

# BACH2 inhibition reverses $\beta$ cell failure in type 2 diabetes models

Jinsook Son,<sup>1,2</sup> Hongxu Ding,<sup>3</sup> Thomas B. Farb,<sup>4</sup> Alexander M. Efanov,<sup>4</sup> Jiajun Sun,<sup>5</sup> Julie L. Gore,<sup>4</sup> Samreen K. Syed,<sup>4</sup> Zhigang Lei,<sup>4</sup> Qidi Wang,<sup>5</sup> Domenico Accili,<sup>1,2</sup> and Andrea Califano<sup>3</sup>

<sup>1</sup>Department of Medicine and <sup>2</sup>Naomi Berrie Diabetes Center, Vagelos College of Physicians and Surgeons, Columbia University, New York, New York, USA. <sup>3</sup>Department of Systems Biology, Columbia University Irving Medical Center, New York, New York, USA. <sup>4</sup>Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, Indiana, USA. <sup>5</sup>Shanghai National Clinical Research Center for Endocrine and Metabolic Diseases, Key Laboratory for Endocrine and Metabolic Diseases of the National Health Commission of the PR China, Shanghai Institute of Endocrine and Metabolic Disease, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China.

**Type 2 diabetes (T2D) is associated with defective insulin secretion and reduced  $\beta$  cell mass. Available treatments provide a temporary reprieve, but secondary failure rates are high, making insulin supplementation necessary. Reversibility of  $\beta$  cell failure is a key translational question. Here, we reverse engineered and interrogated pancreatic islet-specific regulatory networks to discover T2D-specific subpopulations characterized by metabolic inflexibility and endocrine progenitor/stem cell features. Single-cell gain- and loss-of-function and glucose-induced  $\text{Ca}^{2+}$  flux analyses of top candidate master regulatory (MR) proteins in islet cells validated transcription factor BACH2 and associated epigenetic effectors as key drivers of T2D cell states. BACH2 knockout in T2D islets reversed cellular features of the disease, restoring a nondiabetic phenotype. BACH2-immunoreactive islet cells increased approximately 4-fold in diabetic patients, confirming the algorithmic prediction of clinically relevant subpopulations. Treatment with a BACH inhibitor lowered glycemia and increased plasma insulin levels in diabetic mice, and restored insulin secretion in diabetic mice and human islets. The findings suggest that T2D-specific populations of failing  $\beta$  cells can be reversed and indicate pathways for pharmacological intervention, including via BACH2 inhibition.**

## Introduction

Type 2 diabetes (T2D) is characterized by pancreatic  $\beta$  cell function and mass abnormalities (1). The clinical features of  $\beta$  cell failure include limited adaptive potential of  $\beta$  cell mass (2, 3), increased glucagon tone (4), rapid progression in response to incipient hyperglycemia (5), and limited response to treatment (6, 7). In this regard, whether  $\beta$  cell failure is reversible remains arguably the key question bearing on the issue of disease modification (8). Understanding islet cell dysfunction is critical to design mechanism-based, durable, and safe disease-modifying interventions (9). Yet, the limited ability to access the endocrine pancreas in vivo has hindered our ability to functionally interrogate islet cells. To wit, despite intensive, multicenter efforts to standardize procurement and analysis of human islets, there is no consensus on a diabetic gene expression signature (10), and mechanistic understanding of  $\beta$  cell failure pathways is still lacking. Indeed,

while these studies have demonstrated islet cell gene expression heterogeneity (11–14), clear definition of a diabetic islet cell signature remains elusive.

Single-cell-based expression profiling can reveal disease determinants, but its ability to detect subtle differences, such as those between a diabetic and nondiabetic cell type (15), is hampered by the fact that differential gene expression may not be assessed for up to 80% of the genes (gene dropout), due to low sequencing depth (16). To address this problem, we have developed the metaVIPER algorithm — an extension to single cells of the VIPER algorithm (17) — which can accurately measure the activity of approximately 6,500 regulatory and signaling proteins in single cells, based on the enrichment of their transcriptional targets (regulon) in differentially expressed genes (18). metaVIPER can accurately identify rare cell subpopulations that otherwise escape detection by expression-based analyses, by measuring protein activity in single cells (19).

In this work, we applied metaVIPER to identify master regulatory (MR) proteins controlling the potentially noisy gene expression signature of diabetic islet cells. In a critical validation of the computational analysis, we show that single-cell-based, CRISPR-mediated inhibition of metaVIPER-inferred MR proteins, such as BACH2, can effectively reverse T2D-specific  $\beta$  cell features. Moreover, immunohistopathology of diabetic pancreata demonstrates an enrichment of BACH2-expressing cells, as predicted by the metaVIPER algorithm. Finally, we show that treatment with a BACH inhibitor improves disease phenotypes in diabetic mice

► **Related Commentary:** <https://doi.org/10.1172/JCI154482>

**Authorship note:** JS and HD contributed equally to this work.

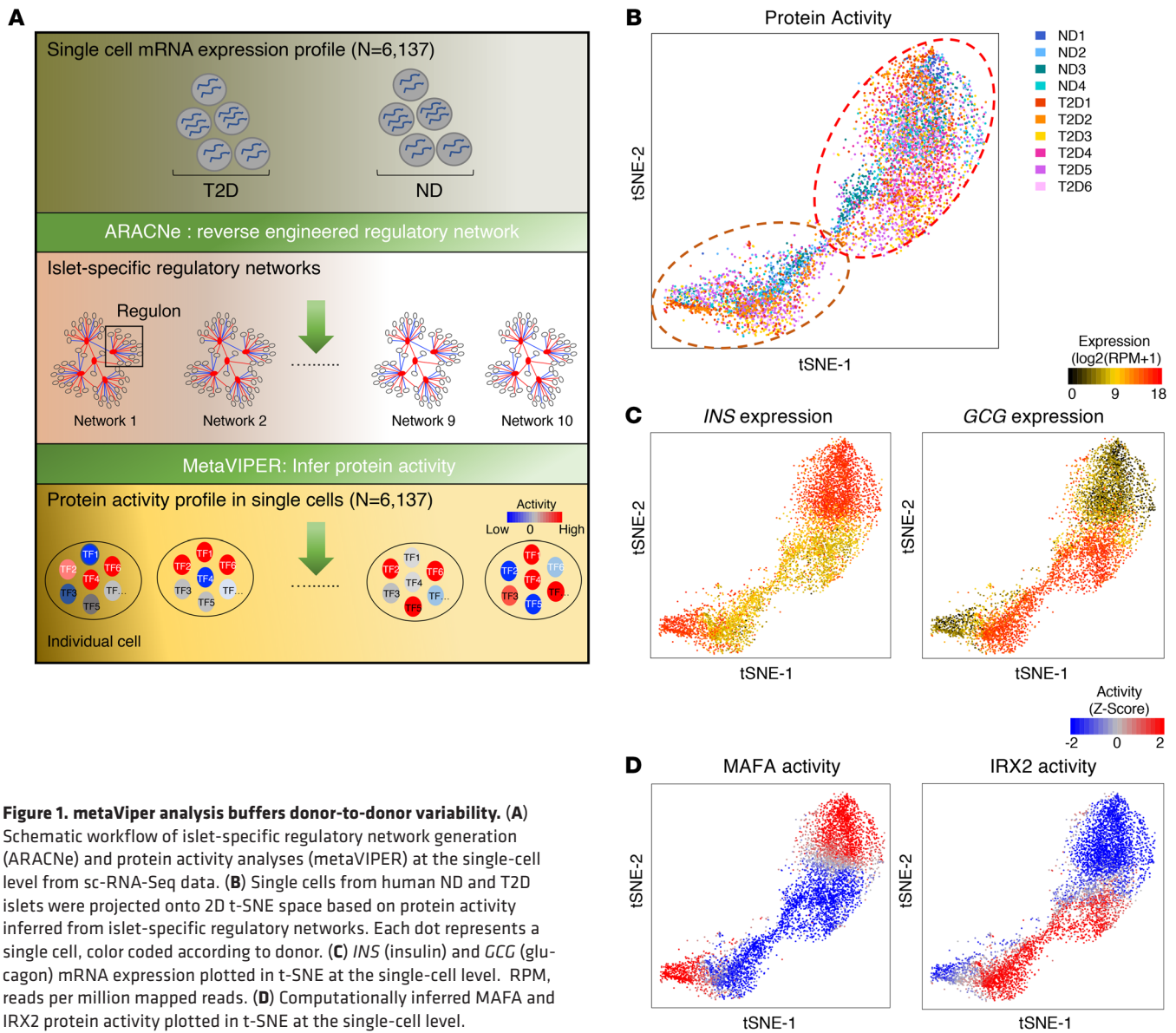
**Conflict of interest:** AC is founder, SAB member, and shareholder of DarwinHealth Inc., which has licensed the VIPER and metaVIPER software. DA is founder, director, and chairman of the SAB of Forkhead Biotherapeutics, corp. Columbia University is also a shareholder of DarwinHealth Inc.

**Copyright:** © 2021, American Society for Clinical Investigation.

**Submitted:** August 4, 2021; **Accepted:** October 28, 2021; **Published:** December 15, 2021.

**Reference information:** *J Clin Invest.* 2021;131(24):e153876.

<https://doi.org/10.1172/JCI153876>.



**Figure 1. metaViper analysis buffers donor-to-donor variability. (A)**

Schematic workflow of islet-specific regulatory network generation (ARACNe) and protein activity analyses (metaVIPER) at the single-cell level from sc-RNA-Seq data. **(B)** Single cells from human ND and T2D islets were projected onto 2D t-SNE space based on protein activity inferred from islet-specific regulatory networks. Each dot represents a single cell, color coded according to donor. **(C)** *INS* (insulin) and *GCG* (glucagon) mRNA expression plotted in t-SNE at the single-cell level. RPM, reads per million mapped reads. **(D)** Computationally inferred MAFA and IRX2 protein activity plotted in t-SNE at the single-cell level.

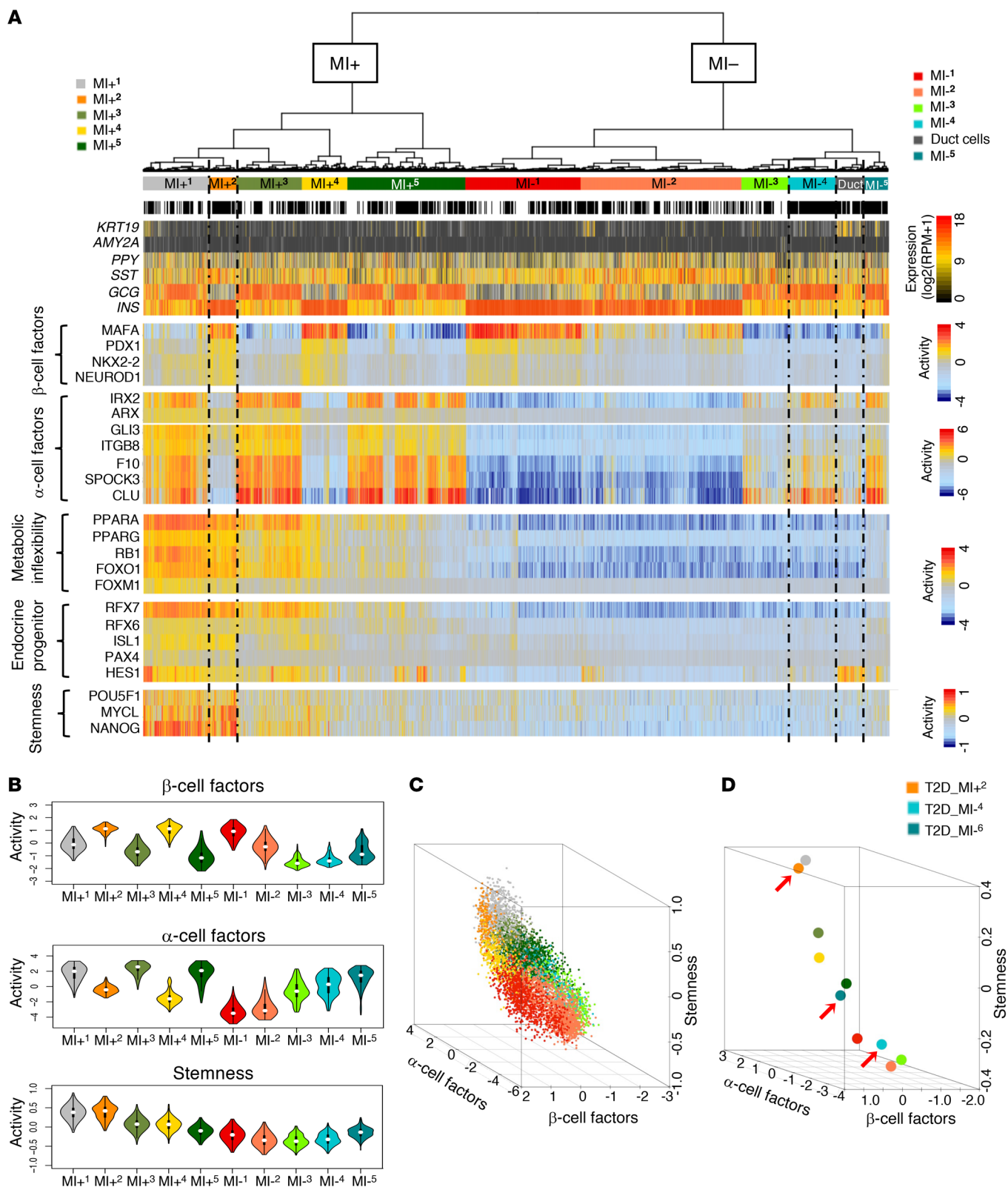
and insulin secretion in diabetic islets of mice and humans, thus providing proof of concept of the potential translational value of the findings. Thus, single-cell protein-activity-based analyses can discover diabetes-causing phenotypes, identify their MR proteins, and guide their pharmacological targeting. This work highlights the clinical potential of disease modification as a conceptual advance in diabetes research.

## Results

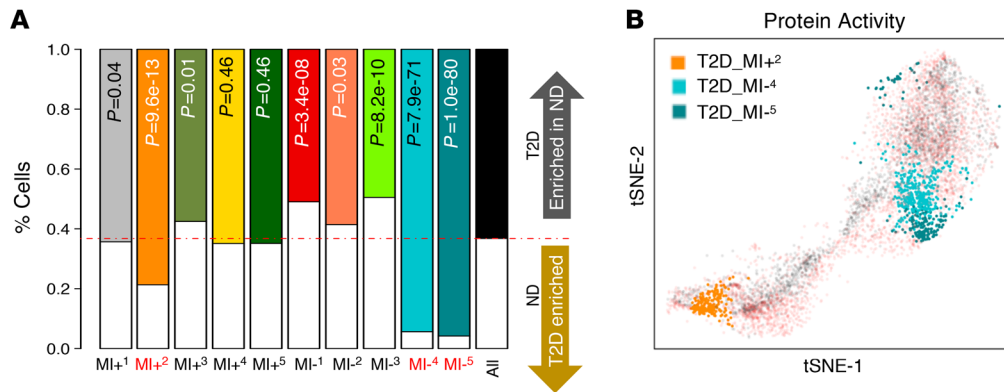
*Single-cell protein activity analysis of pancreatic islet cells.* To capture T2D-specific cellular states, we used single-cell RNA sequencing (scRNA-Seq) profiles from 4 nondiabetic controls (ND) and 6 T2D donors (Supplemental Table 1; supplemental material available online with this article; <https://doi.org/10.1172/JCI1153876DS1>) to build islet-specific transcription regulatory networks and to identify the transcription factors (TFs) and cofactors (co-TFs) that control their differential gene expression signatures (Figure 1A). To measure differential TF activity in individual cells, we first used

ARACNe — a highly validated reverse-engineering algorithm (20) — to build tissue-specific regulatory networks from single cells, as shown in Ding et al. (18). Use of single-cell data greatly exceeded ARACNe's requirements for large data sets ( $n \geq 100$  samples), resulting in a comprehensive, first-of-a-kind pancreatic islet cell regulatory network, assembled from 6,137 scRNA-Seq profiles from T2D or ND islets. The resulting network is composed of regulons for 1,813 TFs, 969 co-TFs, and 3,370 signal transduction (ST) proteins. Since direct physical regulation of gene expression signatures is implemented by TFs and co-TFs, we used the network to measure the differential activity of TF and co-TF proteins in each single cell, compared with the average of all single cells, using metaVIPER (ref. 18 and Figure 1A).

t-SNE (t-distributed stochastic neighbor embedding) analysis based on metaVIPER-measured protein activity effectively removed most donor-to-donor batch effects, resulting in 2 distinct, yet donor-independent, cell clusters (Figure 1B and Supplemental Figure 1). The 2 clusters did not represent the 2 major



**Figure 2.** iterClust classifies ND and T2D islet cells into different biological states. **(A)** iterClust analyses performed using ND and T2D islet cells. The resulting cluster architecture was visualized as a heatmap. Each subgroup is color coded. Each bar denotes a single cell. Black bars represent T2D cells and white bars ND cells. *KRT19*, *AMY2A*, *PPY*, *SST*, *GCG*, and *INS* mRNA expression is plotted at the single-cell level. metaViper-inferred protein activity for β or α cell factors, metabolic inflexibility, endocrine progenitor, and stemness markers is plotted at the single-cell level. RPM, reads per million mapped reads. **(B)** Violin plots showing the distribution of cells in each cluster based on integrated activity of β cell factors (top), α cell factors (middle), or stemness markers (bottom). **(C)** 3D plot showing integrated β cell factor, α cell factor, and stemness activity on the x, y, and z axes, respectively, at the single-cell level. **(D)** 3D plot as in **C** but based on the average cell behaviors of each cluster. Red arrows indicate T2D-enriched clusters, MI+<sup>2</sup> (T2D-β-like) and MI-<sup>4</sup> and MI-<sup>5</sup> (both T2D-α-like).



**Figure 3. T2D-enriched subclusters in human islets.** (A) Bar plots presenting the percentage of ND or T2D cells in each subcluster. A dashed line represents the proportion of cells from ND or T2D islets analyzed by scRNA-Seq. *P* values were derived from Fisher's exact test. (B) Single cells from human ND and T2D islets were projected onto 2D t-SNE space based on protein activity inferred from islet-specific regulatory networks. ND cells are shown in gray and T2D cells are shown in red as background. T2D cells in T2D-enriched clusters, MI+<sup>2</sup> (T2D-β-like) and MI-<sup>4</sup> and MI-<sup>5</sup> (both T2D-α-like), were color coded as indicated.

canonical islet populations (β and α cells), as shown by integrative assessment of *INS* (insulin) and *GCG* (glucagon) mRNA levels (Figure 1C), as well as MAFA and IRX2 protein activity (Figure 1D). Yet, *INS* and *GCG* mRNA expression aligned perfectly with metaVIPER-measured MAFA and IRX2 activity, respectively, confirming metaVIPER's accuracy in marker activity measurement.

Statistically significant superiority of protein activity versus gene-expression-based clustering was recently demonstrated (21), including at the single-cell level (22). Thus, to further cluster islet cells, including both ND and T2D cells, we performed unsupervised hierarchical clustering analysis using iterative clustering (iterClust) algorithms (23) with the full, metaVIPER-inferred protein activity matrix. The analysis yielded 11 clusters (Figure 2A and Supplemental Figure 2), which were further characterized by assessing established lineage and functional markers, including (a) hormone mRNA expression, (b) activity of TFs characteristic of either β or α cell identity (24), (c) activity of metabolic inflexibility/stress-response drivers (25), and (d) activity of endocrine progenitor and stem-like-cell TFs (ref. 26 and Supplemental Table 2). The latter were chosen based on experimental evidence from animal studies of their role in disease progression (25, 27, 28).

Protein-based cluster analysis identified 2 main clusters, characterized by differential activity of metabolic inflexibility/stress-response markers (25, 27) (MI+ and MI-) — a stage in the progression of β cell failure (25) — including PPARα and PPARγ (Figure 2A and Supplemental Figure 3). Expression of the corresponding genes was virtually undetectable in most cells (Supplemental Figure 4), thus confirming metaVIPER's superior sensitivity and complementarity, compared with mRNA measurements (Supplemental Figure 5). Regulatory proteins, including TFs and co-TFs, are often expressed at low levels in terminally differentiated cells (29). More importantly, their mRNA expression is a poor predictor of their activity, due to complex posttranscriptional and posttranslational events. scRNA-Seq data have low sequencing depth, and as a result most genes (generally >80%) cannot be detected by even a single read (dropouts; ref. 18). metaVIPER effectively overcomes this limitation, because the differential

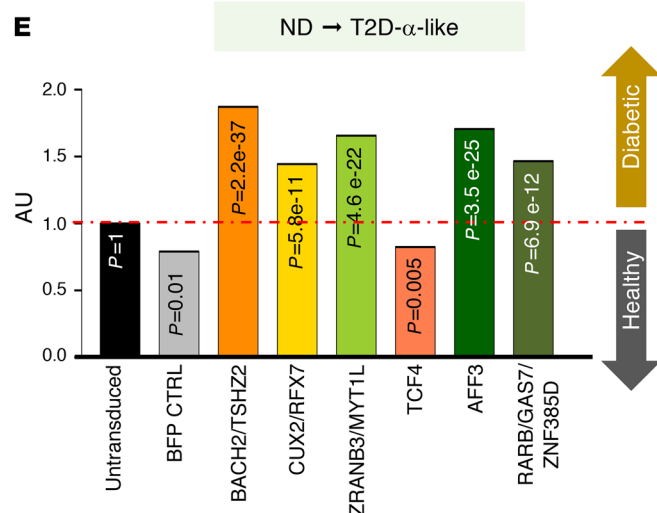
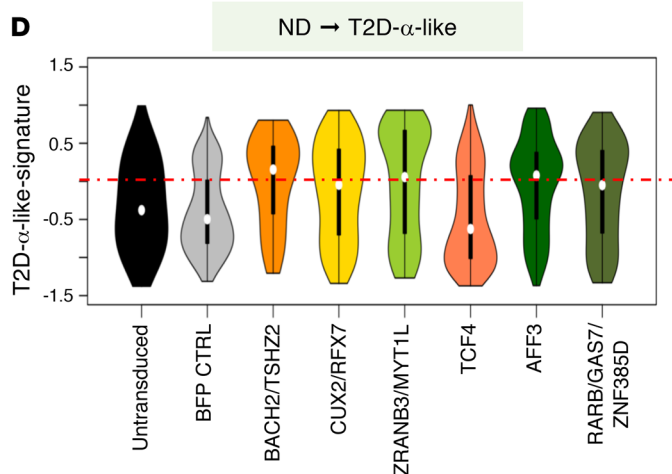
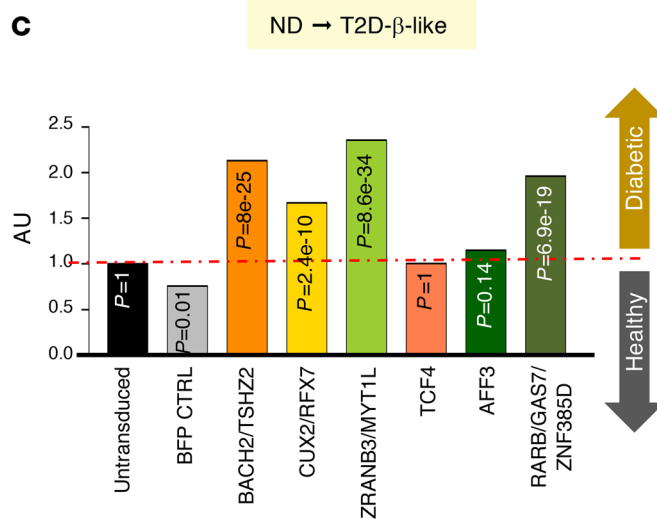
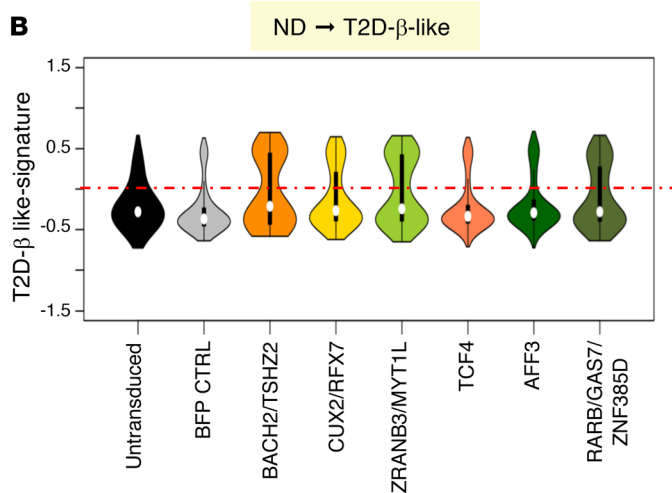
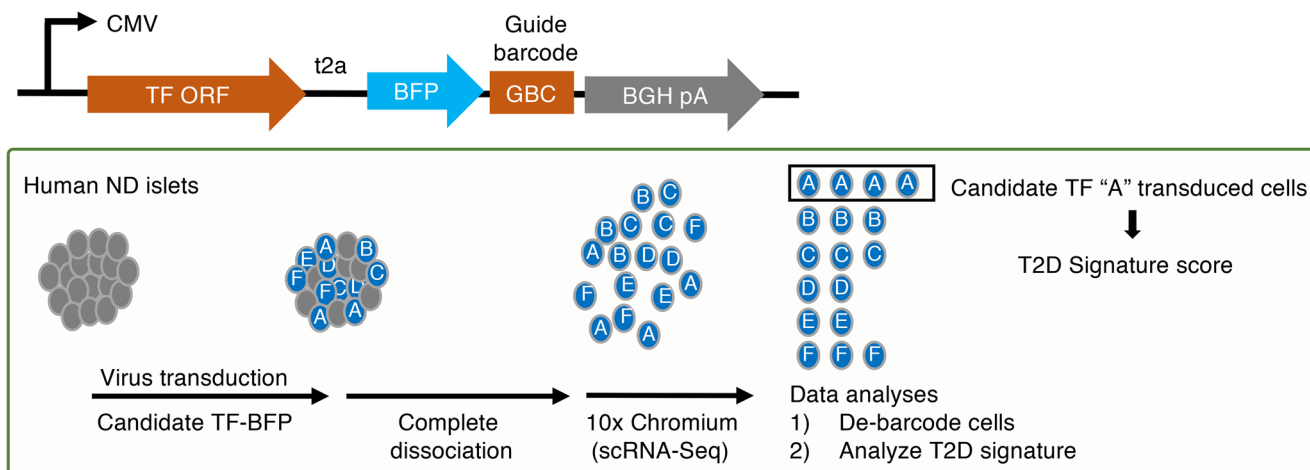
activity of a protein is measured from the differential expression of 50 or more of its transcriptional targets, akin to a highly multiplexed gene reporter assay, thus allowing reliable protein activity assessment even from low-depth profiles (18, 22).

MI+ cells showed clear evidence of a metabolic stress response (30), denoted by high FOXO1, HIF1α, HSF1, and TP53 activity (Supplemental Figure 3, A and B); endocrine progenitor features, denoted by high RFX6 (31, 32) and RFX7 activity (Figure 2A and Supplemental Figure 3C); and cell stemness, denoted by high NANOG, MYCL, and POU5F1 activity in a subset of RFX6/7-positive MI+ cells (Figure 2A and Supplemental Figure 3D). Thus, the 2 main functional clusters in ND and T2D islets are defined by differential activity of metabolic inflexibility/stress-response and endocrine progenitor/stem-cell-like markers (25, 27).

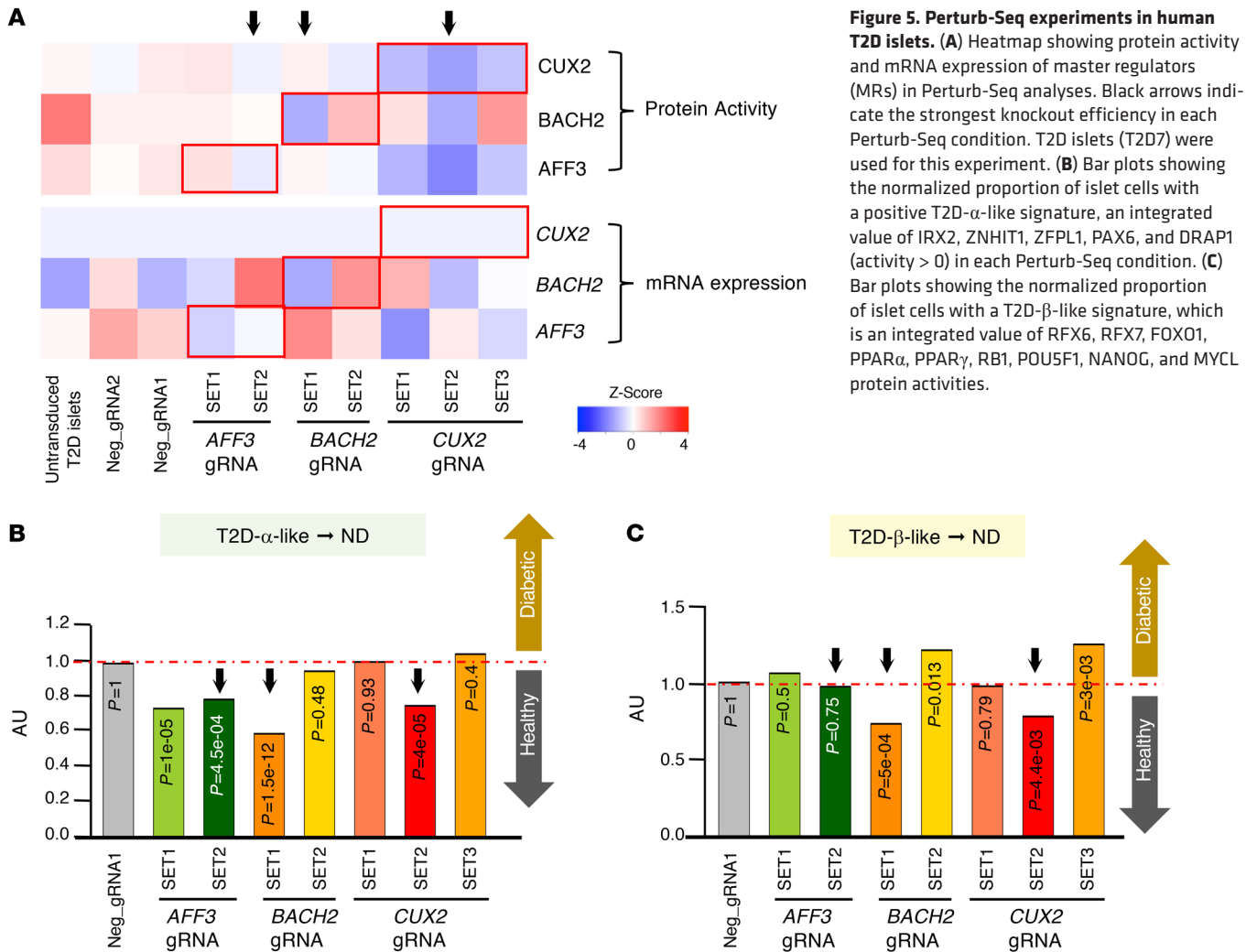
Iterative clustering further divided the MI+ cluster into 5 distinct subclusters — labeled MI+<sup>1</sup> to MI+<sup>5</sup> — and the MI- cluster into 6 distinct subclusters — labeled MI-<sup>1</sup> to MI-<sup>6</sup>. However, MI-<sup>6</sup> appeared to contain *KRT19*-positive pancreatic ductal cells and was thus excluded from further analyses. Four clusters (MI+<sup>2</sup>, MI+<sup>4</sup>, MI-<sup>1</sup>, and MI-<sup>2</sup>) displayed β cell features. Specifically, MI-<sup>1</sup> contained the healthiest β cells, as indicated by the highest levels of *INS* expression and β cell-specific TF activity, as well as by inactive metabolic inflexibility/stress response, progenitor/stem cell, and α cell identity (Figure 2, A-D); MI-<sup>2</sup> had similar, yet less pronounced features, while MI+<sup>2</sup> and MI+<sup>4</sup> differed from the other 2 clusters based on activation of metabolic inflexibility/stress-response and progenitor/stem cell features (Figure 2, A-D, and Supplemental Videos 1 and 2).

In contrast, 6 clusters (MI+<sup>1</sup>, MI+<sup>3</sup>, MI+<sup>5</sup>, MI-<sup>3</sup>, MI-<sup>4</sup>, and MI-<sup>5</sup>) had α cell or α cell-like features. Specifically, MI+<sup>5</sup> showed the strongest α cell identity, with high *GCG* expression and IRX2 activity, and slight activation of metabolic inflexibility/stress response (Figure 2, A-D); MI+<sup>3</sup> had similar features but stronger activation of metabolic inflexibility/stress response, while MI+<sup>1</sup>, also showed high progenitor/stem cell activities (Figure 2, A-D, and Supplemental Videos 1 and 2). Finally, MI-<sup>5</sup>, MI-<sup>4</sup>, and MI-<sup>3</sup> showed gradually decreasing α cell activity, despite preserved *GCG* expression and absence of metabolic inflexibility/stress response or progenitor cell features (Figure 2, A-D).

**A** Adenovirus encoding candidate TF and unique DNA barcode



**Figure 4. scGOF-Seq experiments in human islets.** (A) Schematic drawing of scGOF-Seq plasmids and experimental procedure. ND islets (ND5) were used for this experiment. (B) Violin plots showing the distribution of cells following transduction with each individual candidate or combination thereof analyzed according to T2D-β-like signature, an integrated value of RFX6, RFX7, FOXO1, PPARα, PPARγ, RB1, POU5F1, NANOG, and MYCL protein activities. Nontransduced and BFP-transduced ND islets serve as negative controls. (C) Bar plots showing the normalized proportion of islet cells with a positive T2D-β-like signature (activity > 0) in each scGOF-Seq condition over nontransduced control. (D) Violin plots showing cells with a T2D-α-like signature, which is an integrated value of IRX2, ZNHIT1, ZFPL1, PAX6, and DRAP1. (E) Bar plots showing the normalized proportion of islet cells with a T2D-α-like signature as in B. P values in C and E were derived from Fisher's exact test.

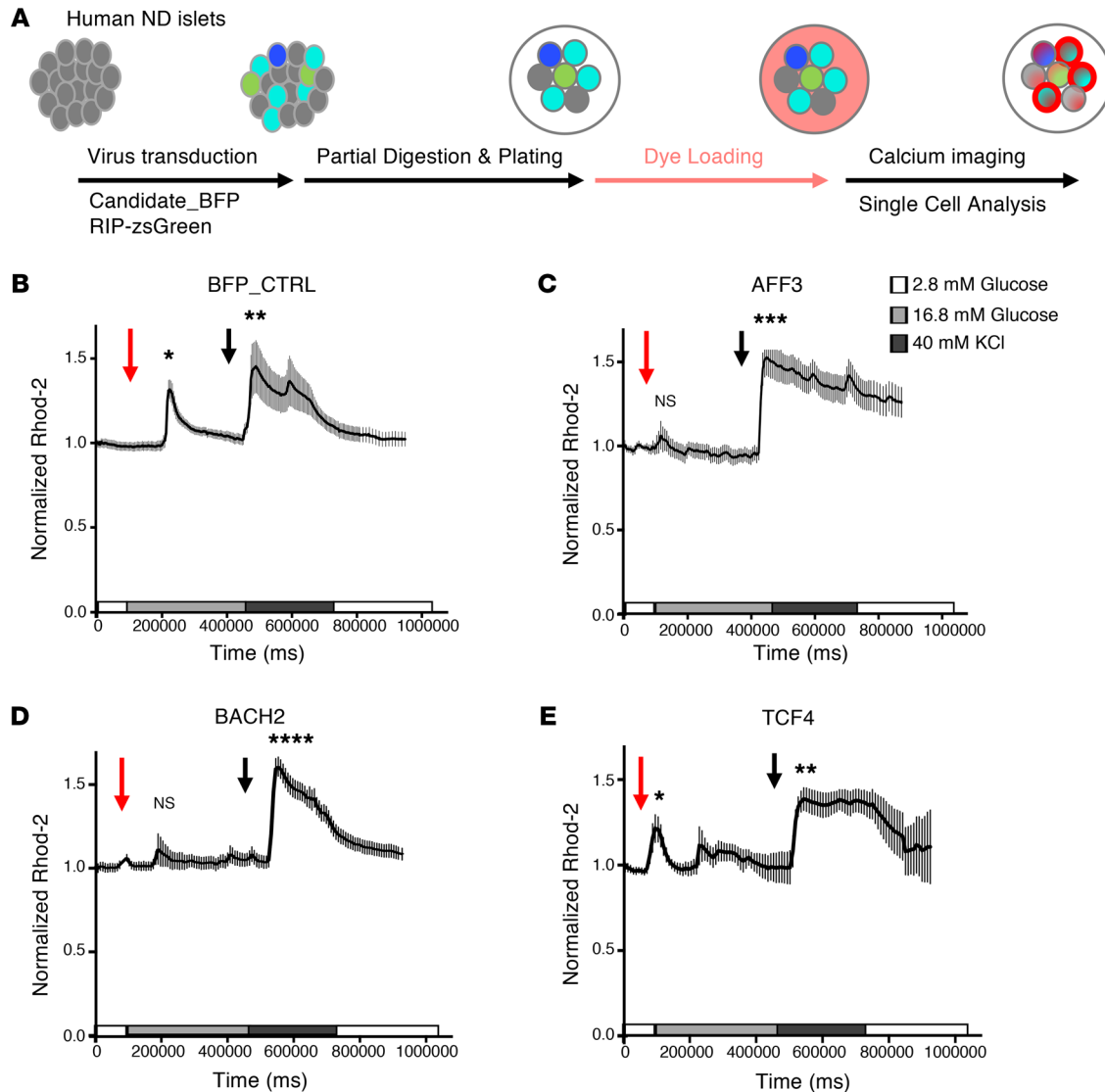


Single-cell protein activity analysis reveals T2D-enriched islet subpopulations. Next, we assessed whether the 10 molecularly distinct clusters were differentially represented in T2D versus ND cells and evaluated the statistical significance of their association (Figure 3A). Indeed, t-SNE plots revealed highly distinctive cluster representation patterns and differential subpopulation representation (Supplemental Figures 1 and 2). Notably, cluster analysis was performed without separating ND and T2D cells and using a common, islet-specific regulatory network, thus allowing for the unbiased, unsupervised determination of T2D- versus ND-specific cells across detected clusters.

The analysis revealed that ND islets were highly enriched in “healthy”  $\beta$  cells (MI<sup>-1</sup>), and  $\alpha$ -like cells characterized by weak  $\alpha$  cell markers but inactive metabolic inflexibility/stress response and progenitor/stem cell features (MI<sup>-3</sup>) ( $P = 3.4 \times 10^{-8}$  and  $8.2 \times 10^{-10}$ , respectively, by Fisher’s exact test). In contrast, T2D islets were highly enriched in 3 clusters: MI<sup>+2</sup>, MI<sup>-4</sup>, and MI<sup>-5</sup> ( $P = 9.6 \times 10^{-13}$ ,  $7.9 \times 10^{-71}$ , and  $1.0 \times 10^{-80}$ , respectively; Figure 3, A and B). Among these, MI<sup>+2</sup> is composed of  $\beta$  cells with an active metabolic inflexibility/stress response and progenitor/stem-cell-like features (Figure 2, A–D, and Supplemental Videos 1 and 2). These cells are distinguished from the  $\beta$  cell cluster MI<sup>-2</sup> by their higher MAFA and

metabolic inflexibility/stress-response activities (Figure 2, A and B). The biphasic distribution of MAFA activity among different  $\beta$  cell clusters is reminiscent of the rodent islet stress response associated with diabetes progression (30). However, this also heralds the onset of metabolic inflexibility (PPAR $\alpha/\gamma$ ; ref. 25) and dedifferentiation (27), as indicated by RFX6 activation (31). MI<sup>-4</sup> and MI<sup>-5</sup> combine weak  $\alpha$ -like features with a muted stress response. However, MI<sup>-5</sup> showed elevated  $\beta$  cell and stemness TF activities compared with MI<sup>-4</sup> (Figure 2, B and D, and Supplemental Videos 1 and 2), consistent with a putative  $\beta$ - to  $\alpha$ -like-cell reprogramming. Although we cannot establish directionality of the transition based on these data alone, functional testing described below supports the  $\beta$ - to  $\alpha$ -like model. In summary, this analysis identified endocrine cell subpopulations enriched in T2D. They include  $\beta$  cells characterized by metabolic inflexibility/stress response and progenitor/stemness signatures (MI<sup>+2</sup>), as well as cells with  $\alpha$ -like features (MI<sup>-4</sup> and MI<sup>-5</sup>), consistent with cell reprogramming.

Different driver proteins elicit distinct T2D cell state transitions. To identify candidate MR proteins that mechanistically regulate the transcriptional identity of cells in T2D-enriched clusters, we assessed the most activated and inactivated proteins (candidate MR proteins), based on metaVIPER analysis of the 2 most rep-

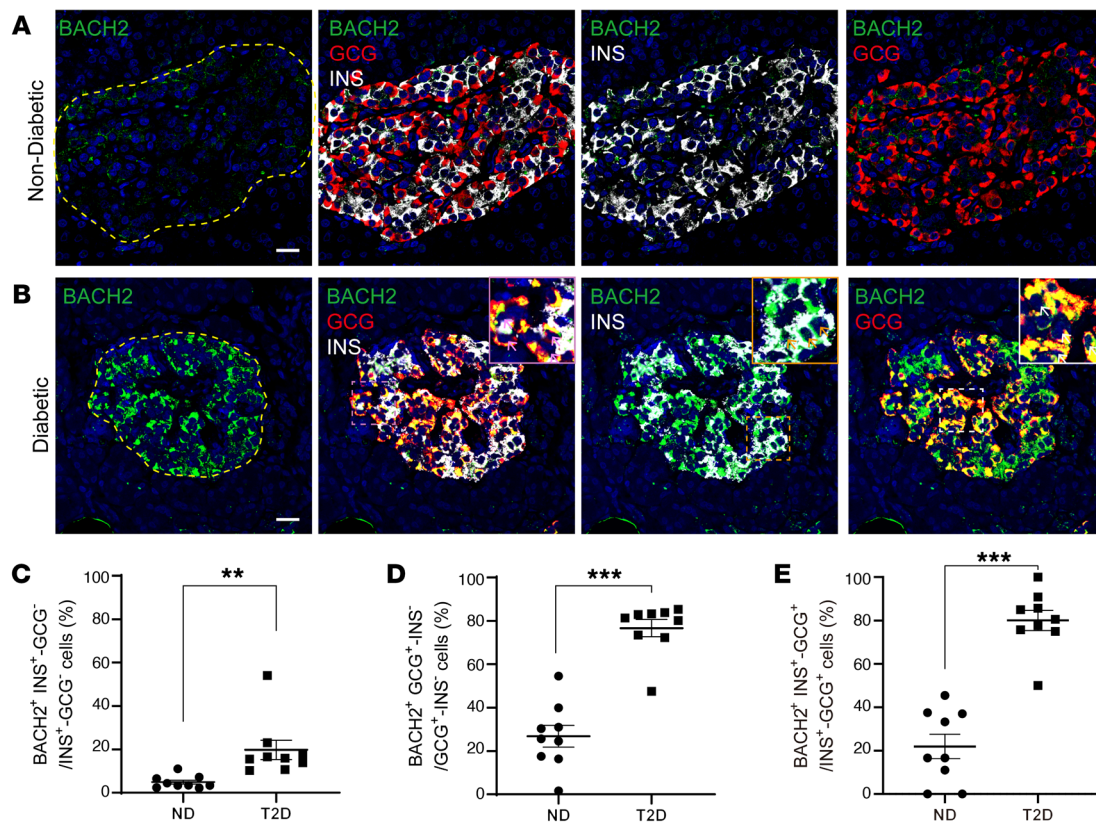


**Figure 6. Single- $\beta$  cell calcium microfluorimetry.** (A) Schematic drawing of the single-cell  $\text{Ca}^{2+}$  imaging procedure. (B–E) Representative traces of  $\text{Ca}^{2+}$  flux measured by Rhod-2 loading in Ad-BFP– (B), Ad-AFF3– (C), Ad-BACH2– (D), or Ad-TCF4–transduced (E) primary human  $\beta$  cells from 3 technical repeats using 2 ND donors (ND7 and ND8). Red arrows indicate the timing of addition of 16.8 mM glucose, and black arrows indicate addition of 40 mM KCl. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$  by ANOVA with Dunnett’s method.

representative T2D-enriched clusters,  $\text{MI}^{+2}$  (T2D- $\beta$  like) and  $\text{MI}^{-5}$  (T2D- $\alpha$  like), versus the healthiest ND  $\beta$  and  $\alpha$  cells ( $\text{MI}^{-1}$  and  $\text{MI}^{+5}$ , respectively; Supplemental Data 1).

Based on their activity in T2D-enriched clusters and association with T2D susceptibility loci (Supplemental Table 3), we selected 15 candidate MR proteins for experimental validation using gain-of-function (GOF) assays in ND islets (Supplemental Table 4). For this purpose, we modified the Perturb-Seq protocol (33, 34) (single-cell GOF sequencing [scGOF-Seq]) (Figure 4A), in which the transduced gene can be mapped to an individual cell using a DNA barcode, allowing its identity and the effect of the gene manipulation to be read out by scRNA-Seq, for further metaVIPER analysis (Supplemental Figure 6 and Supplemental Table 3). We transduced islets with pooled adenoviruses ( $n = 3$  in each pool) to simultaneously screen multiple candidates and test potential epistatic interactions among the cotransduced TFs.

We measured the MR candidates’ ability to reprogram ND cells toward a T2D transcriptional identity, as represented by the  $\text{MI}^{+2}$  ( $\beta$ -like with metabolic inflexibility/stress response and progenitor/stem-like features) and  $\text{MI}^{-5}$  (putatively converted,  $\alpha$ -like) subtypes (T2D- $\beta$ -like and T2D- $\alpha$ -like in Figure 2D). As an integrated metric of the  $\text{MI}^{+2}$  (T2D- $\beta$ -like) state, we used a protein activity signature composed of RFX6, RFX7, FOXO1, PPAR $\alpha$ , PPAR $\gamma$ , RB1, POU5F1, NANOG, and MYCL — henceforth referred to as T2D- $\beta$ -like signature. As a metric of the  $\text{MI}^{-5}$  (T2D- $\alpha$ -like) state, we used a protein activity signature composed of IRX2, ZNHIT1, ZFPL1, PAX6, and DRAP1 — henceforth referred to as T2D- $\alpha$ -like signature (Figure 4). Results are presented as violin plots illustrating the shift from the basal (nontransduced or BFP-transduced) to the scGOF-Seq cell population, and as bar graphs (Figure 4 and Supplemental Figure 7). For some candidate MR combinations, the number of transduced cells was too low to allow statistical evaluation.



**Figure 7. BACH2 immunostaining in human pancreata.** (A and B) Coimmunostaining of ACH2 (green), insulin (white), glucagon (red), and DAPI (blue) in human ND (A) or T2D pancreata (B). INS<sup>+</sup>GCG<sup>+</sup> bihormonal cells expressing BACH2 are indicated by the pink arrows. Orange arrows indicate BACH2-positive β cells and white arrows indicate BACH2-positive α cells. Scale bars: 20 μm. (C–E) Ratio of BACH2-positive cells to INS<sup>+</sup>GCG<sup>-</sup> β cells (C), GCG<sup>+</sup>INS<sup>-</sup> α cells (D), and INS<sup>+</sup>GCG<sup>+</sup> bihormonal cells (E). All data are expressed as mean ± SEM. \*\**P* < 0.01, \*\*\**P* < 0.001 by 2-tailed, unpaired Student's *t* test performed between the 2 groups (*n* = 9 per group).

The strongest emergence of the T2D-β-like signature was observed in cells transduced with the BACH2/TSZH2 and RARB/GAS7/ZNF385D combinations, as well as with ZRANB3/MYTL1 and CUX2/RFX7 (Figure 4, B and C). Next, we asked which TFs can drive conversion to a T2D-enriched α-like state. AFF3 showed the strongest effect, followed by the BACH2/TSHZ2 and ZRANB3/MYTL1 combinations (Figure 4, D and E; see also Supplemental Figure 7). These data suggest that active reprogramming to an α-like-cell state is characteristic of T2D.

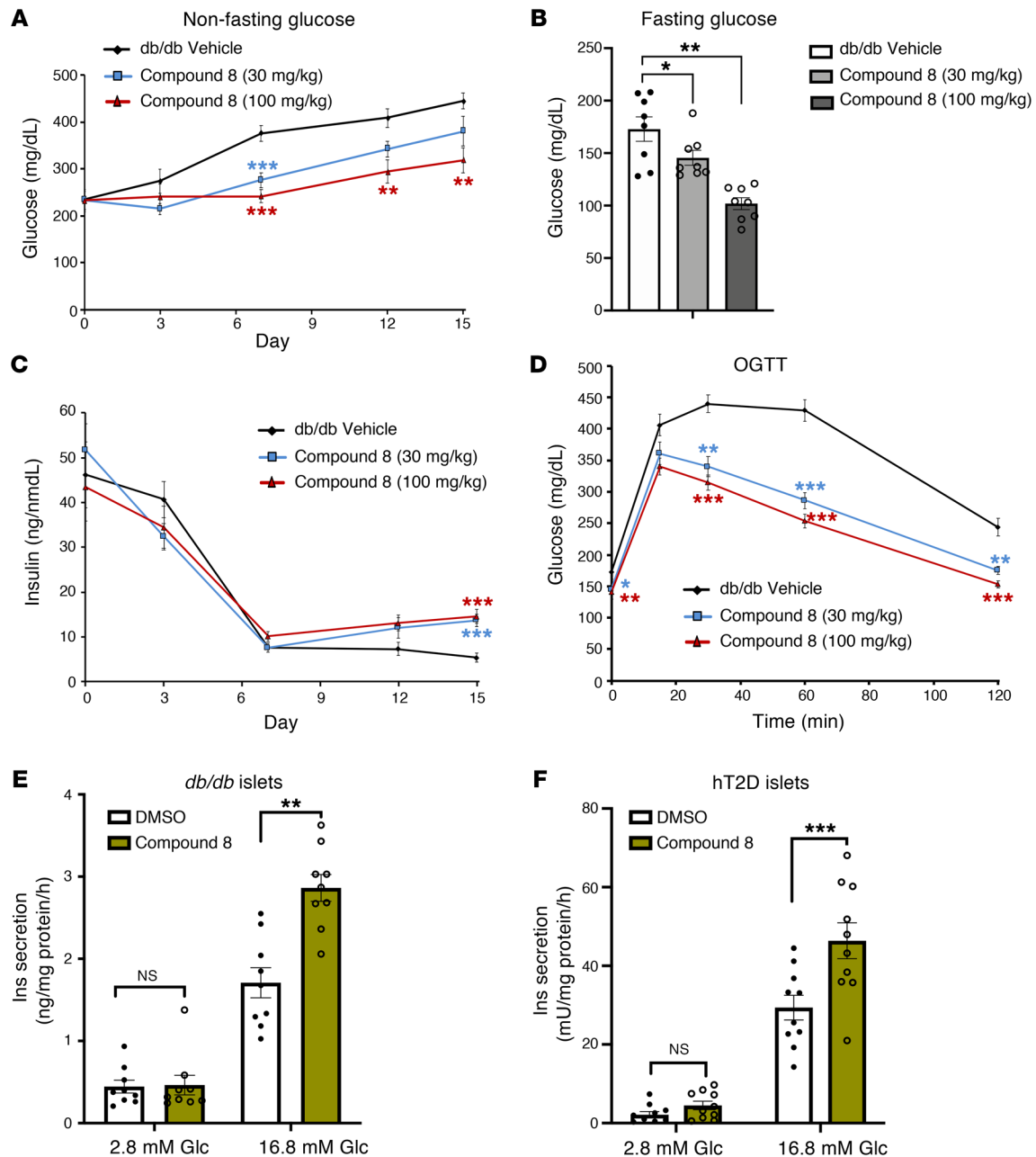
To determine the mechanisms driving cell fate transitions to a T2D-β-like signature, we analyzed TFs and co-TFs whose activity changed in coordinated fashion upon conversion. We found that BACH2, FOXO1, MYTL1, NFATC3, RFX7, and TCF4 were coregulated in all conditions associated with a T2D-β-like endpoint signature (Supplemental Figure 8). While FOXO1 or TCF4 ectopic expression did not drive the T2D-β-like signature, their activity significantly increased upon BACH2- or ZRANB3-dependent conversion, suggesting that these 6 TFs form a hierarchical module whose coordinated activity drives T2D-β-like conversion. This observation is consistent with RNA-Seq analyses of β cell-specific FoxO1-knockout mice, which showed decreased *Bach2* mRNA levels (25, 28). We confirmed the data in a replication experiment using islets from a different donor (Supplemental Figure 9), obtained on different dates with 2 independent adenovirus infection experiments. We

present the 2 experiments separately (Figure 4 for ND5 and Supplemental Figure 9 for ND6; Supplemental Table 4) to circumvent batch effects that may arise from a merged analysis. Each experiment was internally controlled using negative control guide RNAs (gRNAs), and the two are statistically independent.

*Reversibility of diabetic islet cell signature by gene perturbation in T2D islets.* Next, we asked whether targeted, CRISPR-mediated ablation of selected MR activity could reverse disease features in T2D islets (Supplemental Table 4), using a modified Perturb-Seq approach (ref. 34 and Supplemental Figure 10A). We generated individual adenoviruses encoding 4 different gRNAs for each of the top 3 candidate MRs emerging from the GOF assays, including AFF3, BACH2, and CUX2. Following transduction, we measured the extent of cotransduction with each gRNA by decoding gRNA barcodes (Supplemental Table 5) and plotted the effect of different gRNA combinations for each candidate (MR) as SET1 to SET3, where each SET corresponds to a different gRNA for each gene (Supplemental Figure 10B). We measured mRNA levels and computed protein activity signatures to identify sets that reproducibly reprogrammed cells to an ND-like state (Figure 5A, blue color).

AFF3 SET2 decreased *AFF3* mRNA abundance and converted T2D-α-like cells back to an ND-like state but had no effect on the T2D-β-like signature (Figure 5, B and C). BACH2 SET1 decreased *BACH2* mRNA and reversed both T2D-α-like and T2D-β-like sig-





**Figure 8. The effect of BACH2 inhibitor treatment in diabetic mouse models.** (A) Nonfasting blood glucose levels in male *db/db* mice treated with compound 8 or vehicle control throughout 2 weeks of treatment. (B) Overnight-fasted glucose levels in *db/db* mice after 2 weeks of compound 8 treatment. (C) Plasma insulin levels in *db/db* mice treated with compound 8 as in A. (D) Oral glucose-tolerance test (OGTT) in *db/db* mice treated with compound 8 or vehicle control ( $n = 8$  mice per group). (E) Insulin secretion at 2.8 mM or 16.8 mM glucose in islets isolated from *db/db* mice following 24 hours of treatment with 1  $\mu$ M compound 8 ( $n = 10$  per group). (F) Insulin secretion as in E but in hT2D islets ( $n = 10$  per group, T2D8 and T2D9). All data are expressed as mean  $\pm$  SEM. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  by ANOVA with Dunnett's method performed between the 2 groups.

natures toward an ND state. Finally, CUX2 SET2 inhibited CUX2 mRNA and reversed T2D- $\alpha$ -like and T2D- $\beta$ -like signatures. In each instance, the extent of knockdown correlated with the effects (Figure 5, B and C). These assays strongly corroborate the findings from the scGOF studies.

We also separately analyzed restoration of  $\beta$  and  $\alpha$  cell features using *INS* and *GCG* mRNA as an indirect measure (Supplemental Figure 11A), as well as MAFA and IRX2 activities (Supplemental Figure 11, B-E). This subanalysis was consistent with the findings in Figure 5.

**Functional effects of MRs on fate-converted cells.** We next tested whether converting ND  $\beta$  cells (Supplemental Table 4) to a T2D-enriched signature impairs  $\beta$  cell function, as assessed by measuring intracellular  $\text{Ca}^{2+}$  flux in response to glucose in single  $\beta$  cells identified by the *INS* reporter. For these experiments, we selected BACH2 (due to its effects on T2D- $\beta$ -like signatures and genetic link to diabetes susceptibility; ref. 35), AFF3 (due to its effects on T2D  $\alpha$ -like cell conversion), and TCF4 (as a negative control). We gated  $\beta$  cells by cotransducing RIP-zsGreen togeth-

er with adenovirus expressing each candidate TF (36) to monitor  $\text{Ca}^{2+}$  flux (Figure 6A).  $\beta$  Cells transduced with BFP showed a brisk  $\text{Ca}^{2+}$  response to glucose or KCl (Figure 6B). The glucose response was blunted in cells transduced with AFF3 or BACH2, while the KCl response was intact, indicating that AFF3 or BACH2 GOF impairs  $\beta$  cell glucose sensing but do not alter membrane depolarization (Figure 6, C and D). This is consistent with the loss of  $\beta$  cell features observed in scGOF-Seq experiments (Figure 4D). In contrast, TCF4 GOF had no statistically significant effect on  $\text{Ca}^{2+}$  influx compared to Tag-BFP (Figure 6E), consistent with the scGOF-Seq results. These data further validate our approach and strengthen the observations on AFF3 and BACH2.

**Increased BACH2 expression in human T2D pancreata.** To test the clinical relevance of these observations, we performed immunofluorescence analysis of pancreatic samples obtained from ND and T2D patients (Supplemental Table 6) with anti-BACH2 antibodies. BACH2-immunoreactive cells are rare in ND islets, while T2D pancreata show strong BACH2 immunoreactivity both in  $\alpha$  and  $\beta$  cells (Figure 7, A and B). Quantitative analysis of the results showed a 4-fold increase in BACH2-positive  $\beta$  cells and a 3-fold increase in BACH2-positive  $\alpha$  cells in T2D (Figure 7, C and D, and Supplemental Table 6). Also, the percentage of BACH2-positive cells among bihormonal cells ( $\text{INS}^+\text{GCG}^+$ ) increased significantly in T2D islets ( $22\% \pm 5.6\%$  vs.  $80\% \pm 4.6\%$ , ND vs. T2D,  $P < 0.0001$ ; Figure 7E and Supplemental Table 6). These data are consistent with metaVIPER, showing activation of the BACH2 program in association with dedifferentiation and  $\beta$  to  $\alpha$  transition.

**A BACH2 inhibitor improves glycemia in diabetic mice.** Finally, we tested whether inhibition of BACH2 using BACH2 inhibitors affects diabetes in experimental animals. *Bach2* mRNA expression was significantly increased in diabetic mouse  $\beta$  cells compared with controls (Supplemental Figure 12A), consistent with human data (Figure 7). To test the effect of BACH2 inhibition, we administered compound 8 (Supplemental Figure 12B and ref. 37), a dual BACH2/BACH1 inhibitor (Supplemental Figure 12, C and D) to *db/db* mice for 2 weeks and monitored glucose levels throughout. Starting on day 7, we observed a glucose-lowering effect (Figure 8A) as well as decreased fasting glucose in compound 8-treated animals in a dose-dependent manner (Figure 8B). Whereas plasma insulin levels dropped in vehicle-treated mice, consistent with progressive  $\beta$  cell failure, this trend was reversed in compound 8-treated mice, so that insulin levels gradually increased, consistent with improved  $\beta$  cell function upon BACH2 inhibition (Figure 8C). Moreover, oral glucose-tolerance tests showed that compound 8 treatment improved glucose tolerance in *db/db* mice compared with vehicle-treated controls (Figure 8D and Supplemental Figure 12, E and F). It is possible that some of these effects are related to peripheral actions of compound 8. Thus, to analyze the direct effect on islets, we examined the effect of compound 8 on diabetic islets isolated from *db/db* mice or obtained from diabetic organ donors (Supplemental Table 4). In both instances, treatment with compound 8 resulted in a 50% increase in glucose-stimulated insulin secretion and total insulin content, consistent with restoration of  $\beta$  cell function (Figure 8, E and F, and Supplemental Figure 13). These data are consistent with the Perturb-Seq data in human islets (Figure 5) and support the conclusion that BACH2 inhibition can reverse  $\beta$  cell failure.

## Discussion

Clinical progression of T2D is characterized by rapid degradation of  $\beta$  cell function (5) that becomes refractory to treatment, requiring combination therapy and resulting in life-threatening complications. GLP-1-based agents have  $\beta$  cell protective effects (38), but do not appear to significantly delay monotherapy failure, or reverse established  $\beta$  cell failure (39, 40).

Through identification and functional characterization of diabetes-enriched islet cell populations, we provide proof of principle of disease reversibility at the cellular level and validate it by cellular studies in primary  $\beta$  cells of diabetics, immunohistochemistry, and animal experimentation. These observations dovetail with the clinical reversibility of  $\beta$  cell failure in the early phases of T2D (1), or in response to diet (41, 42), but also add a pharmacological dimension to the concept. Compound 8 was originally developed as a BACH1 inhibitor to treat multiple sclerosis but our data showed that it inhibits both BACH1 and BACH2 (Supplemental Figure 12). BACH1 and BACH2 are highly homologous, with significant similarity in the bZIP and BTB domains. However, their expression pattern is different. While BACH1 is expressed ubiquitously, BACH2 is expressed in B cells, T cells, alveolar macrophages, and neural cells (43). Our study suggests that BACH1/2 inhibitors (44, 45) may be repurposed in exploratory animal studies to evaluate their effectiveness against  $\beta$  cell failure. BACH2 expression has been linked to genetic susceptibility to type 1 diabetes (35), and changes to its chromatin structure have been reported in diabetic islets (46). In addition, BACH2 is regulated by the Akt/mTOR pathway and has emerged from a previous analysis of  $\beta$  cell dedifferentiation (28). However, it has not been identified as a mechanistic determinant of T2D cell states. BACH2 binds to the Maf response element (MARE) with small MAFs, and inhibits genes involved in antioxidant responses, which are activated by NRF2 when BACH2 is absent (43). Therefore, inhibition of BACH2 in T2D can increase NRF2-dependent antioxidant response genes, decreasing cell damage and increasing  $\beta$  cell function.

Unbiased analysis of cell transition states identified 3 salient features of diabetic cell populations: (a)  $\alpha$  cell-like features, (b) metabolic inflexibility/stress response, and (c) endocrine progenitor/stem-cell-like features. A striking feature of the MR analysis is the existence of intermediate states, reflecting continuity between the main  $\beta$  and  $\alpha$  cell identity, as detected by identification of  $\alpha$  cell subpopulations with increased  $\beta$  cell signature and vice versa. This can be reconciled with the observation that the latter can be reprogrammed into surrogate  $\beta$  cells (47), as well as with the overall plasticity of endocrine islet cell fate. Within this distribution, healthy  $\beta$  cells (cluster MI<sup>-</sup>) are enriched in ND, while “stressed”  $\beta$  cells (i.e.,  $\beta$  cells with markers of metabolic inflexibility/stress response/progenitor/stemness) are enriched in T2D (cluster MI<sup>+</sup>). In the latter cluster, activation of the 2 nuclear hormone receptors PPAR $\alpha$  and  $\gamma$  may appear counterintuitive, since they oversee different lipid metabolic functions, but is in fact consistent with the observation that, as islets fail, synthetic and oxidative branches of lipid metabolism are coactivated (25). We have suggested that activation of PPAR $\gamma$  is in this context compensatory by directing lipids toward synthesis as opposed to oxidation. This provides a potential explanation for the longstanding

clinical observation that PPAR $\gamma$  agonists have beneficial effects on diabetes prevention and  $\beta$  cell function (48). Nevertheless, given the side effects associated with either PPAR $\gamma$  agonism (49, 50) or PPAR $\alpha$  antagonism (51), we do not believe that these data can be exploited pharmacologically. Rather, they provide unbiased validation of the metaVIPER approach by linking it to mechanisms of disease identified in experimental models (25, 52).

Within “stressed” clusters (MI+), T2D islets are enriched in stressed  $\beta$  cells (MI+<sup>2</sup>), whereas ND islets are enriched in stressed  $\alpha$ -like cells (MI+<sup>3</sup>). In this regard, there appears to be an inverse correlation between GCG mRNA and activation of the IRX2-dependent network on the one hand, and metabolic inflexibility/stress response on the other. Cells in which these features are more pronounced prevail in ND islets, whereas cells with more muted  $\alpha$ -like characteristics are enriched in T2D (MI<sup>-3</sup> vs. MI<sup>-4</sup> and MI<sup>-5</sup>), which is somewhat reminiscent of mixed-feature cells identified previously (53). What can this mean? One potential interpretation is that  $\alpha$ -like cells displaying a more robust stress response are more resilient, and hence capable of reversal to  $\beta$  cells (if indeed this is their origin), whereas those lacking a stress response are transitioning toward a functionally quiescent state. To the extent that these processes are reversible, at least up to a point in the progression of the disease (54), there are likely to be subpools of cells undergoing  $\beta$  to  $\alpha$  and  $\alpha$  to  $\beta$  cell transitions. However, the clinically salient features of decreased insulin production and increased glucagon production make the  $\beta$  to  $\alpha$  cell transition a key disease process.

Our functional studies show that analysis of regulatory networks, reverse engineered from single cells of pancreatic islets from T2D and ND individuals, can pinpoint drivers of different aspects of the cellular pathology, and, consistent with our findings in other contexts (55), favor a model in which a small subset of TFs can drive different aspects of diabetic islet pathology. This regulatory model and associated protein activity signatures differ from those identified thus far and could not have been predicted based on mRNA expression, GWAS, or epigenetic analyses (56). The results that emerge are that BACH2 strongly affects dedifferentiation, whereas AFF3 has the strongest effect on driving the  $\alpha$ -like cell phenotype.

In conclusion, the key findings of this work are the functional assessment of unbiased scRNA-Seq data, their functional validation in a disease-relevant context, and the identification of actionable targets for intervention that address the underlying pathophysiology.

## Methods

**Human islet procurement.** Human ND and T2D islets were from the NIH’s Integrated Islet Distribution Program (IIDP). Upon arrival, human islets were plated at a density of 10,000 islet equivalents per 10-cm nontreated tissue culture dish (Corning, 430591) into 10 mL of islet culture medium (Prodo Labs, PIM(S), cPIM-CS001GMP), supplemented with 5 mL PIM(G) Glutamine/Glutathione (Prodo Labs, PIM-G001GMP), 5% PIM(ABS) Human AB Serum (Prodo Labs, PIM-ABS001GMP), and triple antibiotics, PIM(3 $\times$ ), which includes ciprofloxacin (Corning, 61-277-RE, 10 mg/L), gentamicin (Sigma-Aldrich, G1272, 10 mg/L), and amphotericin B (Omega, FG-70, 2500  $\mu$ g/L). Islets were cultured for no longer than 1 week after arrival and medium was replaced every 2 days. On the day of scRNA-Seq, islets were collected, washed in phosphate-buffered saline (PBS) once, and dis-

persed into single cells by mechanical shaking at 37°C using 0.05% trypsin (Gibco, 25300054). For scGOF-Seq or Perturb-Seq, adenovirus-transduced cells were further gated for PacBlue-positive cells at an excitation wave length of 401 nm, and collected at 452 nm.

**scRNA-Seq library preparation using the Fluidigm C1 800 platform.** Sorted cells were suspended in C1 Cell Suspension Reagent (Fluidigm) and loaded onto each inlet of the C1 high-throughput integrated fluidic circuit (HT IFC). The number of cells captured in each chamber was visualized and noted using a phase contrast microscope. Only chambers with single-cell capture were used for analysis. Cells were lysed, and mRNA reverse transcribed and PCR amplified using C1 Single-cell Auto Prep IFC (Fluidigm, protocol 101-4964). The quality and yield of cDNA were determined by Agilent Bioanalyzer using Agilent High Sensitivity DNA Chip. Libraries for sequencing were prepared using Nextera XT DNA library preparation kit (Illumina FC-131-1096) and sequenced with 50 paired-end cycles on an Illumina HiSeq 2500. Each library pool was sequenced in 2 lanes of the Illumina HiSeq 2500.

**10 $\times$  Genomics platform scRNA-Seq library preparation.** Sorted cells for scGOF-Seq or Perturb-Seq were treated with a Chromium Single Cell 3’ Library and Gel Bead Kit v2 (PN-120237), Chromium Single Cell 3’ Chip Kit v2 (PN-120236), and Chromium i7 Multiplex Kit (PN-120262), and analyzed with a 10 $\times$  Genomics Chromium for Single-Cell Library Preparation Instrument, per the manufacturer’s specifications and 150-bp, paired-end sequenced on a HiSeq 4000 to a depth of 90,000 unique molecular identifiers (UMIs) per cell. UMI counts for each cellular barcode were quantified and used to estimate the number of cells successfully captured and sequenced. The Cell Ranger Single Cell Software suite was used for demultiplexing, barcode processing, alignment, and initial clustering of the raw scRNA-Seq profiles.

We used the Chromium instrument and the Single Cell 3’ Reagent kit (v1) to prepare individually barcoded scRNA-Seq libraries following the manufacturer’s protocol (10 $\times$  Genomics). For QC and to quantify the library concentration we used both the BioAnalyzer (Agilent BioAnalyzer High Sensitivity Kit) and qPCR (Kapa Quantification kit for Illumina Libraries). Sequencing with dual indexing was conducted on an Illumina NextSeq machine using the 150-cycle High Output kit. Sample demultiplexing, barcode processing, and single-cell 3’ gene counting was performed with the Cell Ranger Single Cell Software Suite CR2.0.1. Each droplet partition’s contents were tagged with a UMI — a barcode encoded as the second read of each sequenced read pair. We followed the 10 $\times$  Single Cell 3’ Reagent Kit’s v2 protocol as written, using 12 cycles for cDNA amplification and 12 cycles for sample index PCR. Samples were sequenced to a depth of approximately 400 million reads per sample on a NovaSeq 6000 (R1 = 26 bp, R(i) = 8 bp, R2 = 91 bp).

**Plasmid construction.** We synthesized open reading frames (ORFs) of each scGOF-Seq candidate with a Tag-BFP and an 18-nt unique barcode (Supplemental Table 2) (Qinglan Biotech) and cloned them into the pENTR2b vector using KpnI and EcoRV (AFF3, CUX2, FOXO1, GAS7, TSHZ2, and ZFN385D), BamHI and NotI (BACH2, BNC2, EBF1, RARB, RFX7, and TCF4), SalI and NotI (MYT1L and NFATC3), or KasI and NotI (ZNRANB3). For modified Perturb-Seq, we synthesized the DNA fragment containing the hU6 promoter and gRNA scaffold followed by CMV promoter and *Streptococcus pyogenes* Cas9 ORF with a Tag-BFP. We also inserted NotI and HindIII enzyme sites between Tag-BFP and the bGH poly(A) signal (Genewiz). This synthesized DNA fragment was cloned into the pENTR2b vector using KpnI and EcoRV (pENTR2B-Cas9). gRNA was cloned into the pENTR2B-Cas9 vector

using BsmBI (pENTR2B-gRNA-Cas9), and corresponding guide barcode was sequentially cloned into the pENTR2B-gRNA-Cas9 vector using NotI and HindIII (Supplemental Table 3).

**Adenovirus generation.** Recombinant adenoviruses for scGOF-Seq were generated using the pAd/CMV/V5-DEST Gateway recombination system (Life Technologies) after cloning the full-length cDNA into the pENTR vector. Individual adenoviruses were packaged and amplified in HEK-293A cells, and then pools of 3 P1 adenoviruses as detailed below were expanded into high-titer virus. Pool 1: ZNF385D, RARB, and GAS7; Pool 2: EBF1, FOXO1, and TCF4; Pool 3: BACH2, TSHZ2, and NFATC3; Pool 4: ZRANB3, BNYC2, and MYT1L; and Pool 5: AFF3, RFX7, and CUX2. Adenoviruses were purified by PD-10 column (GE Healthcare, 17085101). Titers were determined by plaque assay (PFU). Each virus pool was transduced into HCT116 cells and expression was analyzed by qPCR. FOXO1, TCF4, NFACT3, RFX7, and AFF3 were amplified individually from P1 adenovirus. Adenovirus for modified Perturb-Seq were generated using the pAd/PL-DEST Gateway recombination system (Life Technologies). Individual adenoviruses were packaged and amplified in HEK-293A cells, and then pools of 4 P1 adenoviruses encoding gRNA to target the same TF were expanded into high-titer virus.

**Adenovirus transduction.** Two hundred to 300 human ND islets were handpicked for each transduction condition and placed in 5 mL round-bottom polystyrene test tubes. Thereafter, islets were washed and incubated with 100  $\mu$ L serum-free islet culture medium containing 1 mM EGTA and transduced at an MOI of 20. After a 5- to 6-hour transduction, 1 mL of complete islet culture medium with 5% PIM(ABS) Human AB Serum (Prodo Labs, PIM-ABS001GMP) was added overnight. Islets were then transferred to 60-mm nontreated tissue culture dishes (Thermo Fisher Scientific, FB0875713A), and the medium was replaced with fresh islet culture medium every 2 days for 7 days for scGOF-Seq or Perturb-Seq experiments, and 3 days for single-cell intracellular calcium microfluorimetry. The 2 ND donors for scGOF-Seq were procured on different dates and adenoviral transduction resulted in different sets of transcription factors being transduced. For this reason, we analyzed the 2 data sets separately in Figure 4 and Supplemental Figure 9 for ND5 and ND6 (Supplemental Table 4), respectively, lest batch effects yield a biased analysis.

**Single-cell intracellular calcium microfluorimetry.** Similarly sized human islets from ND donors were handpicked and transduced with adenovirus expressing each candidate cDNA. The day after transduction, islets were partially dispersed using 0.05% trypsin for 5 minutes at 37°C, and then plated on 35-mm glass-bottom dishes with 10-mm microwells (In vitro Scientific, D35-10-0-N) precoated with fibronectin (Sigma-Aldrich, F1141). Cells were washed with islet media and allowed to rest for 2 additional days. On the third day, each plate was washed with KRBH buffer (10 mM HEPES pH 7.4, 140 mM NaCl, 1.5 mM CaCl<sub>2</sub>, 3.6 mM KCl, 0.5 mM NaH<sub>2</sub>PO<sub>4</sub>, 0.5 mM MgSO<sub>4</sub>, 2 mM NaHCO<sub>3</sub>, and 0.1% BSA) and incubated with 2.8 mM glucose-containing KRBH buffer for 30 minutes, and then loaded in the dark with 5  $\mu$ M Rhod-2, AM (Thermo Fisher Scientific, R1244) in KRBH buffer. Cells were washed and transferred into a perfusion chamber placed in the light path of a Zeiss Axiovert fluorescence microscope, and perfused with low glucose (2.8 mM), high glucose (16.8 mM), or KCl (40 mM) in KRBH buffer.  $\beta$  Cells were excited by a Lambda DG-4 150 watt xenon light source (Sutter), using alternating wavelengths of 340 and

380 nm at 0.5-second intervals, and imaged at 510 nm. For each data set, regions of interest corresponding to the locations of 10 to 20 individual cells were selected and images were recorded using an Axio-Cam camera controlled by Stallion SB.4.1.0 PC software (Intelligent Imaging Innovations). Single-cell intracellular Ca<sup>2+</sup> mobilization data are plotted as a function of time.

**Immunohistochemical and morphometric analyses.** Paraffin sections of pancreas far from the margin of pancreatectomy were collected from our previous research (57). In brief, all cases with partial pancreatectomy performed in Ruijin Hospital between 2013 to 2019 were assessed. Cases with a malignant tumor were excluded and then 9 cases of T2D patients and 9 age- and BMI-matched ND subjects were finally included in this study. Detailed information and clinical characteristics for each patient are listed in Supplemental Table 6.

We fixed and processed tissue for immunohistochemistry as previously described (57). We performed immunofluorescence assays using 5- $\mu$ m-thick paraffin sections obtained and processed as described previously (57). All slides were treated by tissue antigen recovery to improve the fluorescent immune detection of various proteins. Slides were incubated at 4°C overnight with a cocktail of primary antibodies diluted in the blocking reagent for chromogenic and fluorescent immunohistochemical assays. Primary antibodies were prepared at the following dilutions: guinea pig anti-insulin (1:800, DAKO, A056401-2), mouse anti-glucagon (1:400, Abcam, ab10988), and rabbit anti-BACH2 (1:100, Sigma, SAB2108650). Detection was performed using secondary antibodies conjugated to Alexa Fluor 488, 594, and 647 (Jackson ImmunoResearch or Life Technologies). Nuclei were stained with DAPI (Vectashield, Vector Labs) as a marker. Images were captured with an Olympus Microscope or Zeiss LSC 710 confocal microscope. We processed the quantification in a blinded fashion using the cell counter of ImageJ software (NIH) to analyze individual cells located throughout the whole cross section. This analysis scores numbers of positive cells for each marker and calculates the number of cells marked by different signals. To process cell counting, we examined insulin-positive cells (271  $\pm$  63 vs. 207  $\pm$  34 in ND vs. T2D; *P* = not significant) and glucagon-positive cells (124  $\pm$  31 vs. 205  $\pm$  39 in ND vs. T2D; *P* = not significant) per group.

**Quantitative analyses of single-cell gene expression.** For scRNA-Seq profiles of primary donors (GSE98887), raw sequencing reads of single cells were aligned to hg19 reference genome by Bowtie2-2.2.6 (58). Aligned reads were sorted and indexed by samtools-1.2 (59). Count matrices were measured with R packages GenomicFeatures (1.24.5; ref. 60), GenomicAlignments (1.8.4; ref. 60), and TxDb.Hsapiens.UCSC.hg19.knownGene (3.2.2). Cells with fewer than 100,000 reads were filtered out as low-quality cells. For Perturb-Seq (GSE137766) and scGOF-Seq (GSE136887) profiles, UMI matrices generation, including quality filtering, were performed using the standard 10 $\times$  Genomics Cell Ranger pipeline (v2.1.1) with hg38 reference genome. Barcode matrices were generated using the standard scPLATE-Seq data processing pipeline, which is available at <https://github.com/califano-lab/scPLATE-Seq>. Specifically, alignment and mapping were performed by STAR (2.5.2a) with flag `--outFilterMultimapNmax 1` to exclude multimapping reads for accurate barcode quantification (61). QC reports of cells quality filtering, including distributions of transcriptome complexity (number of genes) and sequencing depth (number of reads/UMI) are included in Supplemental Figures 14–16.

**Regulatory networks and transcriptional regulator activity analysis.** Islet-specific regulatory networks were reverse engineered by ARACNe (20) on an individual patient basis. ARACNe was run with 200 bootstrap iterations using 1,813 transcription factors (genes annotated in Gene Ontology molecular function database as GO:0003700, “transcription factor activity” or as GO:0003677, “DNA binding” and GO:0030528, “transcription regulator activity” or as GO:00034677 and GO:0045449, “regulation of transcription”); 969 transcriptional cofactors (a manually curated list, not overlapping with the transcription factor list, built upon genes annotated as GO:0003712, “transcription cofactor activity” or GO:0030528 or GO:0045449); and 3,370 signaling pathway related genes (annotated in GO Biological Process database as GO:0007165 “signal transduction” and in GO cellular component database as GO:0005622, “intracellular”, or GO:0005886, “plasma membrane”). Parameters were set to 0 data processing inequality (DPI) tolerance and mutual information (MI) *P* value (using MI computed by permuting the original data set as null model) threshold of  $1 \times 10^{-8}$ . Protein activity profiles were then generated by metaVIPER by integrating across all 11 donor-specific regulatory networks (18).

**Dimension reduction and clustering analysis.** Dimension reduction was done using both gene expression and metaVIPER-inferred activity profiles. For gene expression, cells were first projected to principal component space using Principal Component Analysis (PCA), further projected to 2D t-SNE space. The t-SNE function in the CRAN R package t-SNE-0.1.3 was used for t-SNE analysis. Specifically,  $1 - r$  (cell-wise Pearson correlation in principal component space) was used as distance matrix. For metaVIPER-inferred activity, cells were projected to 2D t-SNE space. The t-SNE function in CRAN R package t-SNE-0.1.3 was used for t-SNE analysis. Specifically, cell-wise activity dissimilarity was used as distance matrix (17).

Clustering analysis was done using the iterClust iterative clustering analysis framework (23). The iterative clustering analysis was performed using the iterClust function in Bioconductor R package iterClust-1.0.2, at metaVIPER-inferred activity level. In total, 3 iterations were done, separating cell populations at different metabolic stress states and cell types sequentially. Using scRNA-Seq data from cells with more than 0.1 million mapped reads, we first projected single cells onto on 2D space with t-SNE.

**ELISA for HMOX1.** HMOX1 protein induction as a measure of inhibition of the BACH1/2 cellular activity was quantified in cell lines expressing different combinations of these transcription factors. HMOX1 protein levels were determined using the DuoSet IC Human Total HO-1/HMOX-1 kit (R&D Systems, DYC3776). Cells were cultured according to the manufacturer’s protocol and seeded (5,000 cells/well) in 96-well poly-D-lysine-coated plates. Two days after seeding, cells were treated with different concentrations of compound 8 in culture media for 6 hours and then lysed. Lysates were stored at  $-80^{\circ}\text{C}$  overnight. ELISA plates were coated overnight with the capture antibody and the next day lysates were tested in the HMOX1 ELISA assay following the manufacturer’s protocol. Optical density of plates was read at 450 nm on a Cytation microplate reader (BioTek).

**In vivo mouse studies with BACH inhibitors.** Five-week-old male C57BLKS/J *db/db* mice were purchased from The Jackson Laboratory. All mice were fed Lab Diet 5008, maintained under approved Animal Care and Use protocols for Lilly Research Laboratories, and housed under a 12-hour light/12-hour dark cycle. At 6 weeks of age, com-

pounds were administered orally (10 mL/kg) daily for 15 days in 1% hydroxyethylcellulose, 0.25% Tween 80, and 0.05% antifoam. Non-fasted glucose and insulin were assessed throughout the study. On day 13, after an overnight fast, oral glucose-tolerance tests were performed with a bolus of glucose (0.3 gm/kg).

**Glucose-stimulated insulin secretion assays.** Islets were preincubated in KRBH buffer for 1 hour at 2.8 mM glucose, followed by incubation in KRBH at 2.8 or 16.7 mM glucose for 1 hour at  $37^{\circ}\text{C}$ . At the end, we collected islets by centrifugation and assayed the supernatant for insulin secretion or lysates for insulin content by ELISA (Merckodia, 10-1247-01). The insulin levels were normalized to protein concentration.

**Statistics.** Figure 3A and Supplemental Figure 3, C and D quantified the alteration of cell population proportions as opposed to the corresponding control groups, thus *P* values from Fisher’s exact test (FET) were used to evaluate the significance level. For example, to test whether cluster A1 in Figure 3A is enriched for ND or T2D cells, we constructed a  $2 \times 2$  table (ND%, T2D%)  $\times$  (A1 cells, all cells). Such table was then used for FET, and the yielded *P* value of 0.64 indicted that cluster A1 is not significantly enriched for ND or T2D cells. The same FET was performed on other clusters in Figure 3A, as well as cell populations presented in Figure 4, C and E, and Supplemental Figures 7–9. For Figure 8, ANOVA with Dunnett’s method was performed between the 2 groups. A *P* value of less than 0.05 was considered significant.

**Study approval.** Animal studies were performed under approved Animal Care and Use protocols by Lilly Research Laboratories. Human studies were approved by the Institutional Review Board of the Ruijin Hospital affiliated to Shanghai Jiao Tong University School of Medicine and were in accordance with the principles of the Declaration of Helsinki II.

**Data availability.** scRNA-Seq data using the Fluidigm C1 system for all donors have been deposited in the NCBI’s Gene Expression Omnibus (GEO) database (GEO GSE98887). scGOF-Seq and Perturb-Seq data using the 10 $\times$  Genomics Chromium platform have been deposited in the GEO database under accession numbers GSE136887 and GSE137766, respectively.

## Author contributions

J Son and HD contributed equally to designing and performing experiments, to analyzing data, and writing the manuscript. TBF, AME, JLG, SKS, and ZL performed BACH2 inhibitor experiments. J Sun and QW performed immunostaining experiments. AC and DA designed experiments, oversaw research, and wrote the manuscript.

## Acknowledgments

We thank the Genome Technology Center at New York University and the Sulzberger Genome Center at Columbia University for help with scRNA-Seq, and the Human Islet and Adenovirus Core of the AECOM/MSSM DRC for help in generating the adenoviruses. FACS experiments were performed in the DRC Flow Cytometry Core (supported by NIH grant S10OD020056). Human pancreatic islets and/or other resources were provided by the NIDDK-funded Integrated Islet Distribution Program (IIDP) (RRID:SCR\_014387) at City of Hope, NIH grant 2UC4DK098085. This work was supported by NIH grants DK64819, CA197745, and DK63608 (Columbia University Diabetes Research Center).

J Son was supported by Kirschstein-NRSA postdoctoral fellowship F32DK117574. We are grateful to members of the Accili and Califano laboratories for insightful discussions of the data. We thank Travis Morgenstern and Henry Colecraft (Columbia University) for training and equipment for calcium flux imaging. We also thank Andrew Stewart, Adolfo Garcia-Ocaña, and Peng Wang for providing RIP-zsGreen adenovirus.

Address correspondence to: Andrea Califano, Department of Systems Biology, Herbert Irving Comprehensive Cancer Center, Department of Medicine, Department of Biochemistry & Molecular Biophysics, Department of Biomedical Informatics, Columbia University Irving Medical Center, 1130 St. Nicholas Ave,

New York, New York 10032, USA. Phone: 212.851.5140; Email: ac2248@cumc.columbia.edu. Or to: Domenico Accili, Naomi Berrie Diabetes Center, Department of Medicine, Columbia University, 1150 St. Nicholas Ave, New York, New York 10032, USA. Phone: 212.851.5331; Email: da230@cumc.columbia.edu. Or to: Jinsook Son, Naomi Berrie Diabetes Center, Department of Medicine, Columbia University, 1150 St. Nicholas Ave, New York, New York 10032, USA. Phone: 212.851.5333; Email: js4730@cumc.columbia.edu.

HD's present address is: Department of Pharmacy Practice & Science, College of Pharmacy, University of Arizona, Tucson, Arizona, USA.

- Ferrannini E. The stunned beta cell: a brief history. *Cell Metab.* 2010;11(5):349–352.
- Rudenski AS, et al. Natural history of pancreatic islet B-cell function in type 2 diabetes mellitus studied over six years by homeostasis model assessment. *Diabet Med.* 1988;5(1):36–41.
- Dor Y, et al. Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation. *Nature.* 2004;429(6987):41–46.
- Gromada J, et al. The  $\alpha$ -cell in diabetes mellitus. *Nat Rev Endocrinol.* 2018;14(12):694–704.
- Weyer C, et al. The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *J Clin Invest.* 1999;104(6):787–794.
- Kahn SE, et al. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. *N Engl J Med.* 2006;355(23):2427–2443.
- Turner RC, et al. Glycemic control with diet, sulfonylurea, metformin, or insulin in patients with type 2 diabetes mellitus: progressive requirement for multiple therapies (UKPDS 49). UK Prospective Diabetes Study (UKPDS) group. *JAMA.* 1999;281(21):2005–2012.
- Taylor R, et al. Understanding the mechanisms of reversal of type 2 diabetes. *Lancet Diabetes Endocrinol.* 2019;7(9):726–736.
- Halban PA, et al. beta-cell failure in type 2 diabetes: postulated mechanisms and prospects for prevention and treatment. *Diabetes Care.* 2014;37(6):1751–1758.
- Solimena M, et al. Systems biology of the IMIDIA biobank from organ donors and pancreatctomized patients defines a novel transcriptomic signature of islets from individuals with type 2 diabetes. *Diabetologia.* 2018;61(3):641–657.
- Lawlor N, et al. Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. *Genome Res.* 2017;27(2):208–222.
- Wang YJ, et al. Single-cell transcriptomics of the human endocrine pancreas. *Diabetes.* 2016;65(10):3028–3038.
- Baron M, et al. A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Syst.* 2016;3(4):346–360.
- Fang Z, et al. Single-cell heterogeneity analysis and CRISPR screen identify key  $\beta$ -cell-specific disease genes. *Cell Rep.* 2019;26(11):3132–3144.
- Wang YJ, Kaestner KH. Single-cell RNA-seq of the pancreatic islets—a promise not yet fulfilled? *Cell Metab.* 2019;29(3):539–544.
- Elyada E, et al. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. *Cancer Discov.* 2019;9(8):1102–1123.
- Alvarez MJ, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet.* 2016;48(8):838–847.
- Ding H, et al. Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nat Commun.* 2018;9(1):1471.
- Obradovic A, et al. Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages. *Cell.* 2021;184(11):2988–3005.
- Basso K, et al. Reverse engineering of regulatory networks in human B cells. *Nat Genet.* 2005;37(4):382–390.
- Paull EO, et al. A modular master regulator landscape controls cancer transcriptional identity. *Cell.* 2021;184(2):334–351.
- Obradovic A, et al. Tumor-specific cell populations in clear cell renal carcinoma associated with clinical outcome identified using single-cell protein activity inference. *Cell.* 2020;80(21):24.
- Ding H, et al. iterClust: a statistical framework for iterative clustering analysis. *Bioinformatics.* 2018;34(16):2865–2866.
- Nica AC, et al. Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. *Genome Res.* 2013;23(9):1554–1562.
- Kim-Muller JY, et al. Metabolic inflexibility impairs insulin secretion and results in MODY-like diabetes in triple FoxO-deficient mice. *Cell Metab.* 2014;20(4):593–602.
- 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.* 2014;15:620.
- Talchai C, et al. Pancreatic  $\beta$  cell dedifferentiation as a mechanism of diabetic  $\beta$  cell failure. *Cell.* 2012;150(6):1223–1234.
- Kim-Muller JY, et al. Aldehyde dehydrogenase 1a3 defines a subset of failing pancreatic  $\beta$  cells in diabetic mice. *Nat Commun.* 2016;7:12631.
- Vaquerez JM, et al. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet.* 2009;10(4):252–263.
- Kitamura YI, et al. FoxO1 protects against pancreatic beta cell failure through NeuroD and MafA induction. *Cell Metab.* 2005;2(3):153–163.
- Smith SB, et al. Rfx6 directs islet formation and insulin production in mice and humans. *Nature.* 2010;463(7282):775–780.
- Akiba K, et al. A heterozygous protein-truncating RFX6 variant in a family with childhood-onset, pregnancy-associated and adult-onset diabetes. *Diabet Med.* 2020;37(10):1772–1776.
- Adamson B, et al. A multiplexed single-cell CRISPR Screening platform enables systematic dissection of the unfolded protein response. *Cell.* 2016;167(7):1867–1882.
- Dixit A, et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell.* 2016;167(7):1853–1866.
- Onuma H, et al. Variants in the BACH2 and CLE-16A gene might be associated with susceptibility to insulin-triggered type 1 diabetes. *J Diabetes Investig.* 2019;10(6):1447–1453.
- Kenty JH, Melton DA. Testing pancreatic islet function at the single cell level by calcium influx with associated marker expression. *PLoS One.* 2015;10(4):e0122044.
- Kassis JN, et al. 2012. Compounds that modulate oxidative stress. World Intellectual Property Organization patent WO/2012/094580, filed June 1, 2012, and published December 7, 2012. <https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2012094580>.
- Holst JJ. From the incretin concept and the discovery of GLP-1 to today's diabetes therapy. *Front Endocrinol (Lausanne).* 2019;10:260.
- Home PD, et al. Three-year data from 5 HARMONY phase 3 clinical trials of albiglutide in type 2 diabetes mellitus: long-term efficacy with or without rescue therapy. *Diabetes Res Clin Pract.* 2017;131:49–60.
- Jones AG, et al. Markers of  $\beta$ -cell failure predict poor glycemic response to GLP-1 receptor agonist therapy in type 2 diabetes. *Diabetes Care.* 2016;39(2):250–257.
- Taylor R, et al. VLCD for weight loss and remission of type 2 diabetes? - Authors' reply. *Lancet.* 2018;392(10155):1307.
- Taylor R, et al. Remission of human type 2 diabetes requires decrease in liver and pancreas fat content but is dependent upon capacity for beta cell recovery. *Cell Metab.* 2018;28(4):547–556.
- Zhou Y, et al. The bach family of transcription

- factors: a comprehensive review. *Clin Rev Allergy Immunol*. 2016;50(3):345–356.
44. Tsuneyoshi T, et al. S-1-Propanylcysteine augments BACH1 degradation and heme oxygenase 1 expression in a nitric oxide-dependent manner in endothelial cells. *Nitric Oxide*. 2019;84:22–29.
45. Ahuja M, et al. Distinct Nrf2 signaling mechanisms of fumaric acid esters and their role in neuroprotection against 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine-induced experimental Parkinson's-like disease. *J Neurosci*. 2016;36(23):6332–6351.
46. Bysani M, et al. ATAC-seq reveals alterations in open chromatin in pancreatic islets from subjects with type 2 diabetes. *Sci Rep*. 2019;9(1):7785.
47. Furuyama K, et al. Diabetes relief in mice by glucose-sensing insulin-secreting human  $\alpha$ -cells. *Nature*. 2019;567(7746):43–48.
48. Kernan WN, et al. Pioglitazone after ischemic stroke or transient ischemic attack. *N Engl J Med*. 2016;374(14):1321–1331.
49. Qiang L, Accili D. FGF21 and the second coming of PPAR $\gamma$ . *Cell*. 2012;148(3):397–398.
50. Kim-Muller JY, Accili D. Cell biology. Selective insulin sensitizers. *Science*. 2011;331(6024):1529–1531.
51. Abu Aboud O, et al. PPAR $\alpha$  inhibition modulates multiple reprogrammed metabolic pathways in kidney cancer and attenuates tumor growth. *Am J Physiol Cell Physiol*. 2015;308(11):C890–C898.
52. Kim-Muller JY, et al. FoxO1 deacetylation decreases fatty acid oxidation in  $\beta$ -cells and sustains insulin secretion in diabetes. *J Biol Chem*. 2016;291(19):10162–10172.
53. Cinti F, et al. Evidence of  $\beta$ -cell dedifferentiation in human type 2 diabetes. *J Clin Endocrinol Metab*. 2016;101(3):1044–1054.
54. Savage PJ, et al. Diet-induced improvement of abnormalities in insulin and glucagon secretion and in insulin receptor binding in diabetes mellitus. *J Clin Endocrinol Metab*. 1979;48(6):999–1007.
55. Califano A, Alvarez MJ. The recurrent architecture of tumour initiation, progression and drug sensitivity. *Nat Rev Cancer*. 2017;17(2):116–130.
56. Miguel-Escalada I, et al. Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nat Genet*. 2019;51(7):1137–1148.
57. Sun J, et al.  $\beta$ -cell dedifferentiation in patients with T2D with adequate glucose control and non-diabetic chronic pancreatitis. *J Clin Endocrinol Metab*. 2019;104(1):83–94.
58. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–359.
59. Li H, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–2079.
60. Lawrence M, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol*. 2013;9(8):e1003118.
61. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.