidence to support the

162   assignment of effector genes at T2D loci. We applied three complementary approaches which have

163   been developed to combine genetic association results for T2D with diverse sources of genetic and

164   genomic data and thereby generate lists of the "effector" gene(s) most likely to mediate the genetic

165   association at each signal[58–60]. All methods are described and the predicted effector gene lists

166   provided on the T2D Knowledge Portal (https://t2d.hugeamp.org) [61]. The union of the effector gene

167   assignments from the three methods generated a list of 336 candidate effector transcripts at T2D loci,

168   with no candidates identified by all three methods. The intersection of this list with the set of screen

169 hits, yielded 20 genes (3% of screening hits and 6% of predicted effector transcripts, Supplementary

170 Table 2) for which our screen provided biological evidence for a role in a disease relevant phenotype.

171 None of these 20 genes were identified by all 3 prediction methods. Of these 20 genes, 5 (25%) are

172 assigned as "causal" by the curated *heuristic* effector gene prediction method with a further 4 (20%)

173 assigned between strong, moderate and possible.

174 To compare how our perturbation screen performed against other commonly applied approaches for

175 effector transcript prioritization, we explored, a relatively modest albeit the largest to date, eQTL

176 study of 420 human islet donors[19]. In this study, colocalization between T2D-GWAS and islet cis-

177 eQTL signals was supported by two methods (COLOC and RTC) at 20 loci, rising to 39 when

178 statistical support for colocalization was required from only one of the approaches. Whilst both

179 approaches identified similar numbers of genes, no gene was common between them, suggesting that

180 the islet eQTLs are unlikely to manifest as reduced insulin content and additional cellular phenotypes

181 (e.g. insulin secretion or ER stress) may underly their effect. These observations support the inclusion

182 of multiple approaches for optimal effector gene prioritization and encourage both larger eQTL

183 studies to detect further signals, which could then intersect with hits from this screen, and an

184 expansion of targeted and/or genome-wide screens for additional cellular phenotypes (e.g. insulin

185 secretion) across disease relevant cell types.

186 **Loss of the T2D associated gene *CALCOCO2* reduces insulin content**

187 We focused our functional studies on the screening hit *CALCOCO2* (encoding Calcium Binding and

188 Coiled-Coil Domain protein) on the basis that its role in the human beta cell has not been explored

189 (Fig. 4a)[18,62–65]. *CALCOCO2* is a ubiquitin-binding autophagy receptor and plays a crucial role in

190 selective autophagy by linking the degradation target to the autophagic machinery[66,67]. So far, the

191 gene has been shown to initiate autophagy for invading pathogens (xenophagy) and damaged

192 mitochondria (mitophagy)[66–68]. *CALCOCO2* maps to the T2D GWAS locus usually named for the

193 gene *TTLL6,* but fine-mapping data highlighted a coding variant within *CALCOCO2* (rs10278,

194 p.Pro347Ala) as the most likely causal variant[9]. Physiological clustering (based on patterns of

195 association of the T2D-risk allele with other diabetes-related phenotypes) has revealed that this locus

8

196   is likely to exert its primary effect on disease risk through the modulation of beta cell function[9]. In this

197   CRISPR screen, KO of *CALCOCO2* resulted in consistent enrichment of *CALCOCO2* targeting

198   sgRNAs in the low insulin content sample, corresponding to a decrease in insulin content (Fig. 4b).

199   None of the other genes at the *CALCOCO2* GWAS locus showed an effect on insulin content in the

200   screen, further supporting *CALCOCO2* as the likely effector transcript at this T2D risk locus

201   (Supplementary Fig S5).

202   We obtained independent confirmation of the effect of *CALCOCO2* loss on insulin content through

203   siRNA-based silencing in EndoC-βH1. Silencing of *CALCOCO2* was highly efficient and induced a

204   mean mRNA and protein reduction of 78.1% and 80.0%, respectively (Fig. 4c-e). Reduction of

205   *CALCOCO2* did not affect cell viability (Fig. 4f). Intracellular insulin content was assessed using two

206   independent antibody-based detection methods. Consistent with the results of the CRISPR screen and

207   highlighting its sensitivity, insulin content was significantly reduced in *CALCOCO2* silenced cells by

208   24.0% and 38.0% as measured by an alphaLISA and western-blot based detection method,

209   respectively (Fig. 4g-i). This reduction in insulin content was not due to decreased insulin gene

210   expression nor was it due to altered proinsulin processing as the ratio of proinsulin to insulin was

211   unaffected (Fig. 4j-m).

CALCOCO2 is required for human beta cell function

213   While we have shown that insulin content is reduced in EndoC-βH1 upon loss of *CALCOCO2*, altered

214   glucose stimulated insulin secretion cannot be concluded based on changes in insulin content alone.

215   We therefore independently assessed the effects of *CALCOCO2* knockdown on insulin secretion in

216   EndoC-βH1. Across different glucose and K_{ATP} modulator conditions (tolbutamide and diazoxide),

217   insulin secretion was significantly reduced by 39.3% upon silencing of *CALCOCO2* (Fig. 4n). While

218   total insulin secretion was affected, the glucose stimulation index from low to high glucose was

219   unchanged, indicating an appropriate cellular response to changes in glucose concentrations (Fig. 4o).

220   Normalization of insulin secretion to insulin content reduced the difference between *CALCOCO2*

221   silenced cells and controls to 30.2%, indicating that the level of insulin content played a partial role in

222   influencing the insulin secretory response (Fig 5p,q).

9

223   We extended our studies to determine *CALCOCO2* expression and localization in primary human

224   pancreatic islets. The gene is expressed at a similar level as other critical beta cell genes such as

225   *KCNJ11*, in both EndoC-βH1 and primary human islets[69]. Immunofluorescence studies in human

226   pancreas sections confirmed protein localization to the cytoplasm of beta cells through colocalization

227   with INS (Fig. 5a). Consistent with its ubiquitous role across cell types, exocrine ductal and acinar

228   cells also stained positive for CALCOCO2 (Supplementary Fig S6).

229   To determine whether the effects of *CALCOCO2* loss can be replicated beyond *in vitro* models of

230   human beta cells and in other secretory islet cell types, we performed shRNA knockdown of

231   *CALCOCO2* in primary human islets from two independent human donors. Islets were dispersed,

232   lentivirally-transduced with shRNA targeting *CALCOCO2* (sh*CALCOCO2*) and reaggregated to form

233   pseudoislets (Fig. 5b). Successful transduction of pseudoislets was confirmed by assessing GFP co-

234   expression (Fig. 5c). *CALCOCO2* expression was on average reduced by 54.0% compared to

235   pseudoislets transduced with non-targeted control shRNA (shNT, Fig. 5d). In line with results in

236   EndoC-βH, intracellular insulin demonstrated a modest but consistent reduction of 15.7% (Fig. 5e). If

237   the lower silencing efficiency in pseudoislets is taken into account, insulin content is affected to the

238   same extent in both the experimental model and primary human tissue. Upon incubation with a range

239   of glucose concentrations, pseudoislets failed to demonstrate significantly affected insulin secretion

240   (Fig. 5f,g). Stimulation with high glucose and the secretion potentiator IBMX increased secretion of

241   insulin upon normalization to content in sh*CALCOCO2*, this effect was lost upon normalization to

242   genomic DNA and therefore reflects the impact of changes in insulin content on secretion. We

243   similarly assessed glucagon content and secretion to identify potential effects of *CALCOCO2* in alpha

244   cells (Fig. 5h-j). While sh*CALCOCO2* showed some trend towards blunted glucagon response to

245   glucose starvation, none of the secretory conditions or glucagon content were significantly changed

246   upon *CALCOCO2* reduction. Together, our findings show that *CALCOCO2* regulates human islet beta

247   cell function.

248 **The T2D-associated genes *QSER1* and *PLCB3* do not affect insulin content**

249 To confirm that this screen correctly identified T2D associated genes which were likely to exert their

250 effect on disease risk through modulating insulin content and beta cell function, we examined

251 additional genes which have been predicted to be causal for T2D on the T2D Knowledge Portal but

252 were not identified as screening hits (Fig. 6a)[59]. The prioritized genes *QSER1* and *PLCB3* both

253 contain coding variants consistent with a causal role in T2D but their mechanism of effect or tissue of

254 action has not yet been resolved [9,70]. *PLCB3* plays a role in the amplification pathway in the beta cell

255 but has also been associated with insulin action in a multi-trait physiological clustering approach[9,71,72].

256 *QSER1* has recently been identified as a DNA methylation regulator associated with pancreatic

257 differentiation defects[73]. In the screen, neither of these genes showed a consistent sgRNA depletion or

258 enrichment in either insulin content populations and were not classified as hits (Fig. 6b,c). We again

259 used siRNA knockdown in our targeted follow-up studies and reduced gene expression by 78.3% and

260 51.2% upon targeting of *QSER1* and *PLCB3*, respectively (Fig. 6d). Consistent with our negative

261 screening results, knockdown of either gene did not affect intracellular insulin content (Fig. 6e). In

262 addition, INS expression and insulin secretion under different glucose conditions or KATP modulator

263 conditions were unchanged (Fig. 6f, Supplementary Fig S6). Together, this genome-wide CRISPR

264 screen was able to correctly identify the insulin regulatory function of *CALCOCO2* whilst also

265 distinguishing between cellular phenotypes related to islet-cell development (Q*SER1*) or insulin action

266 (*PLCB3*).

267

Discussion

Recent efforts to identify effector transcripts at GWAS association signals have concentrated on the integration of disease relevant transcriptomic and epigenomic datasets or focused on detailed mechanistic studies at a single locus [8,11,78,18,19,23,63,74–77]. In most cases these studies are unable to distinguish between pleiotropic effects of variants on regulatory elements and gene expression which are inconsequential for the disease phenotype in question. It is increasingly recognized that genome-wide perturbation datasets in disease relevant cell types could close the gap between effector transcript and disease relevant biology and thus enable focused translational efforts. For T2D, there is substantial evidence to support a central role for defects in pancreatic islet-cell dysfunction along with an understanding of disease relevant cellular phenotypes thus opening the possibility to generate relevant genome-wide cellular perturbation datasets. Here, we present a genome-wide pooled CRISPR loss-of-function screen applying a functionally relevant FACS readout to perform a comprehensive characterization of regulators of insulin content in the human beta cell line EndoC-βH1.

Our pooled genome-wide CRISPR screening approach was successful in identifying robust modulators of insulin content as shown through the detection of not only the insulin gene (*INS*) itself but also genes involved in monogenic types of diabetes or known regulators of insulin transcription and secretion such as the transcription factor *NKX2.2* or the glucose transporter *SLC2A2*. Through the integration of our screening hits with the results from three different currently available effector gene prioritization tools, 20 genes were identified as effector transcripts working through beta cell dysfunction. Our data add to a growing number of genome-wide multi-omic datasets which can be harnessed to close the gap between genetic discovery and biological mechanisms and serve as a model for future cellular studies not only in human beta cells but also in other disease relevant cell types. Furthermore, our unbiased genome-wide data will be relevant for other traits/diseases where the pancreatic beta cell plays a role, including type 1 diabetes.

We initially identified the T2D risk associated gene *CALCOCO2* as a CRISPR screening hit and a novel positive regulator of insulin content. We selected this gene for further study on account of its potential to underlie a T2D GWAS signal at the *CALCOCO2* locus and due to a lack of knowledge on

12

295   its role in beta cell biology. Our focused functional studies using complementary siRNA-based

296   knockdown not only confirmed its effect on insulin content but also revealed a role in insulin

297   secretion. We therefore provide, for the first time, functional data connecting *CALCOCO2* to beta cell

298   function, a likely effect on disease pathogenesis and tissue of action, expanding the evidence for a

299   causal role beyond its genetic coding association with T2D risk [9,70,79]. Detailed mechanistic studies

300   will be the focus of future investigations but *CALCOCO2*'s role as a selective autophagy receptor hint

301   at defective insulin secretion because of dysregulated autophagy. *CALCOCO2* is an essential receptor

302   to induce Parkin-mediated mitophagy (the degradation of dysfunctional mitochondria) and the

303   observed effect on insulin secretion is consistent with other studies such as for the mitophagy

304   regulator *Clec16a* [68,80–83]. In addition, the mitophagy regulators *SQSTM1* and *TBK1* were also

305   identified as screening hits, further highlighting the relevance of this cellular pathway. In addition to

306   its association with T2D risk based on GWAS and its intracellular role, *CALCOCO2* has also been

307   implicated in studies of the plasma proteome, demonstrating that reduced circulating levels of

308   *CALCOCO2* are associated with increased T2D risk [84,85]. Plasma proteins are key druggable targets

309   and disease biomarkers, positioning *CALCOCO2* at the center of investigations focusing not only on

310   understanding the underlying functional mechanism of T2D but also potential therapeutic or

311   diagnostic strategies [86].

312   Our CRISPR screen in the human beta cell line EndoC-βH1 identified more than half of all protein

313   networks that were also identified in a previous CRISPR screen assessing insulin content in a mouse

314   insulinoma cell line, highlighting conserved interspecies regulatory networks [87]. While shared protein

315   networks were identified in both CRISPR screens, only a small proportion of screening hits (7% or 43

316   genes) were overlapping, highlighting the importance of performing the screen in a human cellular

317   model system but also potentially diverging experimental strategies. Of the 20 genes from our

318   CRISPR screen that were overlapping predicted effector transcripts, only one gene and the primary

319   phenotypic readout in both screens, *INS*, was also identified by the CRISPR screen in the mouse

320   insulinoma cell line [87].

321   The reason many investigators select to work in rodent beta cells lines is one largely of experimental

322   ease. Although the human beta cell line EndoC-βH1 demonstrates functional characteristics closely

323   resembling human primary cell it is also associated with difficult growth and culture characteristics [20–

324   22]. To obtain perturbation evidence in this physiologically relevant cellular model, we overcame the

325   associated obstacles and variations across passages through highly consistent culturing conditions,

326   integration of two independent replicates, stringent analysis parameters focusing on high

327   reproducibility and rigorous proof of concept studies.

328   This is the first step towards the generation of comprehensive genome-wide perturbation assays in

329   disease relevant human cell lines for T2D and represents a proof of concept that cellular screens can

330   augment genomic efforts linking variants to regulatory elements and transcripts to bridge the gap

331   between gene expression and disease relevant cellular biology. This genome-wide effort in a human

332   cell line should accelerate efforts for further screens, both loss and gain of function, for additional

333   disease relevant cellular phenotypes in authentic human models.

334   In summary, we have developed and performed a genome-wide pooled CRISPR screen in a model of

335   human beta cells, providing a comprehensive perturbation database to associate genes of interest with

336   a direction of effect, tissue of action and functional mechanism. We have successfully applied the

337   screening hits as a prioritization tool for causal genes at T2D GWAS loci and highlight *CALCOCO2*

338   as a novel modulator of insulin secretion and beta cell function with a likely causal role in T2D.

14

Methods

Cell culture

HEK293T were routinely passaged in DMEM 6429 (Sigma-Aldrich) containing 10% fetal calf serum, 100 U/ml penicillin and 100 µg/ml streptomycin. EndoC-βH1 were routinely passaged in 5.5 mM glucose containing growth medium (Sigma-Aldrich) as previously described[88]. All cells were mycoplasma free (Lonza) and grown at 37°C and 5% CO_2.

Human islet procurement

Deidentified human islets and pancreas samples were obtained from three healthy, nondiabetic organ donors procured through the Integrated Islet Distribution Network (IIDP) and the Alberta Diabetes Institute Islet Core (donor information in Supplementary Table 3).

Cloning of individual sgRNAs

Single CRISPR sgRNAs were cloned into plentiCRISPRv2 (Addgene #52961) as previously described[20,89]. Briefly, BsmBI compatible tails (5'CACCGX3' and 5'AAACYC3') were added to complementary sgRNA oligonucleotides and annealed (Supplementary Table 4). PlentiCRISPRv2 was digested using BsmBI (Fermentas), ligated with annealed sgRNA oligonucleotides using Quick Ligase (NEB) followed by transformation and amplification in Stbl3 competent cells (Invitrogen). Correct sgRNA integration was validated through Sanger Sequencing and plasmids were further used to produce lentivirus.

Pooled sgRNA library amplification

Toronto human knockout pooled library (TKOv3) containing 71,090 based on a lentiCRISPRv2 backbone was a gift from Jason Moffat (Addgene #90294). The library was transformed and amplified using 25 µl Endura Competent Cells (Lucigen) and 100 ng plasmid libraries per electroporation reaction as previously described[24]. Transformation efficiency was determined through serial dilution to ensure sgRNA representation and plasmids were subsequently extracted using Plasmid Mega kits (Qiagen). Library representation was confirmed through sequencing on a NextSeq500 (Illumina) using 75 base pair (bp) single end reads.

Lentiviral production and transduction

365

366   Lentivirus for individually cloned sgRNAs and pooled libraries was produced and titered as

367   previously described[20]. Briefly, CRISPR plasmids were co-transfected in 60 % confluent HEK293T

368   with packaging vectors pMD2.G (Addgene #12259), psPAX2 (Addgene #12260) using jetPRIME

369   transfection reagents (Polyplus transfection). Viral supernatant was collected at 48 h and 72 h post

370   transfection and concentrated using ultracentrifugation. The functional viral titer was determined in

371   EndoC-βH1 through measuring the percentage of survival after transduction with different viral

372   dilutions and antibiotic selection. Cell viability was assessed using the CyQUANT Direct Cell

373   Proliferation assay (Invitrogen). The required amount of lentivirus to achieve 26 % survival which is

374   representative of an MOI of 0.3 was calculated and used in subsequent small-scale or genome-wide

375   screen transductions. EndoC-βH1 were transduced for 6 h and selected for 7 days in 4 μg/ml

376   puromycin with media changes as required.

Genome-wide CRISPR screen in EndoC-βH1

377

378   Two independent genome-wide CRISPR screens were performed consecutively in EndoC-βH1 with

379   independent lentiviral CRISPR library transductions, FACS sorting and sgRNA sequencing. Each

380   screen was performed at a coverage of 500 cells per sgRNA and a MOI of 0.3. A total of 670 million

381   and 744 million cells were transduced in replicate one and two, respectively. The cells were

382   transduced as described, incubated in 4 μg/ml puromycin for 7 days with media changes as required to

383   remove dead cells followed by collecting, staining and FACS analysis.

INS intracellular staining and fluorescence-activated cell sorting

384

385   EndoC-βH1 cells were harvested and incubated with LIVE/DEAD Fixable Far Red Dead Cell Stain

386   (Thermo Fisher) for 30 min at room temperature to distinguish live from dead cells and washed in 1%

387   BSA in PBS. The cells were fixed and permeabilized using the BD Cytofix/Cytoperm kit (BD

388   Biosciences) for 20 min at 4°C and washed using Perm/Wash Buffer (BD Biosciences). Staining with

389   primary antibodies was performed overnight at 4°C followed by incubation with suitable secondary

390   antibodies diluted in Perm/Wash Buffer (Supplementary Table 4). The samples were filtered through

391   a 70 μm cell strainer and sorted on a FACSAria III (BD Biosciences) using a 100 μm nozzle. Isotype

16

392   and transduction controls stained with each antibody alone were analyzed alongside the samples.

393   INS^{LOW} and INS^{HIGH} samples were gated based on live and single cells with lower INS levels

394   compared to EV control cells (INS^{LOW}) and the highest 10 % of INS-expressing cells (INS^{HIGH}).

395   Sorted cell samples were stored at -80 °C until DNA extraction. Flow cytometry data was analyzed

396   using Flowjo 10.6 (BD Biosciences).

Preparation of genomic DNA for next-generation sequencing

398   DNA was extracted from frozen FACS sorted samples using QIAamp Blood Maxi/Mini Kit (Qiagen)

399   and processed for sequencing in a two-step PCR approach. Integrated sgRNAs were amplified using

400   Q5 polymerase (NEB) with 2.5 μg DNA input per reaction and an optimized number of cycles per

401   sample according to manufacturer's instructions. Specific Illumina TruSeq adapters were attached to

402   each sample using Q5 polymerase (NEB) with an optimized number of cycles (Supplementary Table

403   4). PCR products were run on a 2 % gel and purified using QIAquick Gel extraction kit (Qiagen).

404   Each Illumina library sample was qPCR-based quantified using the KAPA Library Quantification Kit

405   (Roche) prior to pooling and multiplexed sequencing on a NEXTSeq500 (Illumina) as 75 bp reads

406   with standard Illumina sequencing primer and PhiX to approximately 20% spike-in.

Analysis of pooled CRISPR screen

408   Raw fastq sequencing read files were merged for each sample using the "cat" command. To identify

409   enriched and depleted sgRNAs in this CRISPR screen, we used the previously described MAGeCK

410   (v0.5.9.2) algorithm[25]. Briefly, the "count" command was used to extract and count sgRNA sequences

411   from the raw fastq files based on the TKOv3 library input yielding a mean read count of 981 aligned

412   reads per sgRNA (Supplementary Fig S3). sgRNAs with low read counts of less than 10 and sgRNAs

413   mapping to genes that were not expressed in EndoC-βH1 were excluded. The "test" command was

414   used with the paired module to assess and analyse sgRNA enrichment or depletion between INS^{LOW}

415   and INS^{HIGH} populations. The analysis median normalizes read counts across samples to account for

416   varying read count depth and applies a mean-variance model to identify significant sgRNA

417   enrichment or depletion. MAGeCK multiple-testing adjusted sgRNA-level enrichment scores were

418   the basis for gene-level hit selection. We applied additional stringent criteria to only prioritise hits

17

419   with highly consistent effects across both replicates requiring genes to have an FDR<0.1 for at least

420   two out of their four sgRNAs in both independent replicates. CRISPR screen analysis was performed

421   in Python 3.8 and R 3.5.

422   Gene set enrichment and pathway analysis

423   Enriched GO and KEGG pathways within the screening hits were determined using the Database for

424   Annotation, Visualization and Integrated Discovery (DAVID) with homo sapiens as list and

425   background input[37]. Protein connectivity networks based on physical and functional interactions were

426   identified among screening hits using STRING v11[39]. Only interactions with a high confidence score

427   of ≥0.9 were selected.

428   Gene silencing in EndoC-βH1

429   Forward siRNA-based silencing in EndoC-βH1 was performed at 24 h after plating using ON-

430   TARGET plus SMARTpool siRNAs for *CALCOCO2, QSER1, PLCB3* or a non-targeting control

431   (Dharmacon) at 15 nM final concentration. siRNAs were preincubated with 0.4% Lipofectamine

432   RNAiMAX (Invitrogen) in Opti-MEM reduced serum-free medium (Gibco) for 15 min at room

433   temperature to form transfection complexes before dropwise addition to the culture. Cells were

434   assayed or harvested 72 h post transfection.

435   CALCOCO2 silencing in primary human islets

436   Lentiviral constructs coding for shRNAs targeting human *CALCOCO2* were obtained from

437   Dharmacon and virus was produced as described before. Primary islets were dispersed into single

438   cells by enzymatic digestion (Accumax, Invitrogen) and transduced with 1×10^9 viral units/1 mL

439   lentivirus. Transduced islet cells were cultured in ultra-low attachment well plates for five days prior

440   to further analysis.

441   Gene expression analysis

442   RNA for gene expression analysis from EndoC-βH1 was extracted using TRIzol reagent (Invitrogen)

443   and synthesized into complementary DNA using the GoScript Reverse Transcriptase System

444   (Promega), both according to manufacturer's instructions. RNA for gene expression analysis from

445   primary human pseudoislets was extracted using the PicoPure RNA isolation kit (Life Technologies)

446   and cDNA was synthesized using the Maxima first strand cDNA synthesis kit (Thermo Scientific),

447   according to manufacturer's instructions. Quantitative qPCR (qPCR) was performed using TaqMan

448   real-time PCR assays on a 7900HT Fast Real-Time PCR System (all Applied Biosystems). Ct-values

449   were analysed using the ΔΔCt method and target genes were normalized to the combined average of

450   the housekeeping genes peptidylprolyl isomerase A (PPIA), glycerinaldehyd-3-phosphat-

451   dehydrogenase (GAPDH) and TATA-box binding protein (TBP). *CALCOCO2* expression in EndoC-

452   βH1 and primary islets was extracted from previously published and analyzed RNA sequencing

453   data[69,90].

454   SYBR Green based quantitative PCR

455   sgRNA integrations in the small-scale screen were quantified using SYBR Green based quantitative

456   PCR and primer targeting the respective sgRNA sequences (Supplementary Table 4). DNA from

457   FACS sorted EndoC-βH1 was extracted as described. Each sample was prepared using 20 ng of total

458   DNA and SYBR Green PCR Master mix (Bio-Rad), following manufacturer's instructions. The

459   samples were amplified and analysed as described for gene expression experiments.

460   Insulin secretion assay in EndoC-βH1

461   Silenced EndoC-βH1 were starved overnight at 48 h post transfection in culture medium containing

462   2.8 mM glucose followed by 30 min incubation in 0 mM glucose the next morning. Insulin secretion

463   was initiated through static incubations in the indicated glucose or secretagogues conditions for 1 h.

464   Insulin containing supernatant was harvested and adherent cells were lysed in ice-cold acid ethanol to

465   release intracellular insulin. Insulin was quantified using the Insulin (human) AlphaLISA Detection

466   Kit and the EnSpire Alpha Plate Reader (both Perkin Elmer) based on 1:10 and 1:200 dilutions for

467   supernatant and insulin content, respectively. Secreted insulin was normalized to the level of

468   intracellular insulin content or cell count, which was measured prior to cell lysis using the CyQUANT

469   Direct Cell Proliferation assay (Invitrogen).

470   In vitro insulin and glucagon secretion assays in primary pseudoislets

471   Batches of 25 pseudoislets were used per donor for *in vitro* secretion assays as previously described

472   [91]. For insulin secretion assays, pseudoislets were incubated at 2.8, 5.6, 16.7 and 16.7 mM + IBMX

473   glucose concentrations for 60 min each and supernatants were collected. For glucagon secretion

474   assays, pseudoislets were incubated at 7, 1, 1 mM +10 mM Arginine glucose concentrations for 60

475   min each and supernatants were collected. Human insulin in the supernatants and pseudoislet lysates

476   were quantified using a human insulin ELISA kit (Mercodia). Human glucagon in the supernatants

477   and pseudoislet lysates were quantified using a human glucagon ELISA kit (Mercodia). Secreted

478   insulin and glucagon levels are presented as a percentage of gDNA, quantified from the pseudoislet

479   lysates.

480   Western Blot analysis

481   Whole-cell protein extracts were obtained from cell pellets through lysis in RIPA buffer (50 mM Tris

482   pH 7.4, 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS) containing 1x

483   protease inhibitor cocktail (Roche). Protein concentrations were quantified using the DC protein assay

484   (Bio-Rad), 10 μg of total protein were mixed with sample buffer (4x Laemmli Buffer (Bio-Rad), 10%

485   β-mercaptoethanol (Sigma-Aldrich)) and boiled for 10 min at 80°C. Denatured samples were run on a

486   4-20% Criterion TGX Stain-Free Precast Gel for 15 min at 300 V in Tris–glycine–SDS buffer and

487   transferred to a 0.2 μm polyvinylidene difluoride (PVDF) using the Trans-Blot Turbo Transfer

488   System (all Bio-Rad). Unspecific antibody binding was blocked through incubation with 3% BSA in

489   Tris-buffered saline containing 0.1% Tween-20 (TBST) for 1 h at room temperature. Primary

490   antibody incubations were performed at 4°C overnight followed by incubation with HRP-conjugated

491   secondary antibodies for 1 h at room temperature (Supplementary Table 4). Blots were imaged using

492   the ChemiDoc MP Imaging System (Bio-Rad) and reprobed as described using loading control

493   antibodies of appropriate size. Image Lab 6.0 software (Bio-Rad) was used to quantify protein bands

494   and proteins of interest were normalized to their respective loading control on the same blot.

495 Immunofluorescence analysis

496 Human pancreatic tissue was obtained from subjects in post-mortem examinations in line with local

497 and national ethics permission. Pancreata were fixed in 4% paraformaldehyde overnight,

498 cryoembedded and sections of 4 μm were prepared. Tissue sections were dipped in distilled water,

499 boiled for 20 min in target retrieval solution (Dako), permeabilized for 10 min at room temperature

500 using 1% PBS-Triton X-100 (Sigma-Aldrich) and unspecific antibody binding was blocked using 1%

501 BSA (Roche), 0.2% non-fat milk, 0.5% Triton-X 100, 1% DMSO in PBS (all Sigma-Aldrich).

502 Primary antibody incubations were performed overnight at 4°C followed by incubation with

503 secondary Alexa Fluor-conjugated antibodies at room temperature for 1 h (Supplementary Table 4).

504 The slides were mounted using VectaShield mounting media containing DAPI (Vector Laboratories)

505 and imaged on a Zeiss AxioM1 (Zeiss) microscope using an x20 and x40 objective.

506 Statistics

507 Statistical analyses were performed in Prism 8.1 (GraphPad Software). The number of biological

508 independent replicates are shown as individual data points and the error bars represent the standard

509 error of the mean. If appropriate, fold changes were plotted but statistical analysis was performed on

510 log-transformed values. Normally distributed variables were compared between two or more groups

511 using two-sample Student's t-test or one-way analysis of variance (ANOVA) followed by Sidak's

512 multiple comparison test, respectively. Samples that were normalized to their respective control

513 within a single replicate were analyzed using a one-sample Student's t-test.

514

515 **Data availability**

516 Fastq sequencing files from the CRISPR screen have been deposited in the European Nucleotide

517 Archive (ENA) at EMBL-EBI. The accession number will be made available upon publication.

Acknowledgments

Author contributions

A.K.G., S.K.T. and A.L.G. conceived the study. A.K.G., E.N-G. and S.K.T. performed CRISPR screening optimization experiments. A.K.G., E.N-G., R.B. and S.N. performed the CRISPR screen. E.N-G. and R.B performed sequencing library preparation. A.K.G. analyzed the CRISPR screening data. A.W-A contributed to the CRISPR screening analysis. A.K.G. and A.L.G. performed data integration analysis. A.K.G. performed the characterization of *CALCOCO2, QSER1* and *PLCB3* in beta cells. A.K.G. and R.J.B performed islet immunostainings. R.J.B., A.K.G. and V.R. performed islet shRNA knockdown experiments. S.K.K, D.E. and A.L.G. supervised the research. A.K.G. and A.L.G. wrote the manuscript. All authors approved the final draft of the manuscript.

Competing Interest

A.K.G. is now an employee of AstraZeneca, S.K.T. is now an employee of Vertex Pharmaceuticals and A.W-A. is now an employee of Genomics plc. All experimental work by A.K.G., S.K.T., and A.W-A. was carried out under employment at the University of Oxford. A.K.G. holds stocks in Astra

544   Zeneca. A.L.G.'s spouse holds stock options in Roche. The other authors declare no competing

545   interests.

546

References

1. Mahajan, A. *et al.* Trans-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *medRxiv* 2020.09.22.20198937 (2020). doi:10.1101/2020.09.22.20198937

2. Krentz, N. A. J. & Gloyn, A. L. Insights into pancreatic islet cell dysfunction from type 2 diabetes mellitus genetics. *Nat. Rev. Endocrinol.* 1–11 (2020). doi:10.1038/s41574-020-0325-0

3. Fulco, C. P. *et al.* Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).

4. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* (2021). doi:10.1038/s41586-021-03446-x

5. Claussnitzer, M. *et al.* Leveraging Cross-Species Transcription Factor Binding Site Patterns : From Diabetes Risk Loci to Disease Mechanisms. *Cell* **156**, 343–358 (2014).

6. Morris, J. A. *et al.* Discovery of target genes and pathways of blood trait loci using pooled CRISPR screens and single cell RNA sequencing. *bioRxiv* 2021.04.07.438882 (2021). doi:10.1101/2021.04.07.438882

7. Dimas, A. S. *et al.* Impact of Type 2 Diabetes Susceptibility Variants on Quantitative Glycemic Traits Reveals Mechanistic Heterogeneity. *Diabetes* **63**, 2158–2171 (2014).

8. Thurner, M. *et al.* Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. *Elife* **7**, e31977 (2018).

9. Mahajan, A. *et al.* Refining the accuracy of validated target identification through coding variant fine- mapping in type 2 diabetes. *Nat. Genet.* **50**, 559–571 (2018).

10. Udler, M. S. *et al.* Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. *PLoS Med.* **15**, e1002654 (2018).

11. Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143 (2014).

12. Benner, C. *et al.* The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression. *BMC Genomics* **15**, 620 (2014).

13. Fiaschi-Taesch, N. M. *et al.* Human Pancreatic β-Cell G1/S Molecule Cell Cycle Atlas. *Diabetes* **62**, 2450–2459 (2013).

14. Henquin, J.-C., Dufrane, D., Kerr-Conte, J. & Nenquin, M. Dynamics of glucose-induced

578   insulin secretion in normal human islets. *Am. J. Physiol. Metab.* **309**, E640–E650 (2015).

579   15.   De Vos, A. *et al.* Human and rat beta cells differ in glucose transporter but not in glucokinase

580   gene expression. *J. Clin. Invest.* **96**, 2489–2495 (1995).

581   16.   Braun, M. *et al.* Voltage-Gated Ion Channels in Human Pancreatic Beta-Cells:

582   Electrophysiological Characterization and Role in Insulin Secretion. *Diabetes* **57**, 1618–28

583   (2008).

584   17.   Rorsman, P. & Ashcroft, F. M. Pancreatic β-Cell Electrical Activity and Insulin Secretion: Of

585   Mice and Men. *Physiol. Rev.* **98**, 117–214 (2018).

586   18.   Varshney, A. *et al.* Genetic regulatory signatures underlying islet gene expression and type 2

587   diabetes. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2301–2306 (2017).

588   19.   Viñuela, A. *et al.* Genetic variant effects on gene expression in human pancreatic islets and

589   their implications for T2D. *Nat. Commun.* **11**, 4912 (2020).

590   20.   Grotz, A. K. *et al.* A CRISPR/Cas9 genome editing pipeline in the EndoC-βH1 cell line to

591   study genes implicated in beta cell function. *Wellcome Open Res.* **4**, 150 (2019).

592   21.   Hastoy, B. *et al.* Electrophysiological properties of human beta-cell lines EndoC-βH1 and -

593   βH2 conform with human beta-cells. *Sci. Rep.* **8**, 1–16 (2018).

594   22.   Lawlor, N. *et al.* Multiomic Profiling Identifies cis-Regulatory Networks Underlying Human

595   Pancreatic β Cell Identity and Function. *Cell Rep.* **26**, 788-801.e6 (2019).

596   23.   Thomsen, S. K. *et al.* Type 2 diabetes risk alleles in PAM impact insulin release from human

597   pancreatic β-cells. *Nat. Genet.* **50**, 1122–1131 (2018).

598   24.   Hart, T. *et al.* Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens.

599   *Genes Genomes Genet.* **7**, 2719–2727 (2017).

600   25.   Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale

601   CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).

602   26.   Sussel, L. *et al.* Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due

603   to arrested differentiation of pancreatic beta cells. *Development* **125**, 2213–2221 (1998).

604   27.   Doyle, M. J. & Sussel, L. Nkx2.2 regulates beta-cell function in the mature islet. *Diabetes* **56**,

605   1999–2007 (2007).

606   28.   Sansbury, F. H. *et al.* SLC2A2 mutations can cause neonatal diabetes, suggesting GLUT2 may

607   have a role in human insulin secretion. *Diabetologia* **55**, 2381–2385 (2012).

25

608   29.   Abdulkarim, B. *et al.* A Missense Mutation in PPP1R15B Causes a Syndrome Including
609   Diabetes, Short Stature, and Microcephaly. *Diabetes* **64**, 3951–3962 (2015).

610   30.   Modi, H. *et al.* Autocrine Action of IGF2 Regulates Adult β-Cell Mass and Function. *Diabetes*
611   **64**, 4148–4157 (2015).

612   31.   Casellas, A. *et al.* Insulin-like growth factor 2 overexpression induces β-Cell dysfunction and
613   increases beta-cell susceptibility to damage. *J. Biol. Chem.* **290**, 16772–16785 (2015).

614   32.   Mercader, J. M. *et al.* A Loss-of-Function Splice Acceptor Variant in IGF2 Is Protective for
615   Type 2 Diabetes. *Diabetes* **66**, 2903–2914 (2017).

616   33.   Flannick, J. *et al.* Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls.
617   *Nature* **570**, 71–76 (2019).

618   34.   Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576.e16 (2017).

619   35.   Behan, F. M. *et al.* Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens.
620   *Nature* **568**, 511–516 (2019).

621   36.   Dempster, J. M. *et al.* Agreement between two large pan-cancer CRISPR-Cas9 gene
622   dependency data sets. *Nat. Commun.* **10**, 5817 (2019).

623   37.   Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large
624   gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

625   38.   Ogata, H. *et al.* KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27**,
626   29–34 (1999).

627   39.   Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased
628   coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic*
629   *Acids Res.* **47**, D607–D613 (2019).

630   40.   Wang, L. *et al.* Dichotomous role of pancreatic HUWE1/MULE/ARF-BP1 in modulating beta
631   cell apoptosis in mice under physiological and genotoxic conditions. *Diabetologia* **57**, 1889–
632   1898 (2014).

633   41.   Suriben, R. *et al.* β-Cell Insulin Secretion Requires the Ubiquitin Ligase COP1. *Cell* **163**,
634   1457–1467 (2015).

635   42.   Tsuda, N. *et al.* A novel free fatty acid receptor 1 (GPR40/FFAR1) agonist, MR1704,
636   enhances glucose-dependent insulin secretion and improves glucose homeostasis in rats.
637   *Pharmacol. Res. Perspect.* **5**, e00340 (2017).

638   43.   Veluthakal, R., Madathilparambil, S. V., McDonald, P., Olson, L. K. & Kowluru, A.

639 Regulatory roles for Tiam1, a guanine nucleotide exchange factor for Rac1, in glucose-
640 stimulated insulin secretion in pancreatic beta-cells. *Biochem. Pharmacol.* **77**, 101–113 (2009).

641 44. Taguchi, Y. *et al.* Sphingosine-1-phosphate Phosphatase 2 Regulates Pancreatic Islet β-Cell
642 Endoplasmic Reticulum Stress and Proliferation. *J. Biol. Chem.* **291**, 12029–12038 (2016).

643 45. Longuet, C. *et al.* Extracellularly Regulated Kinases 1/2 (p44/42 Mitogen-Activated Protein
644 Kinases) Phosphorylate Synapsin I and Regulate Insulin Secretion in the MIN6 β-Cell Line
645 and Islets of Langerhans. *Endocrinology* **146**, 643–654 (2005).

646 46. Hennige, A. M. *et al.* Upregulation of insulin receptor substrate-2 in pancreatic β cells prevents
647 diabetes. *J. Clin. Invest.* **112**, 1521–1532 (2003).

648 47. Cantley, J. *et al.* Pancreatic deletion of insulin receptor substrate 2 reduces beta and alpha cell
649 mass and impairs glucose homeostasis in mice. *Diabetologia* **50**, 1248–1256 (2007).

650 48. Fernandez-Ruiz, R., Vieira, E., Garcia-Roves, P. M. & Gomis, R. Protein Tyrosine
651 Phosphatase-1B Modulates Pancreatic β-cell Mass. *PLoS One* **9**, e90344 (2014).

652 49. Bouzakri, K. *et al.* Rab GTPase-Activating Protein AS160 Is a Major Downstream Effector of
653 Protein Kinase B/Akt Signaling in Pancreatic β-Cells. *Diabetes* **57**, 1195 LP – 1204 (2008).

654 50. Sakamoto, K. & Holman, G. D. Emerging role for AS160/TBC1D4 and TBC1D1 in the
655 regulation of GLUT4 traffic. *Am. J. Physiol. Endocrinol. Metab.* **295**, E29-37 (2008).

656 51. Dash, S. *et al.* A truncation mutation in TBC1D4 in a family with acanthosis nigricans and
657 postprandial hyperinsulinemia. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9350–9355 (2009).

658 52. Dash, S. *et al.* Analysis of TBC1D4 in patients with severe insulin resistance. *Diabetologia* **53**,
659 1239–1242 (2010).

660 53. Moltke, I. *et al.* A common Greenlandic TBC1D4 variant confers muscle insulin resistance
661 and type 2 diabetes. *Nature* **512**, 190–193 (2014).

662 54. Ndiaye, F. K. *et al.* Expression and functional assessment of candidate type 2 diabetes
663 susceptibility genes identify four new genes contributing to human insulin secretion. *Mol.*
664 *Metab.* **6**, 459–470 (2017).

665 55. Carrat, G. R. *et al.* Decreased STARD10 Expression Is Associated with Defective Insulin
666 Secretion in Humans and Mice. *Am J Hum Genet* **100**, 238–256 (2017).

667 56. Hu, M. *et al.* Chromatin 3D interaction analysis of the STARD10 locus unveils FCHSD2 as a
668 new regulator of insulin secretion. *bioRxiv* 2020.03.31.017707 (2020).
669 doi:10.1101/2020.03.31.017707

670　57.　Kulzer, J. R. *et al.* A Common Functional Regulatory Variant at a Type 2 Diabetes Locus
671　　　Upregulates ARAP1 Expression in the Pancreatic Beta Cell. *Am J Hum Genet* **94**, 186–197
672　　　(2014).

673　58.　Forgetta, V. *et al.* An Effector Index to Predict Causal Genes at GWAS Loci. *bioRxiv*
674　　　2020.06.28.171561 (2020). doi:10.1101/2020.06.28.171561

675　59.　Type 2 Diabetes Knowledge Portal. Curated T2D effector gene predictions. Available at:
676　　　https://t2d.hugeamp.org/effectorgenes.html. (Accessed: 10th January 2021)

677　60.　Type 2 Diabetes Knowledge Portal. Integrated Classifier Predictions. Available at:
678　　　https://t2d.hugeamp.org/effectorgenes.html. (Accessed: 15th March 2021)

679　61.　Accelerating Medicines Partnership. Type 2 Diabetes Knowledge Portal. (2020). Available at:
680　　　http://www.type2diabetesgenetics.org/. (Accessed: 4th April 2020)

681　62.　Greenwald, W. W. *et al.* Pancreatic islet chromatin accessibility and conformation reveals
682　　　distal enhancer networks of type 2 diabetes risk. *Nat. Commun.* **10**, 2078 (2019).

683　63.　van de Bunt, M. *et al.* Transcript Expression Data from Human Islets Links Regulatory Signals
684　　　from Genome- Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their
685　　　Downstream Effectors. *PLoS Genet.* **11**, 1–21 (2015).

686　64.　Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R. & Soranzo, N. New genetic loci
687　　　implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.*
688　　　**42**, 408 (2010).

689　65.　Ingelsson, E. *et al.* Detailed Physiologic Characterization Reveals Diverse Mechanisms for
690　　　Novel Genetic Loci Regulating Glucose and Insulin Metabolism in Humans. *Diabetes* **59**,
691　　　(2010).

692　66.　Xie, X. *et al.* Molecular basis of ubiquitin recognition by the autophagy receptor CALCOCO2.
693　　　*Autophagy* **11**, 1775–1789 (2015).

694　67.　Boyle, K. B., Ravenhill, B. J. & Randow, F. CALCOCO2/NDP52 initiates selective autophagy
695　　　through recruitment of ULK and TBK1 kinase complexes. *Autophagy* **15**, 1655–1656 (2019).

696　68.　Furuya, N. *et al.* NDP52 interacts with mitochondrial RNA poly(A) polymerase to promote
697　　　mitophagy. *EMBO Rep.* **19**, e46363 (2018).

698　69.　Thomsen, S. K. *et al.* Systematic Functional Characterization of Candidate Causal Genes for
699　　　Type 2 Diabetes Risk Variants. *Diabetes* **65**, 3805–11 (2016).

700　70.　Mahajan, A. *et al.* Fine-mapping of an expanded set of type 2 diabetes loci to single-variant

701 resolution using high-density imputation and islet- specific epigenome maps. *Nat. Genet.* **50**,

702 1505–1513 (2018).

703 71. Thore, S., Dyachok, O., Gylfe, E. & Tengholm, A. Feedback activation of phospholipase C via

704 intracellular mobilization and store-operated influx of Ca2+ in insulin-secreting β-cells. *J. Cell*

705 *Sci.* **118**, 4463 LP – 4471 (2005).

706 72. Thore, S., Dyachok, O. & Tengholm, A. Oscillations of phospholipase C activity triggered by

707 depolarization and Ca2+ influx in insulin-secreting cells. *J. Biol. Chem.* **279**, 19396–19400

708 (2004).

709 73. Dixon, G. *et al.* QSER1 protects DNA methylation valleys from de novo methylation. *Science*

710 *(80-.).* **372**, eabd0875 (2021).

711 74. Dwivedi, O. P. *et al.* Loss of ZnT8 function protects against diabetes by enhanced insulin

712 secretion. *Nat. Genet.* **51**, 1596–1606 (2019).

713 75. Fadista, J. *et al.* Global genomic and transcriptomic analysis of human pancreatic islets reveals

714 novel genes influencing glucose metabolism. *Proc Natl Acad Sci U S A* **111**, 13924–13929

715 (2014).

716 76. Miguel-Escalada, I. *et al.* Human pancreatic islet three-dimensional chromatin architecture

717 provides insights into the genetics of type 2 diabetes. *Nat Genet* **51**, 1137–1148 (2019).

718 77. Rai, V. *et al.* Single-cell ATAC-Seq in human pancreatic islets and deep learning upscaling of

719 rare cells reveals cell-specific type 2 diabetes regulatory signatures. *Mol. Metab.* **32**, 109–121

720 (2020).

721 78. Gaulton, K. J. *et al.* A map of open chromatin in human pancreatic islets. *Nat. Genet.* **42**, 255–

722 259 (2010).

723 79. Bonàs-Guarch, S. *et al.* Re-analysis of public genetic data reveals a rare X-chromosomal

724 variant associated with type 2 diabetes. *Nat. Commun.* **9**, 321 (2018).

725 80. Lazarou, M. *et al.* The ubiquitin kinase PINK1 recruits autophagy receptors to induce

726 mitophagy. *Nature* **524**, 309–314 (2015).

727 81. Fex, M. *et al.* The pathogenetic role of β-cell mitochondria in type 2 diabetes. *J. Endocrinol.*

728 **236**, R145–R159 (2018).

729 82. Soleimanpour, S. A. *et al.* The diabetes susceptibility gene Clec16a regulates mitophagy. *Cell*

730 **157**, 1577–1590 (2014).

731 83. Pearson, G. *et al.* Clec16a, Nrdp1, and USP8 Form a Ubiquitin-Dependent Tripartite Complex

732   That Regulates β-Cell Mitophagy. *Diabetes* **67**, 265 LP – 277 (2018).

733   84.   Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease.
734   *Science (80-.).* **361**, 1–61 (2018).

735   85.   Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma
736   proteome on complex diseases. *Nat. Genet.* **52**, 1122–1131 (2020).

737   86.   Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**,
738   19–34 (2017).

739   87.   Fang, Z. *et al.* Single-Cell Heterogeneity Analysis and CRISPR Screen Identify Key β-Cell-
740   Specific Disease Genes. *Cell Rep.* **26**, 3132-3144.e7 (2019).

741   88.   Ravassard, P. *et al.* A genetically engineered human pancreatic β cell line exhibiting glucose-
742   inducible insulin secretion. *J. Clin. Invest.* **121**, 3589–3597 (2011).

743   89.   Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for
744   CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).

745   90.   Nica, A. C. *et al.* Cell-type, allelic, and genetic signatures in the human pancreatic beta cell
746   transcriptome. *Genome Res.* **23**, 1554–1562 (2013).

747   91.   Peiris, H. *et al.* Discovering human diabetes-risk gene function with genetics and physiological
748   assays. *Nat. Commun.* **9**, 1–11 (2018).
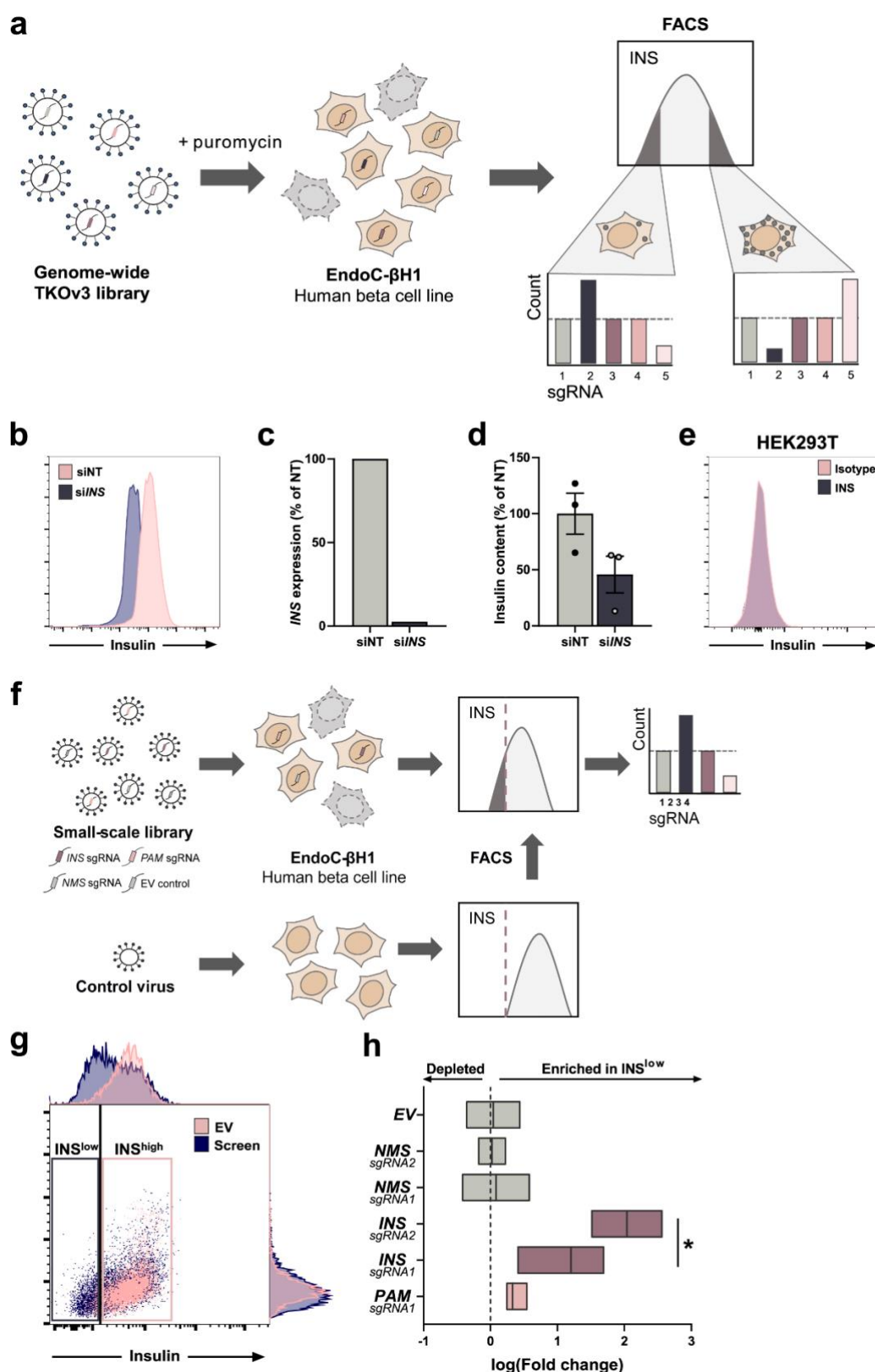
749

750

Figures

Fig 1. Optimization of a CRISPR screen for insulin content in EndoC-βH1. a) Pipeline for a genome-wide CRISPR loss-of-function screen in EndoC-βH1 from viral transduction (left) to

755   antibiotic selection (middle) and a final FACS selection followed by sequencing and enrichment

756   analysis of integrated sgRNAs (right). b) FACS staining for intracellular insulin in *INS* silenced

757   EndoC-βH1 (si*INS*, blue) or their respective non-targeting controls (siNT, pink), c) the associated

758   mRNA expression of *INS* and d) corresponding levels of intracellular insulin as measured by

759   alphaLISA. e) FACS staining using the insulin antibody (blue) or isotype control (pink) in the human

760   embryonic kidney cell line HEK293T. f) Pipeline for a small-scale CRISPR screen in EndoC-βH1

761   targeting only three genes *(NMS, INS* and *PAM)* alongside EV control cells. FACS sorting gates for

762   low insulin (dashed line) were determined based on control cells transduced in parallel. sgRNA

763   enrichment was assessed in INS^{low} compared to INS^{high}. g) FACS staining in EndoC-βH1 comparing

764   insulin staining of EV control cells (pink) and cells from the small-scale CRISPR screen (blue).

765   Respective boxes indicate sorting gates INS^{low} and INS^{high}. h) sgRNA enrichment for all individual

766   sgRNAs in INS^{low} compared to INS^{high} from control (grey) samples and positive controls (*INS* and

767   *PAM*, purple and pink). All data are presented as mean ± SEM from two (g,h) or three independent

768   experiments and representative FACS plots are shown with their respective silencing efficiency. Data

769   were analyzed using one (h) or two-sample t-test (d). NT, non-targeting. LFC, log2(fold change); EV,

770   empty vector. P-value * < 0.05.
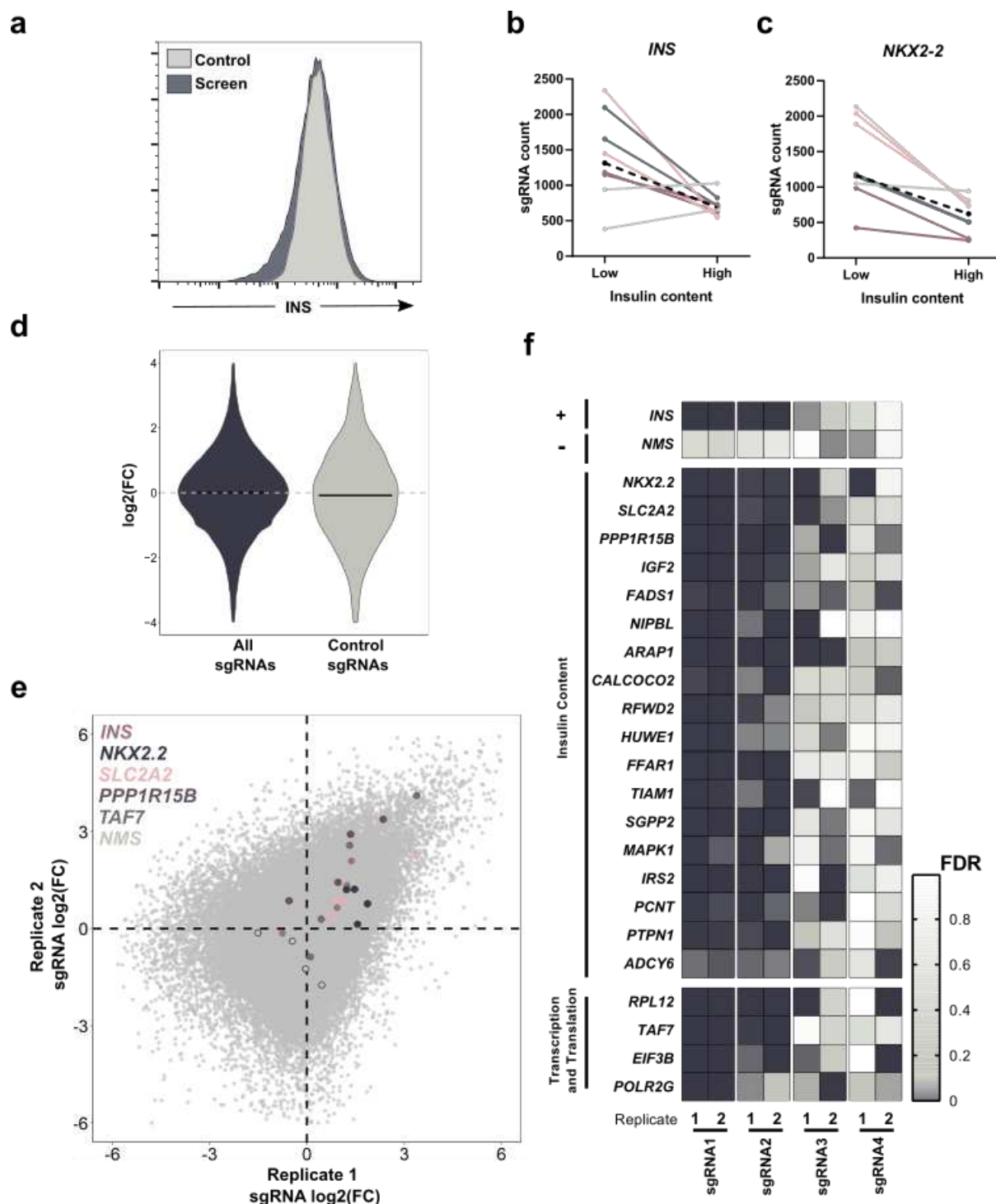
771

772

773

774

Fig 2. A CRISPR screen for insulin content in EndoC-βH1. a) Flow cytometry histograms showing insulin content in EndoC-βH1 transduced with the CRISPR screening library (blue) compared to cells transduced with EV virus (grey). b,c) Changes in sgRNA count from low to high insulin content screening sample with each colour representing the same sgRNA across the two screen replicates for (b) *INS* and (c) *NKX2-2*. The black dashed line represents the median sgRNA count for this gene. d) sgRNA distribution of log2(FC) for all sgRNAs (blue) compared to control sgRNAs targeting LacZ,

33

782   EGFP and luciferase (grey). The black line indicates the median values for each group. e) sgRNA

783   distribution of log2(FC) within each screening replicate with specific sgRNAs of interest (4 per gene)

784   highlighted in individual colours per gene while overall sgRNA distribution is depicted in light grey.

785   f) Individual genes of interest with their respective sgRNAs per replicate, the colour ranges from not

786   significant (white) to highly significant (dark blue) based on the FDR and representing a significant

787   sgRNA enrichment or depletion between INS^{low} and INS^{high}. *INS* (+) and *NMS* (-) on the top panel

788   represent positive and negative control genes, respectively. Data are from two independent genome-

789   wide CRISPR screen replicates, screen FDR values were determined using MAGeCK (e,f) or through

790   a two-sample t-test (d). Log2(FC), log2(fold change), EV, empty vector.
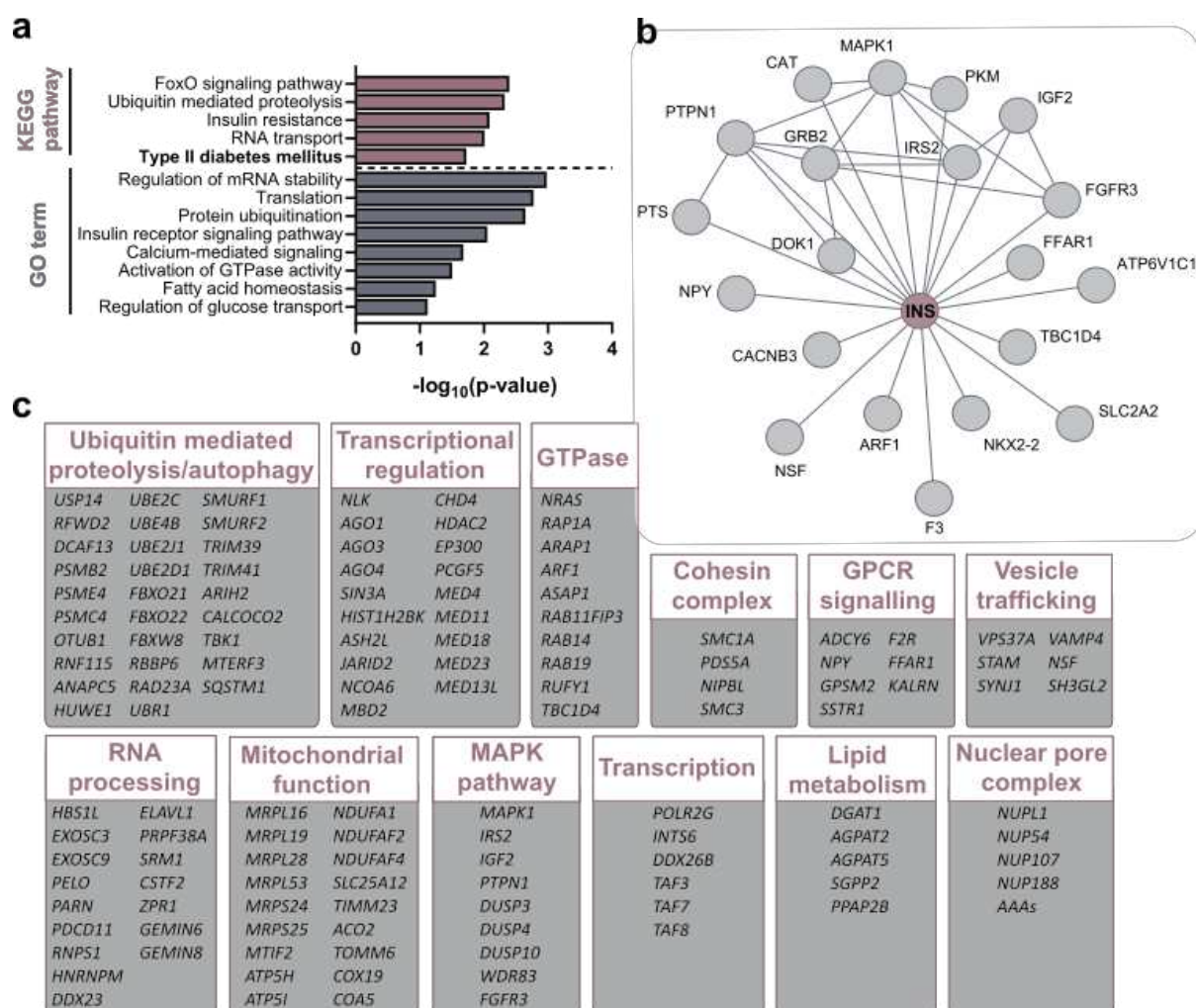
791

792

793

Fig 3. Integration of CRISPR screening hits. a) Pathway enrichment analysis assessing KEGG pathways or GO terms for biological processes. Selected pathways are shown, ranked by p-value. b,c) STRING pathway analysis showing protein-protein associations including physical and functional interactions b) between INS and other screening hits and c) for screening hits clustered into functionally associated groups.
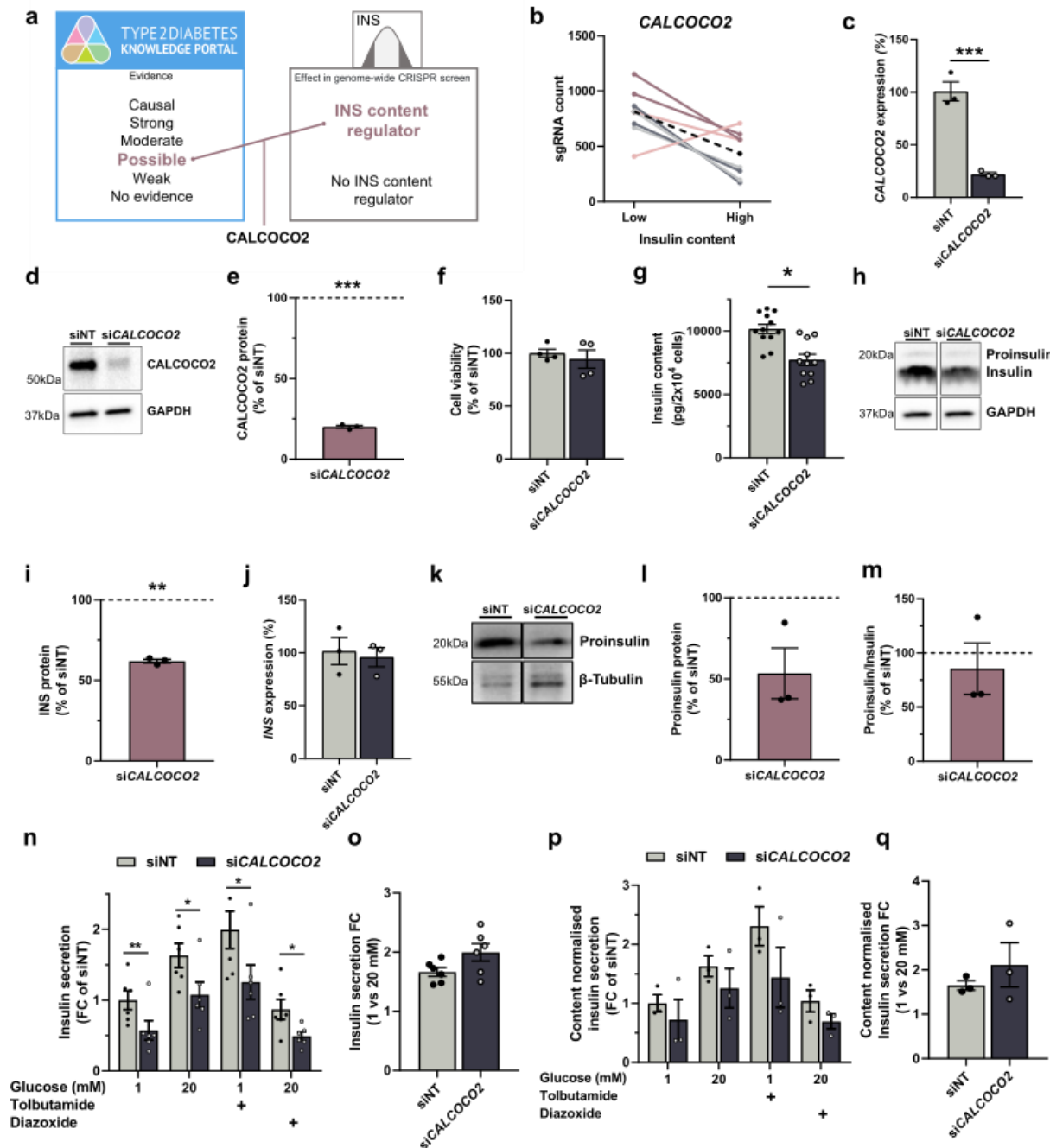
799

800

Fig 4. *CALCOCO2* knockdown reduces insulin content in EndoC-βH1. a) Gene prioritization approach combining genes with low evidence as T2D effector transcripts (as outlined in the integration approach) while being identified as a screening hit, highlighted *CALCOCO2*. b) Changes in sgRNA count from low to high insulin content sample with each color representing the same sgRNA across the two screen replicates. The black dashed line represents the median sgRNA count for this gene. c-m) All data are from si*CALCOCO2* treated EndoC-βH1 compared to non-targeting control cells. c) mRNA expression of *CALCOCO2*. d) Protein level of CALCOCO2 and its loading control GAPDH. e) Quantification of CALCOCO2 western blot data, normalized to GAPDH and siNT control cells. f) Cell count measurements. g) Insulin content in pg per 20 000 cells, measured by

810   alphaLISA. h) Protein level of insulin and its loading control GAPDH. The antibody also detects

811   proinsulin but only gives a weak signal compared to mature insulin. i) Quantification of insulin and

812   western blot data normalised to GAPDH and siNT control cells. j) mRNA expression of *INS*. k)

813   Protein level of proinsulin and its loading control β-Tubulin. l) Quantification of proinsulin western

814   blot data normalised to β-Tubulin and siNT control cells. m) Ratio of proinsulin to insulin for western

815   blot data. n,p) Insulin secretion normalised to siNT (n) or to insulin content and siNT (p) in 1 mM, 20

816   mM, 1 mM+100 μM tolbutamide or 20 mM glucose+100 μM diazoxide. o,q) Insulin secretion fold

817   change from 1 to 20 mM glucose normalised to siNT (o) or to insulin content and siNT (q). The

818   protein level is displayed as percentage of siNT which is highlighted as dotted line at 100%. All data

819   are mean ± SEM from at least three independent experiments. Data were analyzed using two-sample

820   t-test (c), one-sample t-test (e,i,l,m), one-way (f,g,j,o,q) or two-way (n,p) ANOVA with Sidak's

821   multiple comparison test. P-values * < 0.05, ** < 0.01, *** < 0.001. FC, fold change; NT, non-
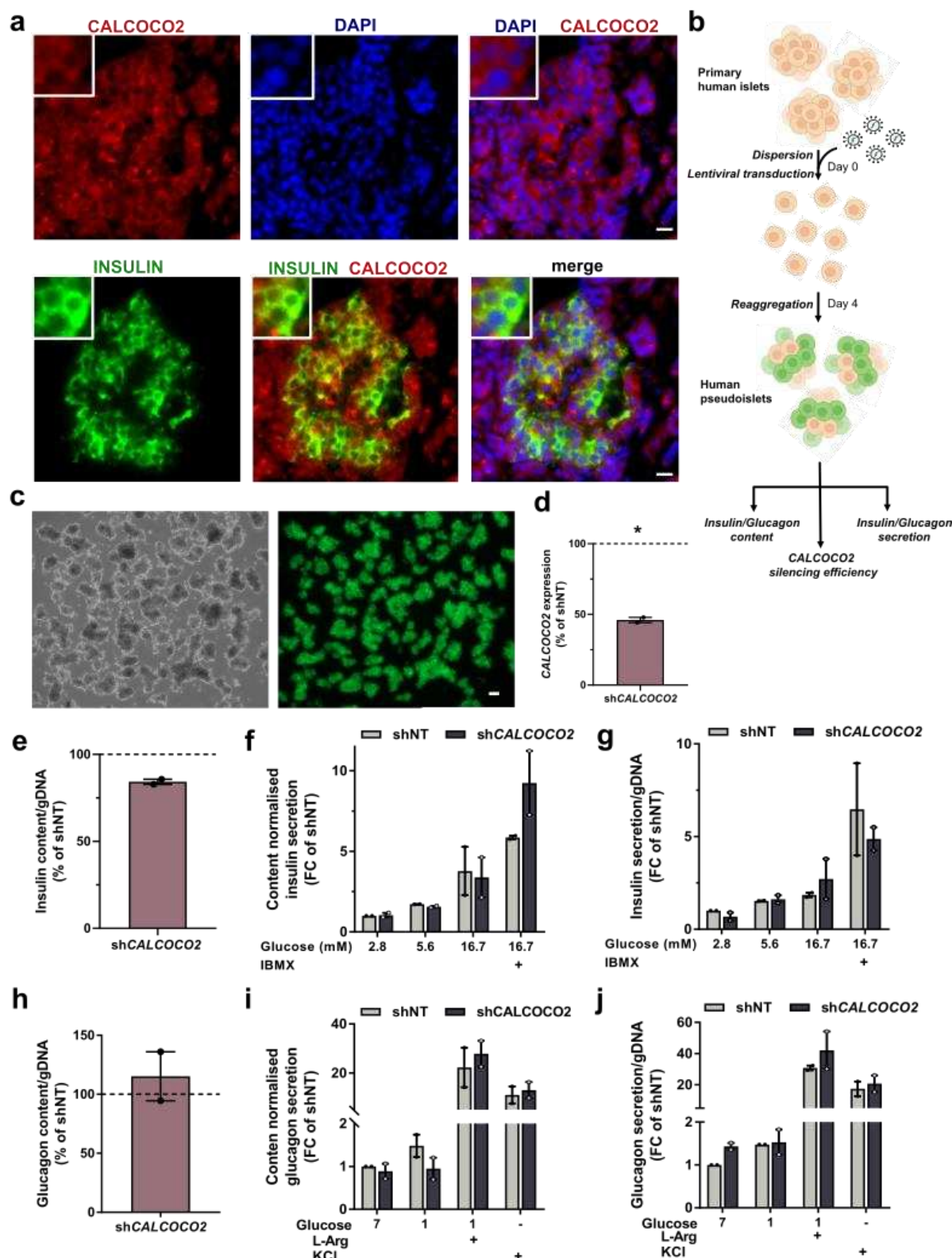
822   targeting.

823

824

825

Fig 5. *CALCOCO2* knockdown reduces insulin content in primary human pseudoislets. a) Immunofluorescence staining of primary human islets in pancreas sections. Sections were double stained for INS (green) and CALCOCO2 (red). Cell nuclei were counterstained with DAPI (blue). Scale bar is 10 μm. b-j) All data are from sh*CALCOCO2* pseudoislets compared to non-targeting

831 control pseudoislets (shNT). b) Pseudoislet transduction and formation from primary human islets. c)

832 sh*CALCOCO2* pseudoislets under bright field (left) and GFP positive pseudoislets (right). Scale bar is

833 100 μm. d) mRNA expression of *CALCOCO2*. e,h) Intracellular insulin (e) and glucagon (h) content.

834 f,g,i,j) Insulin (f,g) or glucagon (i,j) secretion normalised to content (f,i) or to genomic DNA (gDNA)

835 (g,j) in 2.8 mM, 5.6 mM, 16.7 mM or 16.7 mM glucose + IBMX for insulin secretion and 7 mM, 1

836 mM, 1 mM glucose + L-Arginine or KCl for glucagon secretion. All data are mean ± SEM from two

837 independent donor. Data were analysed using two-way ANOVA Sidak's multiple comparison test

838 (f,g,i,j) or one-sample t-test (d,e,h). P-values * < 0.05. FC, fold change; NT, non-targeting.
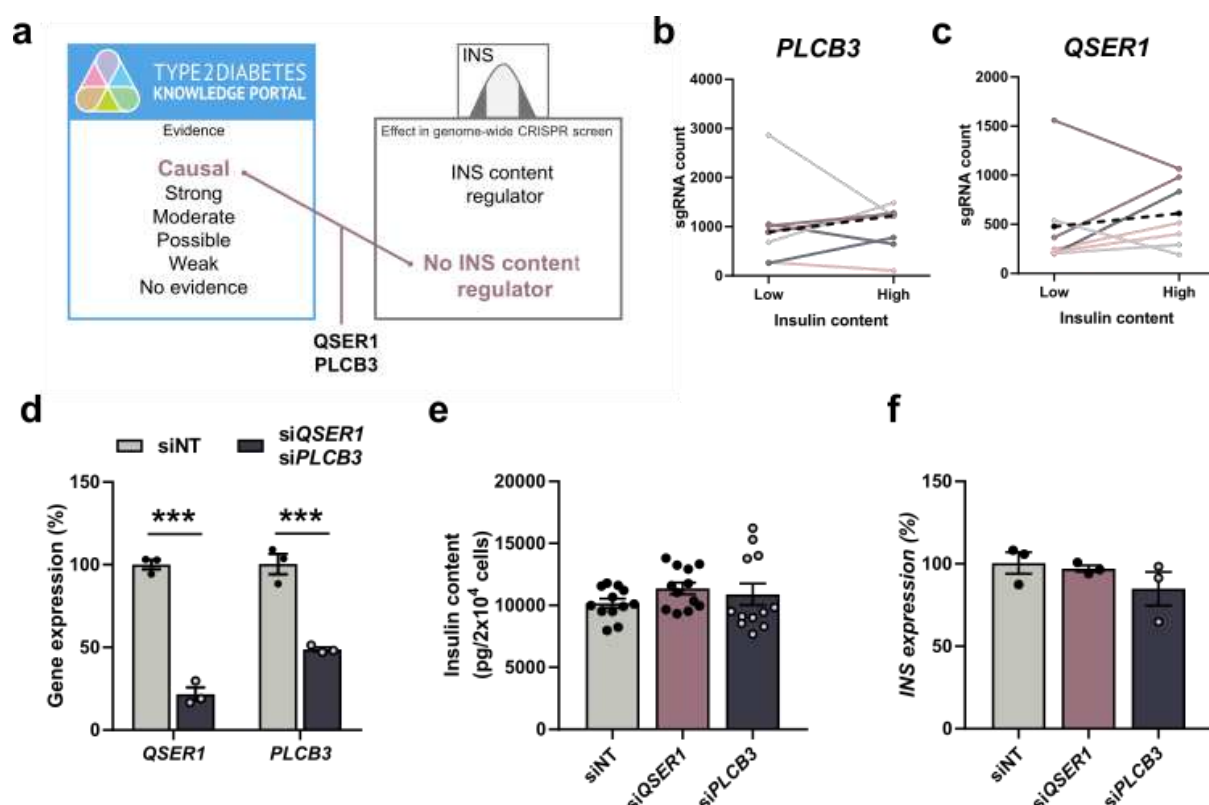
839

840

Fig 6. *QSER1* or *PLCB3* knockdown does not affect insulin content in EndoC-βH1. a) Gene prioritization approach combining genes with causal evidence as T2D effector transcripts (as outlined in the integration approach) while also not being a screening hit, highlighted *QSER1* and *PLCB3*. b) Changes in sgRNA count from low to high insulin content sample with each color representing the same sgRNA across the two screen replicates. The black dashed line represents the median sgRNA count for this gene. d-f) All data are from si*QSER1* or si*PLCB3* treated EndoC-βH1 compared to non-targeting control cells. d) mRNA expression of *QSER1* and PLCB3 in their respective silenced cells (blue) compared to siNT control cells (grey). e) Insulin content in pg per 20 000 cells. f) mRNA expression of *INS*. All data are mean ± SEM from three independent experiments or two independent genome-wide CRISPR screen replicates. Data were analysed using two-sample t-test (d) and one-way ANOVA with Sidak's multiple comparison test (e, f). P-values *** < 0.001. FC, fold change; LFC, log2 (fold change); NT, non-targeting.

854 **Supplementary Information**

855

856 **A genome-wide CRISPR screen identifies regulators of beta cell function**

857 **involved in type 2 diabetes risk**

858 Antje K Grotz, Elena Navarro-Guerrero, Romina J Bevacqua, Roberta Baronio, Soren K Thomsen,

859 Sameena Nawaz, Varsha Rajesh, Agata Wesolowska-Andersen, Seung K Kim, Daniel Ebner, Anna L
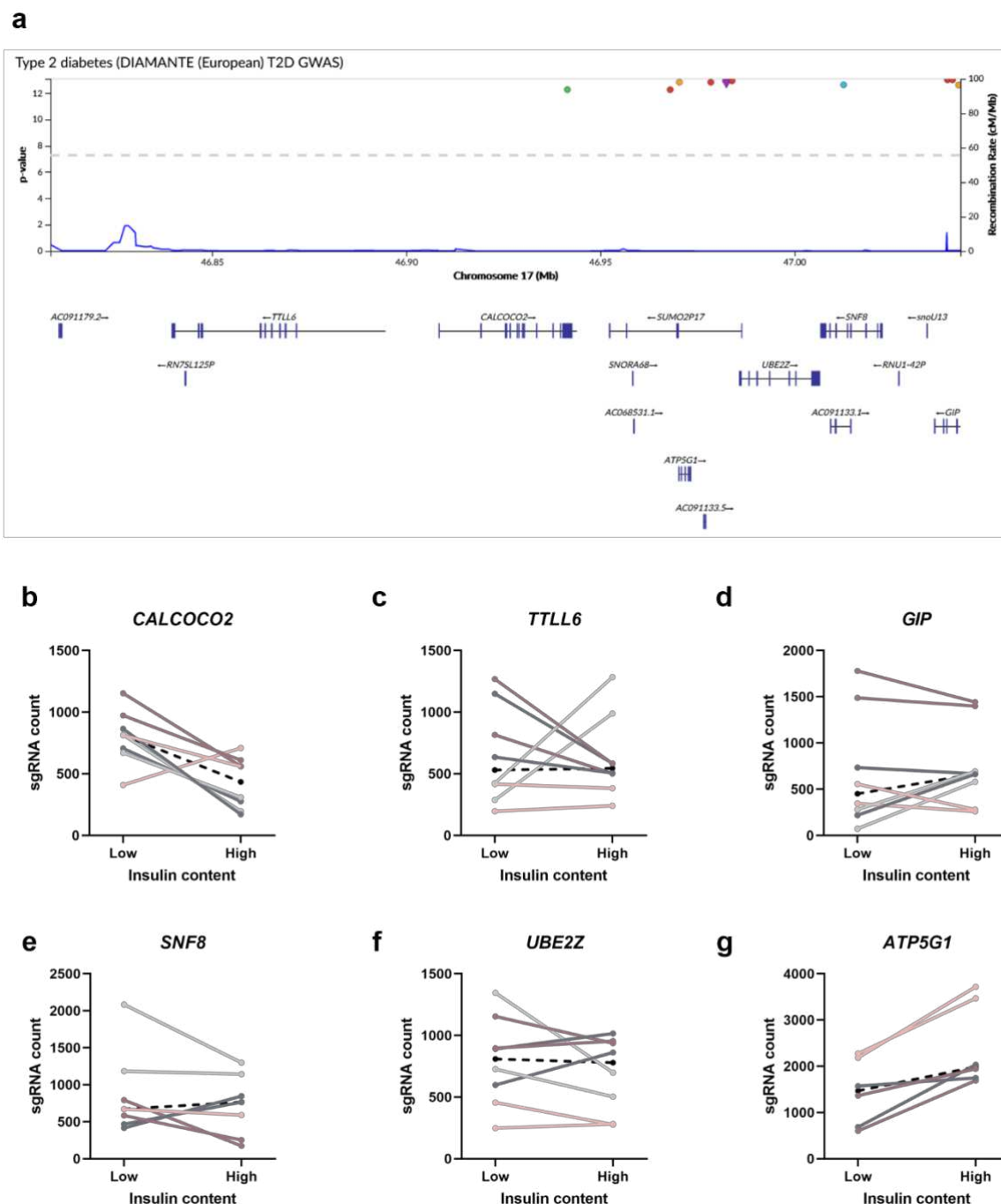
860 Gloyn

861
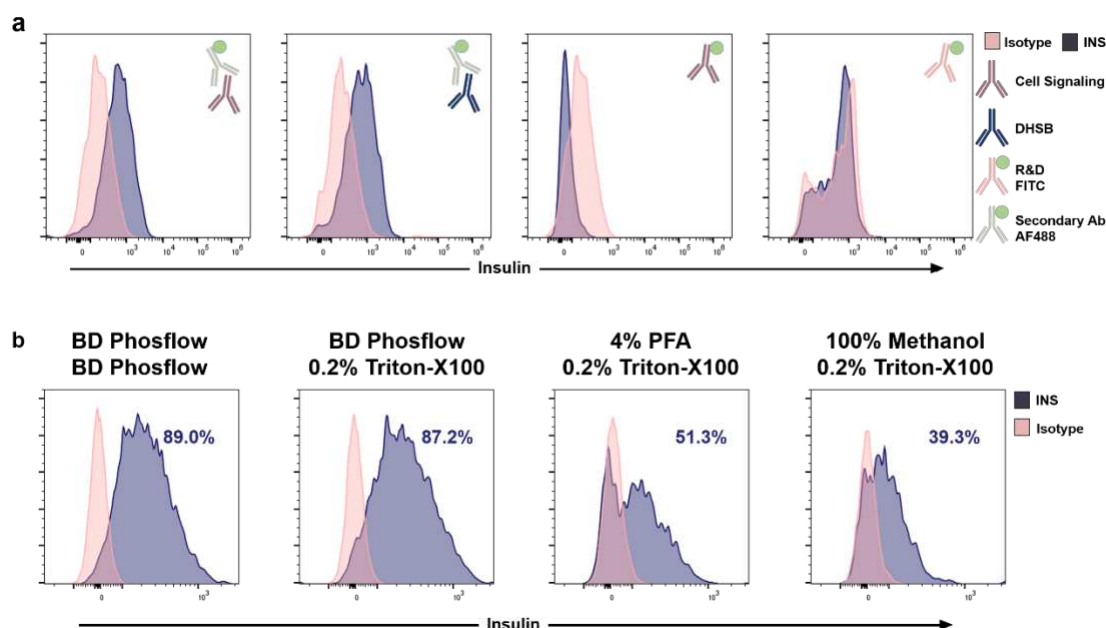
Supplementary Fig S1. FACS optimization for intracellular INS staining. Histograms showing different FACS staining strategies in EndoC-βH1 comparing intracellular insulin (blue graph) and respective isotype controls (pink graph) using a) different monoclonal primary antibodies as indicated (purple, blue, pink antibody) in combination with a secondary Ab (grey) conjugated to AF488 (green) or direct conjugation to FITC (green) and b) using different staining protocols with various fixation (top row) or permeabilization buffers (bottom row) as indicated. Three independent experiments were performed and representative histograms are shown.

871

872
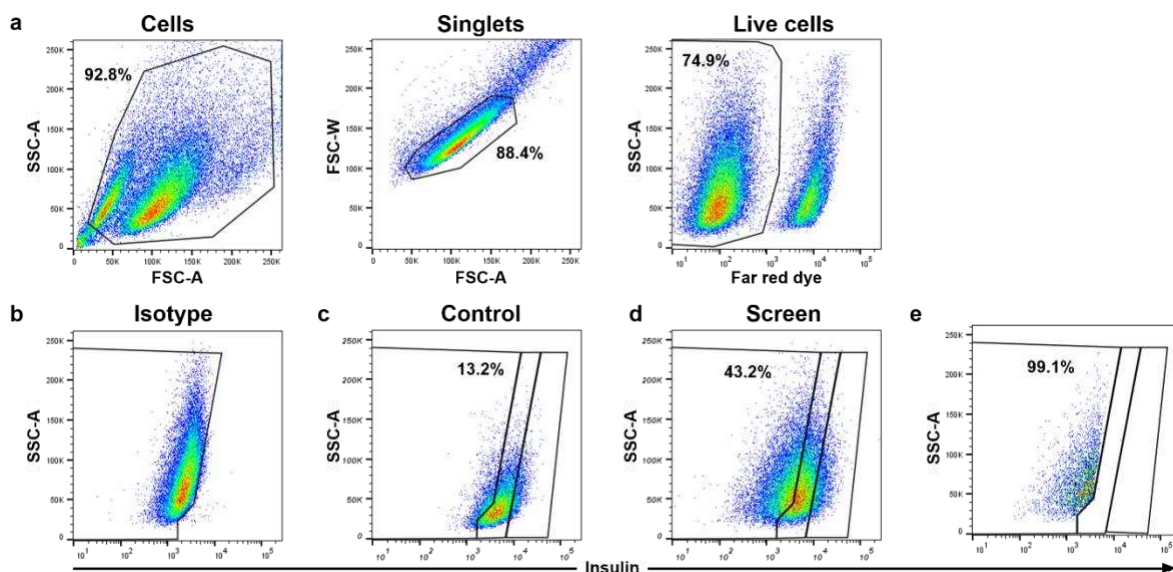
Supplementary Fig S2. FACS sorting of the genome-wide CRISPR screen. Representative FACS images and sorting gates for stained EndoC-βH1 in the genome-wide CRISPR screens. a) Sorting gates assessing cell size, granularity and viability to exclude debris, define singlets and live cells. b-d) Insulin staining of the isotype control (b), control cells (c) and screening cells (d). e) Resorting of low insulin cell population to assess the sample purity. SSC-A, side scatter area; FSC-A, forward scatter area; FSC-W, forward scatter width.

879

880
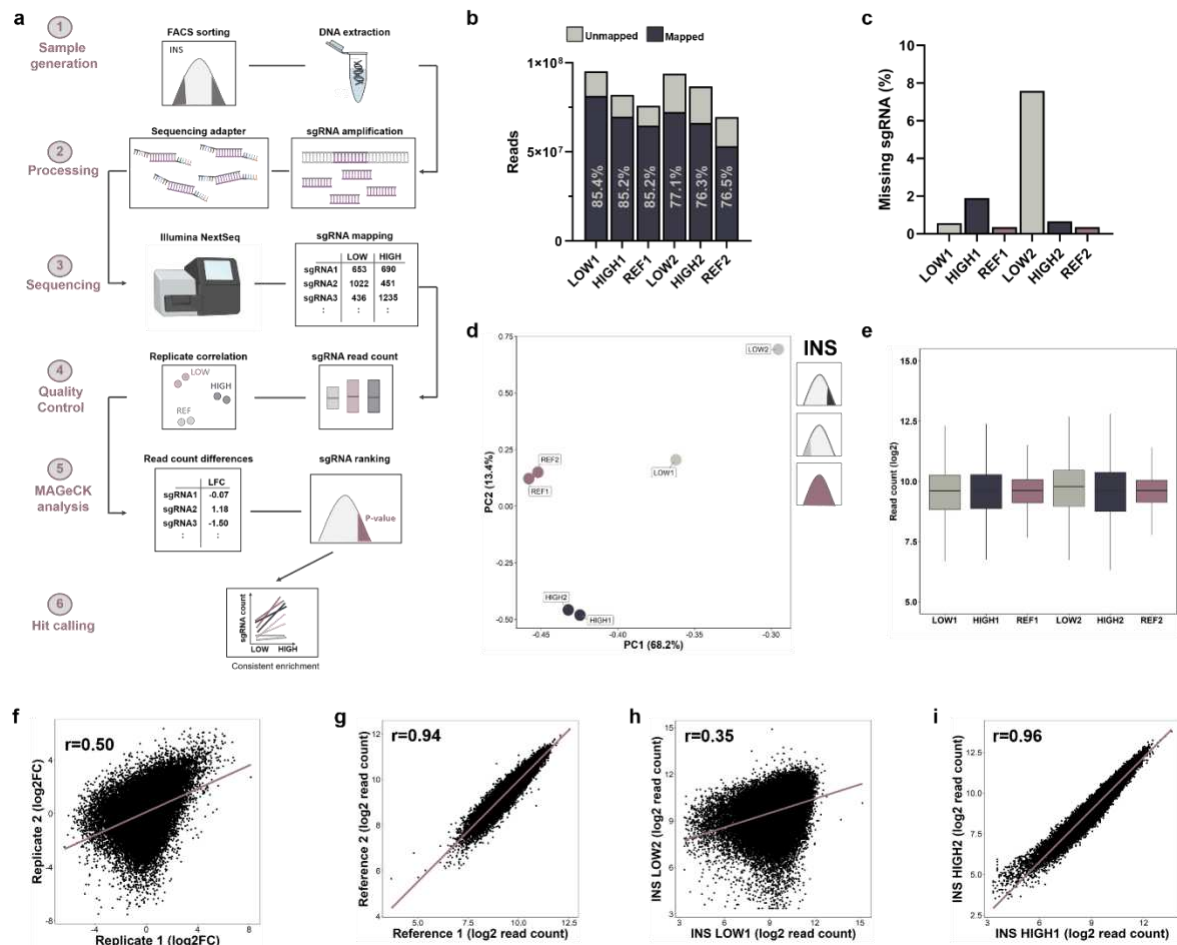
Supplementary Fig S3. CRISPR screen outline and quality control. a) CRISPR screen outline from initial FACS sorting of transduced EndoC-βH1 to final sequencing analysis. b-c,e) Read count and mapped sgRNA quality control for low insulin (LOW1, LOW2), high insulin (HIGH1, HIGH2) and reference samples (REF1, REF2) from both replicates including: (b) total read counts and the proportion of mapped reads and (c) the number of zero-count sgRNAs as proportion of total sgRNAs and (e) normalized read count distributions. d) Principal Component Analysis for the first two components of low *(grey)* and high *(blue)* insulin content and reference samples *(purple)* from both replicates. f-i) Pairwise Pearson correlation analysis of fold changes between low and high insulin content for each replicate (f), reference replicates (g), low insulin content replicates (h) and high insulin content replicates (i). The purple line indicates the linear regression line. Log2FC, log2 (fold change); r, Pearson correlation coefficient.

Supplementary Fig S4. Protein association networks of screening hits. STRING pathway analysis showing protein-protein associations including physical and functional interactions for CRISPR screening hits. Clusters were manually identified, annotated, and individually plotted and selected clusters are shown. Confidence level to determine interactions was set to highest confidence (0.9).

898

Supplementary Fig S5. *CALCOCO2* locus with its candidate genes. a) *CALCOCO2* locus with credible set GWAS association signals *(top)* and neighbouring genes *(bottom)*. b-g) sgRNA count distribution for all 8 sgRNAs per gene (4 sgRNAs per replicate) in the genome-wide CRISPR screen for genes at the *CALCOCO2* locus. Changes in sgRNA count from low to high insulin content sample with each colour representing the same sgRNA across the two screen replicates. The black dashed line represents the median sgRNA count for this gene.
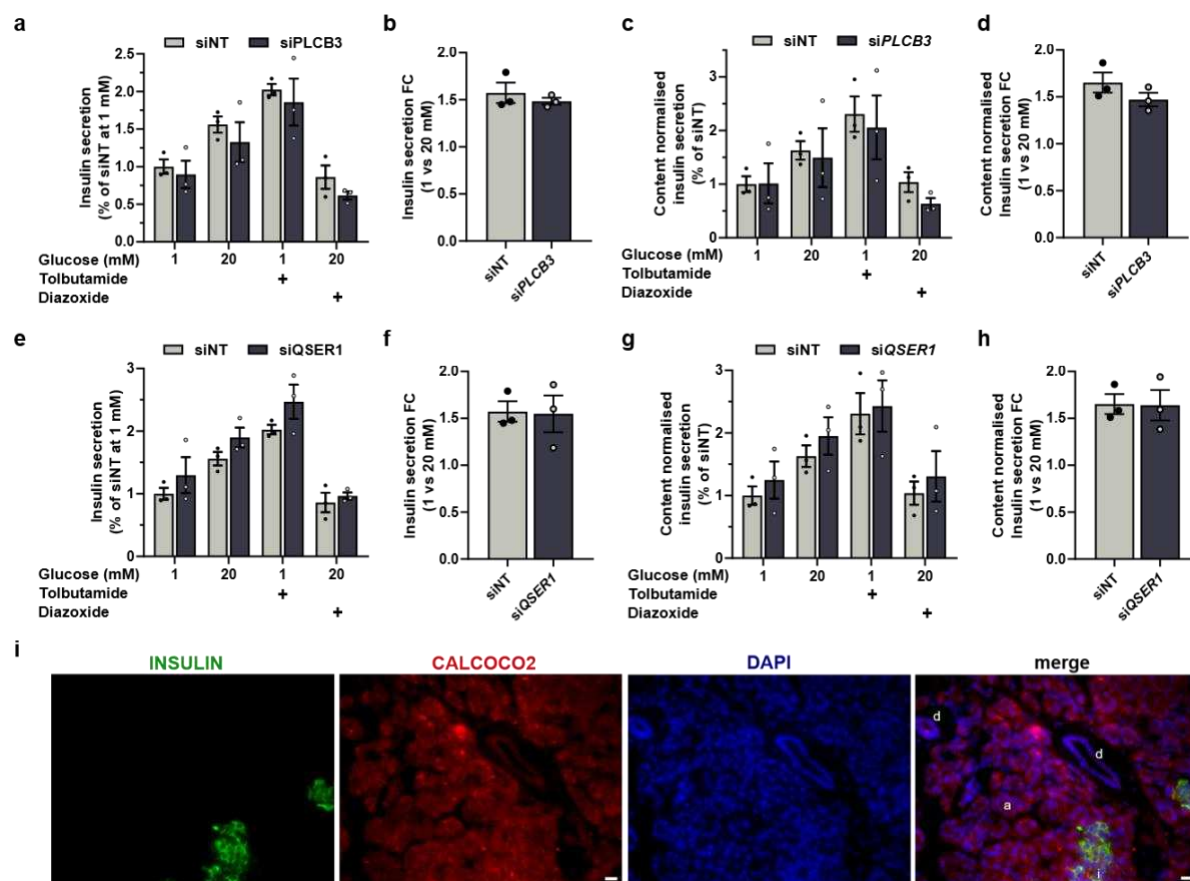
Supplementary Fig S6. *QSER1* and *PLCB3* in EndoC-βH1, CALCOCO2 in the exocrine pancreas. a-h) All data are from siRNA treated EndoC-βH1 (blue) as indicated compared to non-targeting control cells (grey). (a,e) Insulin secretion normalised to siNT or (c,g) to insulin content and siNT in 1 mM, 20 mM, 1 mM+100 μM tolbutamide or 20 mM glucose+100 μM diazoxide. (b,d,f,h) Insulin secretion fold change from 1 to 20 mM glucose. All data are mean ± SEM from three independent experiments. Data were analysed using two-way ANOVA Sidak's multiple comparison test (a,c,e,g) and two-sample t-test (b,d,f,h). P-values *** < 0.001. FC, fold change; NT, non-targeting. i) Immunofluorescence staining of exocrine tissue in pancreas sections. Sections were double immunostained for INS (*green*) and CALCOCO2 (*red*). Cell nuclei were counterstained with DAPI (*blue*). Scale bar is 10 μm. d: ductal cells; a: acinar cells; i: islet.

917 **Supplementary Table 1 – CRISPR Screening Hits.**

918

919 A complete list of the 580 genes meeting the stringent criteria set for reproducibly effecting insulin
920 content levels in the genome-wide CRISPR screen. Genes are shown based on fold change and the
921 direction of effect (increase or decrease) on insulin content indicated.

922

923

924

925 **Supplementary Table 2 – Prioritized causal genes for T2D based on integration with**
926 **T2D effector gene predictions.**

Genome-wide CRISPR screen	eQTL for T2D (Viñuela et al. 2020)
AGPAT2	ADCY5
ARAP1	AP3S2
CALCOCO2	CAMK1D
DCUN1D4	STARD10
FADS1	CEP68
IGF2	DGKB
INS	GPSM1
IRS2	HMG20A
MCCC1	PLEKHA1
MED23	ITGB6
NKX2-2	RNF6
NUDT3	TCF7L2
PELO	UBE2E2
PPP1R15B	NKX6-3
RBMS1	GRB14
SHQ1	HAUS6
SIN3A	IGF2BP2
SLC2A2	KLHL42
TBC1D4	ABCB9
ZNF101	SCD5
	SLC12A8
	SLC7A7
	PDE8B

927 Common genes from effector transcript predictions and the genome-wide CRISPR screen (left) or
928 from a recent human islet eQTL studies for T2D (right)[19].

929

930 **Supplementary Table 3 – Human Tissue Donor Details.**

Tissue source	Sample Identifier	Sex	Age (yrs)	BMI (Kg/m^2)
IIDP	SAMN18021384	Male	56	24.2
IIDP	SAMN18479112	Male	44	24.4
ADI	R398	Female	51	25.4
Other	OD30535	Female	31	48.2

931

932 IIDP, Integrated Islet Distribution Program (https://iidp.coh.org). ADI, Alberta Diabetes Institute
933 http://www.bcell.org/adi-isletcore.html

934

935 **Supplementary Table 4 – Resource Table.**

Primer	Sequence
Sequencing Primer	
PCR1_F	AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG
PCR1_R	TCTACTATTCTTTCCCCTGCACTGTTGTGGGCGATGTGCGCTCTG
R01	*AAGTAGAG#TTCTACTATTCTTTCCCCTGCACTGT
R02	*ACACGATC#ATTCTACTATTCTTTCCCCTGCACTGT
R03	*CGCGCGGT#GATTCTACTATTCTTTCCCCTGCACTGT
R04	*CATGATCG#CGATTCTACTATTCTTTCCCCTGCACTGT
F01	+TAAGTAGAGTCTTGTGGAAAGGACGAAACACCG
F02	+ATACACGATCTCTTGTGGAAAGGACGAAACACCG
F03	+GATCGCGCGGTTCTTGTGGAAAGGACGAAACACCG
F04	+CGATCATGATCGTCTTGTGGAAAGGACGAAACACCG
F05	+TCGATCGTTACCATCTTGTGGAAAGGACGAAACACCG
F06	+ATCGATTCCTTGGTTCTTGTGGAAAGGACGAAACACCG
F07	+GATCGATAACGCATTTCTTGTGGAAAGGACGAAACACCG
F08	+CGATCGATACAGGTATTCTTGTGGAAAGGACGAAACACCG
F09	+ACGATCGATAGGTAAGGTCTTGTGGAAAGGACGAAACACCG
SYBR green qPCR primer	
GAPDH_F	AGGCAACTAGGATGGTGTGG
GAPDH_Rev	TTGATTTTGGAGGGATCTCG
TBP_F	TGTTTGTTTCTTGCGAGTGC
TBP_Rev	CCCGGGAGGGTTCTTAGTAG
PPIA_F	AAGGGCTTTCGTGGAAACTT
PPIA_Rev	CCCCTCCATTCTCATCAAGA
LKO1.5R	GTTGATAACGGACTAGCCT
EV	TTTTTGTACGTCTCTGTTTTAGAGC
PAM	GACGAAACACCGGTTCAGAAC
INS_1	CAATGCCACGCTTCTGCAG
INS_2	CATCTGCTCCCTCTACCAGC
NMS_1	AAACACCGTGCAGGATCACA
NMS_2	ACACCGACAATATCCAAGCCA

sgRNA	Target	Sequence
sgRNA details		
sgRNA 1	*PAM*	GAACTAGCAGGCTAGGGACG
sgRNA 2	*PAM*	GTTCAGAACCATACCACCAG
sgRNA 5	*INS*	CACAATGCCACGCTTCTGCA
sgRNA 6	*INS*	CATCTGCTCCCTCTACCAGC
sgRNA 7	*NMS*	TGCAGGATCACACTGCGACC

sgRNA 8		*NMS*	ACAATATCCAAGCCATCTGA	
Target	**Dilution**	**Species**		**Details**
Western Blot Antibodies				
β-Tubulin	1:2000	Mouse monoclonal		Santa Cruz, sc-365791
GAPDH	1:10 000	Rabbit monoclonal		Abcam, ab37168
INS	1:1000	Mouse monoclonal		Santa Cruz, sc-393887
C-peptide	1:1000	Mouse monoclonal		ExBio, 11-247-C100
CALCOCO2	1:1000	Mouse monoclonal		Santa Cruz, sc-376540
Anti-mouse IgG HRP	1:2500	Rabbit polyclonal		Thermo Fisher, 31460
Anti-rabbit IgG HRP	1:2500	Goat polyclonal		Thermo Fisher, 31450
Immunofluorescence Antibodies				
CALCOCO2	1:100	Mouse monoclonal		Santa Cruz, sc-376540
INS	1:300	Guinea-pig polyclonal		Dako, A0564
Anti-mouse	1:300	Donkey		Invitrogen, #A-31570
Anti-guinea-	1:300	Donkey		Abcam, ab150187
FACS antibodies				
INS	1:10	Rabbit monoclonal		Cell Signaling, #3014
INS	1:10	Rat monoclonal		DHSB, GN-ID4
INS	1:10	Rat monoclonal		R&D, MAB1417
IgG	1:10	Rabbit monoclonal		Cell Signaling, #3900
IgG	1:10	Rat monoclonal		Invitrogen, 02-9602
Anti-rat	1:500	Chicken polyclonal		Invitrogen, A-21470
Anti-rabbit	1:200	Goat polyclonal		Invitrogen, A-11034

936　Key experimental resources such as primer and antibodies are listed. The following primer sequences
937　were abbreviated: CAAGCAGAAGACGGCATACGAGAT (*),
938　GTGACTGGAGTTCAGACGTGTGCTCTTCCGA TCT (#) and
939　AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTT CCGATCT
940　(+).

941

942