The TFIID subunit Taf4 is required for pancreatic beta cell function and identity.

Thomas Kleiber[1,2,3,4,5], Guillaume Davidson[1,2,3,4], Gabrielle Mengus[1,2,3,4], Igor Martianov[1,2,3,4], and Irwin Davidson[1,2,3,4,6,7]#.

1 Institut de Génétique et de Biologie Moléculaire et Cellulaire. BP 163, 67404 Illkirch Cedex, C.U. Strasbourg, France

2 Centre National de la Recherche Scientifique, UMR7104

3 Institut National de la Santé et de la Recherche Médicale, U1258

4 Université de Strasbourg.

5 Danish Cancer Society Research Center, Strandboulevarden 49, 2100 Copenhagen, Danmark

6 Plateforme GenomEast, infrastructure France Génomique.

7 Equipe Labélisée Lingue National contre le Cancer

To whom correspondence should be addressed

E mail: irwin@igbmc.fr

Running Title: Taf4 regulates beta cell function

Key words: diabetes, insulin, glycaemia, trans-differentiation,

The authors declare no potential conflicts of interest.

Abstract.

We selectively inactivated the Taf4 subunit of general transcription factor TFIID in adult murine pancreatic beta cells (BCs). Taf4 inactivation rapidly diminishes expression of critical genes involved in BC function leading to increased glycaemia, lowered blood insulin levels, defective glucose-stimulated insulin secretion and in the longer term reduced BC mass through apoptosis of a subpopulation of BCs. Nevertheless, glycaemia and blood insulin levels are stabilised after 11 weeks with mutant animals showing long term survival. Bulk RNA-seq and ATAC-seq and single cell RNA-seq on isolated Langerhans islets show that Taf4 loss leads to major remodelling of chromatin accessibility and gene expression not only in targeted BCs, but also alpha and delta cells. One week after Taf4-loss cells with mixed BC, alpha and/or delta cell identities were observed as well as a BC population trans-differentiating into alpha cells. Trans-differentiation was associated with the combinatorial action of the Mafb, Mafg, Bptf and Foxp1 transcription factors. Taf4 is therefore essential for BC function and identity and we identify a novel set of factors associated with BC trans-differentiation.

Introduction.

Proper regulation of gene expression by RNA polymerase II (Pol II) requires a complex and dynamic dialogue between transcription factors binding promoters and enhancers, chromatin remodelling and modification complexes and the basal transcription machinery [1-5]. The multi-subunit basal transcription complex TFIID plays a critical role in this communication. TFIID comprises the TATA-box binding protein (TBP) and 13-14 TBP-associated factors (TAFs) and plays a unique role in pre-initiation complex (PIC) formation [6,7]. Recent electron microscopy studies showed that TFIID is organised in three lobes A-C that undergo topological reorganisation during the initial steps of PIC formation as it binds promoter DNA along with TFIIA and TFIIB [7-9]. Lobes A and B are assembled from the histone-fold domain (HFD) containing TAFs (TAF4, 6, 8, 9, 10, 12) and TAF5 that are hence present as two copies in TFIID. Lobe C contains TAF1, TAF2 and TAF7, while lobe A additionally contains the HFD heterodimer TAF11-TAF13 and TBP. TBP makes important interactions with TAF1 and the TAF11-TAF13 heterodimer that are remodelled upon DNA-binding. Moreover, numerous interactions between transcription factors and TAFs have been described that have been suggested to facilitate TFIID promoter binding and PIC formation.

The TAF4-TAF12 heterodimer is crucial for the structural integrity of lobes A and B [10, 11]. In lobe B, a region of the conserved TAF4 C-terminal domain makes contacts with promoter DNA and both TAF4 and TAF12 contact the TFIIA-TBP module suggesting that they promote TBP DNA binding and fix the distance between the TBP binding site and the transcription start site (TSS). The general requirement for TFIID as well as the related SAGA complex for transcription by RNA polymerase II in yeast has recently been shown [12, 13]. Notwithstanding these observations and the potential role of transcript buffering [14], genetic knockout of TAF4, as is the case with several other TAFs, leads to highly specific effects on PIC formation and gene expression, rather than a general effect [11]. This may in part be

explained by redundancy with its paralog TAF4B that also heterodimerizes with TAF12 and integrates into TFIID, thus maintaining TFIID integrity in absence of TAF4 [15]. This is supported by the observation that loss of both Taf4 and Taf4b in MEFs and in mouse ES cells abrogates cell viability [16]. We also used somatic inactivation to address the function of murine Taf4 embryonic and adult murine epidermis or neonatal hepatocytes [11, 15, 17-19] and used germline inactivation to address its role during embryogenesis [16]. In each biological context, Taf4 regulated specific gene expression programs and functions.

In neo-natal hepatocytes, Taf4 is essential for activation of the metabolic gene expression program [11]. The Taf4-Taf12 heterodimer interacts with the nuclear receptor Hnf4a to promote its binding to conserved and functional cis-regulatory elements located close to the TSS. The stable binding of Hnf4a to these elements is thus dependent on concomitant PIC formation revealing that the Hnf4a-Taf4-Taf12 axis coordinates PIC formation with Hnf4a occupation of its cognate regulatory elements. The function of Hnf4a is not however limited to hepatocytes, but it plays important functions in the development and homeostasis of pancreatic beta cells (BCs) [20-26]. To evaluate whether Taf4 is a critical co-factor for Hnf4a in BCs analogous to what we observed in neo-natal hepatocytes, and to more generally assess its function in these cells, we used Insulin-Cre-ER^{T2} transgenics to inactivate Taf4 in murine adult BCs. We show that Taf4 regulates an extensive gene expression program including critical components of the insulin signalling pathway. Taf4 loss also impacts BC identity and we characterize several intermediate and trans-differentiating BC populations defining novel combinatorial sets of transcription factors associated with the transitions between them.

Results.

Inactivation of Taf4 in adult pancreatic beta cells leads to hyperglycaemia.

To inactivate Taf4 selectively in adult pancreatic BCs, mice with the previously described floxed Taf4 alleles [11, 15] were crossed with *Ins1*::Cre-ER^{T2} transgenic drivers to generate *Ins1*::Cre-ER^{T2}::*Taf4*$^{lox/lox}$, *Taf4*$^{lox/+}$ or *Taf4*$^{+/+}$ animals. At 11-13 weeks of age, adult animals were injected with Tamoxifen (Tam) on 3 consecutive days to generate *Taf4*$^{b-/-}$ mice lacking Taf4 in BCs and control animals from the other genotypes that maintained Taf4 expression. Immunostaining of Langerhans islets revealed loss of Taf4 expression already 1 week after injection in *Taf4*$^{b-/-}$ animals, whereas it was clearly visible in Tam injected *Ins1*::Cre-ER^{T2}::*Taf4*$^{+/+}$ controls (Fig. 1A). We noted also the absence of Taf4 in the nuclei of the acinar cells surrounding the islets in both the control and *Taf4*$^{b-/-}$ mutant animals suggesting that these cells normally did not express Taf4 (Fig. 1A and B). No Taf4 expression was detected in Langerhans islets of the *Taf4*$^{b-/-}$ animals 3 and 10 weeks after injection, but strikingly its expression was detected in a sub-population of BCs after 34 weeks (Fig. 1A and B). After 55 weeks, islets were still composed of a mix of Taf4 expressing and non-expressing BCs, showing that even after this prolonged period the Taf4 null population was not completely replaced by Taf4 expressing cells. In each case, cells that lacked or regained Taf4 expression showed cytoplasmic insulin labelling indicating they were insulin producing BCs. This observation suggested that over time the BC population was partially reconstituted from cells that either had not undergone recombination of the Taf4 alleles, from a precursor cell population or trans-differentiation of other Langerhans islet cell populations into BCs. These data showed that Taf4 can be efficiently inactivated in BCs and that Taf4 mutant BCs were replaced only partially over the long term by a novel population expressing Taf4.

We examined evolution of different physiological parameters over a period of 20 weeks following Taf4 inactivation. Compared to Taf4-expressing littermates, the *Taf4*$^{b-/-}$

animals showed progressive reduction in body weight (Fig. 2A). Measurement of blood glucose levels indicated, that while no change was seen 1 week after Taf4 loss, a potent increase in glycaemia was observed after 3 weeks (Fig. 2B). Hyperglycaemia persisted over the 20 weeks, although levels were mildly but reproducibly reduced after 10 weeks compared to what was observed at 3 and 6 weeks. In agreement with this, circulating insulin levels were also reduced between weeks 1-10, but recovered somewhat at later times (Fig. 2C). Measurement of ATP levels, a regulator of insulin secretion, in wild-type and mutant islets showed a strong decrease 3 weeks after Tam injection in the *Taf4*$^{b-/-}$ animals, but a progressive recovery to almost normal levels between 11 and 24 weeks (Fig. 2D). Loss of Taf4 in BCs therefore resulted in increased glycaemia and lowered insulin levels, that were partially, but not fully, restored after 10 weeks. As a result, the *Taf4*$^{b-/-}$ animals did not show a total loss of BC function and no premature death with lifespans comparable to their Taf4 expressing littermates.

Defective glucose tolerance and glucose-stimulated insulin secretion upon Taf4 inactivation.

To assess whether Taf4 loss lead to defective glucose-stimulated insulin secretion (GSIS) and glucose tolerance, *Taf4*$^{b-/-}$ animals and their Taf4-expressing littermates were injected with glucose 3 weeks after Tam injection and the resulting glycaemia and circulating insulin levels were monitored. Under basal conditions the *Taf4*$^{b-/-}$ animals showed the increased glycaemia seen after 3 weeks that was further increased 30 minutes after glucose administration (Fig. 2E). Compared to their Taf4-expressing littermates, glycaemia in the *Taf4*$^{b-/-}$ animals increased to higher levels that persisted over 90 minutes before returning to basal values by 120 minutes. In contrast, control animals showed a lower increase in glycaemia that returned to basal values after 90 minutes. Similarly, glucose administration in

Taf4$^{b-/-}$ animals failed to increase insulin levels as seen in control animals (Fig. 2F). Taf4 inactivation therefore led to diminished GSIS and altered glucose tolerance.

Taf4 inactivation leads to defective insulin maturation and apoptosis of a subset of beta cells.

We examined BC fate following loss of Taf4. Histological examination of pancreatic sections allowed quantification of the number and size of Langerhans islets. At 3 weeks, the number of islets was comparable in *Taf4*$^{b-/-}$ and control animals, however, by 11 weeks their numbers were significantly lower in the knockouts (Fig. 3A). We additionally crossed the *Ins1*::CreER^{T2}::*Taf4*$^{lox/lox}$ animals with mice bearing a Lox-Stop-Lox cassette driving a GFP reporter in the *Rosa26* locus such that Tam administration activated GFP expression in BCs that were subsequently isolated from islets by FACS. After 11 weeks, a 7-fold reduction in BCs was observed following Taf4 loss (Fig. 3A). Consistent with the loss of BCs, at 3 weeks, and more strikingly at 11 weeks, the % of islets of smaller size were increased in the *Taf4*$^{b-/-}$ animals (Fig. 3B and C). Staining of islets by TUNEL at 3 weeks in both *Taf4*$^{b-/-}$ and control islets showed few apoptotic cells at the periphery and only rare apoptotic BCs within the islet (Fig. 3D). After 11 weeks, numerous apoptotic BCs were seen within the islets. Quantification indicated that by 3 or 11 weeks after Tam injection 100% of the *Taf4*$^{b-/-}$ islets comprised apoptotic BCs within the islet, but only 30% of the control animals (Fig. 3E). Taf4 loss therefore led to a progressive loss of BCs through apoptosis that began around 3 weeks, but that was strongly amplified by 11 weeks.

Examination of BC histology by electron microscopy showed numerous densely stained mature insulin granules in the cytoplasm in control BCs (Fig. S1A). However, after 2-3 weeks, the number of mature insulin granules was strongly diminished and replaced by lighter staining pro-insulin granules (Fig. S1B-C). Furthermore, the Golgi apparatus was visibly enlarged in mutant BCs. After 11 weeks two BC populations could be discerned.

Around 70% of BCs displayed only few mature insulin granules in their cytoplasm (Fig. S1D), whereas the other 30% displayed more numerous granules (Fig. S1E). Increased extracellular space between the mutant BCs was observed compared to the tight cell-cell contacts of control BCs (Fig. S1F).

Together the above results showed that loss of Taf4 in BCs led to both acute and longer-term effects. Within 1-3 weeks, diminished insulin maturation was observed along with defective glucose tolerance and GSIS. As a consequence, mutant animals showed persistent hyperglycemia. Over the next weeks, the number of BCs was reduced due to apoptosis, but not all BCs were lost and surviving BCs with mature insulin granules or predominantly pro-insulin granules were observed at later times accounting for stabilised blood glucose and insulin levels and long-term survival. Moreover, by 34-53 weeks, islets were partially repopulated by Taf4-expressing BCs. Thus, after an initial acute effect of Taf4 loss on the existing BC population, the surviving BCs adapted and partially restored their insulin signalling capacity and at later times were partially replaced by a new Taf4-expressing BC population.

Taf4 regulated gene expression programs in beta cells.

To investigate how Taf4 knockout in BCs impacted gene expression as the basis for the observed phenotypes, we performed RNA-seq from WT islets and from islets, 1, 3 and 5 weeks after Tam injection (Supplemental Dataset 1). Already 1 week after Tam injection, more than 600 genes were down-regulated and 390 up-regulated (Fig. S2A and B). At 3 and 5 weeks, the number of de-regulated genes strongly increased such that by 5 weeks more than 2500 genes were down-regulated and 900 up-regulated (Fig. S2A and B). More than 300 genes were down-regulated under all conditions together with a large overlap of more than 800 between weeks 3 and 5. At 1 week, several genes critical for BC function such as *Slc2a2*, *Trpm5*, *Ins1* and genes involved in calcium signalling and cell contact were down regulated

(Fig. S2C and Fig. 4A). In contrast, up-regulated genes were strongly enriched in inflammatory/stress response including numerous chemokines and *Reg3a*, *Reg3b* and *Reg3g*, genes with important roles in islet stress response and regeneration (Fig. S2C and Fig. 4A). Up regulated genes were also strongly enriched for oxidative phosphorylation and DNA metabolism including repair, replication and general transcription factors. Expression of the alpha and delta cell makers *Gcg* and *Sst* was also rapidly up-regulated and then attenuated with time (Fig. S2C).

While overall an increasing number of genes was deregulated with time, we identified sets of genes whose normal expression was partially restored. Expression of a subset of genes strongly down-regulated after 1 week was progressively restored to more normal levels after 3 and 5 weeks (Fig. S2D and Fig. 4B). Similar results were seen with up-regulated genes the most striking example of which are the inflammatory/stress response genes up-regulated after 1 week but restored to almost normal after 3 and 5 weeks.

After 5 weeks, down regulated genes remained enriched in those involved in cell adhesion and calcium signalling, many of which were already de-regulated at 1 and/or 3 weeks (Fig. 4B). Similarly, genes involved in oxidative phosphorylation and DNA metabolism remained up-regulated along with novel pathways such as the unfolded protein response (UPR). In addition, genes involved in diverse aspects of RNA metabolism were also up-regulated along with a collection of basal transcription factors and apoptotic genes in keeping with the increased apoptosis described above. Down-regulated genes additionally comprised components of multiple signalling pathways such as FGF, SMAD/TGFB and Notch (Supplemental Dataset 1).

Up-regulation of genes involved in chromatin organisation, RNA metabolic functions and transcription were already observed after 3 weeks (Fig. S2D). These observations suggested that the initial stress response immediately after Taf4 inactivation was followed by

chromatin reorganisation and compensation for loss of *Taf4* by up-regulation of *Taf4b* and *Taf12*. Indeed, *Taf4b* expression was much lower than that of *Taf4* in wild-type islets. Note that as previously described, the deletion of the C-terminal Taf4 exons had little effect on *Taf4* mRNA levels and thus *Taf4* did not appear as a downregulated transcript.

We assessed the effect of Taf4 inactivation on chromatin accessibility using ATAC-seq on isolated islets. After 1 week, extensive changes in chromatin accessibility were observed with a global increase in accessibility upon Taf4 loss. When focussing on the proximal promoter, the number of sites enriched in the WT was higher than those enriched in the mutant (Fig. S3A). A similar result was seen after 5 weeks, where again more promoter proximal sites were enriched in the WT than in the mutant (Fig. S3B). To identify transcription factors whose occupancy was modified, we carried out *in silico* footprinting analyses of the differentially accessible regions at the proximal promoter [27]. Binding of Hox-domain containing transcription factors including critical BC factors such as Nkx6-1 and Hnf1b were enriched in the WT, whereas factors like Yy2, Arnt2, Bhlhe41 and Mycn were enriched 1 week after Taf4 loss (Fig. S3C). Similarly, after 5 weeks, enriched binding of several factors including Yy2, Nrf1 was observed (Fig. S3D).

Together the ATAC-seq and RNA-seq data indicated that Taf4 loss led to a reduction in chromatin accessibility at proximal promoters, modified transcription factor binding with reduced binding of critical BC factors together with reduced expression of a collection of genes essential for BC function and enhanced binding of a new set of transcription factors.

Taf4 inactivation results in loss of beta cell identity and trans-differentiation into alpha cells.

While the above data provided an overall view of gene expression changes in islets, they did not assess whether Taf4 loss affected only BC gene expression or if it impacted gene

expression in other cell types as suggested by increased *Gcg* and *Sst* expression. To address this, we performed single cell (sc)RNA-seq from purified dissociated Langerhans islets from control Taf4-expressing mice and *Taf4*$^{b-/-}$ mice 1 and 5 weeks after Tam injection.

Analyses of around 12000 cells from Taf4-expressing control islets identified the major endocrine populations, with 4 BC clusters expressing high levels of *Ins1* and *Ins2* (clusters 0-2 and 4 in Fig. 5A). Cluster 3 represented *Gcg*-expressing alpha cells, cluster 6 *Sst*-expressing delta cells and cluster 5 an acinar population (Fig. 5A). The 4 BC sub-populations could be distinguished by their expression signatures with a strong ER stress signature in cluster 2, while cluster 0 represented potential secretory cells with high expression of *Slc2a2*, *Abcc8* and *G6pc2* and lower *Ins1*/*2* expression that was more prominent in clusters 1 and 4 (Fig. S4).

We analyzed 10 666 cells 5 weeks after Taf4 inactivation distinguishing again 4 BC populations with high *Ins2* expression (clusters 2-4 and 6, Fig. 5B), but reduced *Ins1* expression as noted in the bulk RNA-seq. Cluster 1 corresponded to *Sst*-expressing delta cells, while clusters 0 and 5 corresponded to *Gcg*-expressing alpha cells. We aggregated the WT and week 5 data sets allowing a global comparison of all cell populations under the 2 conditions (Fig. 5C, D and Fig. S5) clearly distinguishing BC populations from WT and mutant conditions based on high *Ins1* expression in WT (clusters 0, 5, 6 and 9) from the *Ins2* high/*Ins1* low mutant populations (clusters 2, 7, 8 and 11). The WT and mutant cell populations segregated in a UMAP representation (Fig. 5D) where WT and mutant BCs were distinguished by strongly reduced expression of the BC markers *Ucn3* and *Slc2a2* as seen in bulk RNA-seq (Fig. S6A). High expression of stress markers *Ddit3* and *Herpud1* characterized WT BCs in cluster 5 and mutant clusters 2 and 8 that together accounted for a large proportion of mutant BCs (Figs. S5 and S6A). The aggregate also revealed 3 distinct alpha cell populations, cluster 3 from WT and clusters 4 and 10 from the mutant with a strong

stress response signature in cluster 4 (Fig. S5 and S6A). Delta cells from the two conditions adopted a more homogenous population (cluster 3).

The important changes in gene expression following Taf4 inactivation in BCs allowed clear distinction between the WT and mutant BC populations and segregation of alpha cell populations from WT and mutant with 2 populations in the mutant. Delta cell gene expression on the other hand was less affected with little difference in the populations of the two conditions.

We analyzed 2295 cells 1 week after Taf4 inactivation. Contrary to WT and week 5 where 4 BC populations were distinguished, at week 1, 2 populations showed highest *Ins2* expression (clusters 0 and 5, Fig. 6A). Cluster 5, the most strongly differentiated BC population, had highest expression of *Ins1* and *Ins2* and low levels of *Gcg* and *Sst*. *Ins2* was also strongly expressed in clusters 1 and 2 indicating their BC identity. *Ins1* expression was reduced in cluster 0, while it remained higher in cluster 5. More importantly, cluster 0 showed significant *Gcg* and *Sst* expression, although lower than in the bone-fide alpha and delta cell populations, clusters 3 and 7, respectively, that segregated from the BC populations in the UMAP and tSNE representations (Fig. 6B-C and Fig. S6B). Moreover, cluster 4 expressed intermediate levels of *Ins2*, *Gcg* and *Sst* and segregated between the BC, alpha and delta cell populations on the UMAP and between the BC and alpha populations in the tSNE. The small cluster 6 also showed characteristics of a mixed identity population expressing *Ins1/2* and intermediate levels of *Gcg*.

Upon Taf4 inactivation, a population with mixed identities (cluster 0) expressing *Ins1* and *Ins2*, but also *Gcg* or *Sst* and sometimes all three was seen along with intermediate cells in clusters 1 and 4, with cluster 4 cells trans-differentiating to an alpha and perhaps also delta cell identity (Fig. 6C and Fig. S6B). This model was further supported by the ontology of the differentially regulated genes (Supplemental Dataset 2). Cluster 5 was marked by a small

number of differentiated BC markers, cluster 2 showed translational activity, ER stress and high oxidative phosphorylation signatures, whereas in contrast, signatures for transcription regulation, chromatin modification and RNA metabolism were hallmarks of clusters 1 and 4 consistent with the idea that these cells were undergoing transcriptional reprogramming (see also Fig. S6C).

We used the SCENIC program [28] to identify transcription regulatory networks active in the different cell populations. The SCENIC-based tSNE (Fig. 7A) that groups cells based on their regulon activities differed from the expression based tSNE (Fig. 6B). With SCENIC, the BCs clusters 2 and 5 grouped together, intermediate clusters 1 and 4 grouped closely, while cluster 0 grouped close to alpha cells (Fig. 7A). Consistent with the idea that cluster 5 were differentiated BCs, they showed higher activity of the Mafa, Nkx6-1 and Neurod1 regulons (Fig. 7B). These regulons were active in cell sub-populations of clusters 0 and 1, but absent from alpha and delta cells. Cluster 2 was marked by activity of Ddit3 and Atf5, that promotes BC cell survival under stress [29], consistent with the elevated stress response in this population (Fig. 7C). In contrast, Mafb regulon activity was seen in alpha cells and in clusters 0, 1 and 4 confirming they represented abnormal populations with mixed identities, with some cells showing both Mafa and Mafb activities (Fig. 7B). Transition from cluster 1 to cluster 4 was marked by a switch away from BC identity factors to the alpha cell determinant Mafb.

Strikingly, clusters 0, 1 and 4 showed activity of the Arnt2, Bhlhe41, Myc and Nrf1 regulons corresponding to enhanced binding predicted from ATAC-seq footprinting (Fig. 7D). The abnormal cell populations were therefore marked by activity of regulons not normally active in BCs. Clusters 1 and 4 were specifically marked by high activity of the small Maf inhibitor Mafg and the Bptf subunit of the NuRF complex coherent with the chromatin remodeling ontology signature of this cluster (Fig. 7E). Moreover, the Foxp1 regulon was active in differentiated alpha and delta cells as well as scattered cells in clusters 0 and 1, but

labeled most cells of cluster 4 (Fig. 7E). Quantitatively, Mafb, Myc, Arnt2, Mafg, Bptf and Foxp1 regulons showed highest combinatorial activity in cluster 4 again consistent with a major transcriptional reprogramming associated with trans-differentiation (Fig. 7F). We also noted that trans-differentiation did not appear to involve transition via a precursor state as low *Ngn3* (Neurog3) expression was seen in the delta cell population, but was not significantly expressed in clusters 1 and 4 (Fig. 6A).

Together these analyses showed that within 1 week of Taf4 inactivation, several abnormal cell populations appeared, cells with mixed identities, an intermediate sub-population with characteristics of de-differentiation and a population trans-differentiating into alpha cells (Fig. 7G). The abnormal cell populations were associated with reduced activity of BC-identity factors and gain of Mafb activity as well as of several other transcription factors not normally active in BCs. The trans-differentiating population was specifically marked by Mafg and Bptf activity and acquisition of Foxp1 activity, a novel and specific alpha/delta cell marker. As a result, by 5 weeks two alpha cell populations were observed potentially corresponding to the normal and trans-differentiated populations. However, at 5 weeks, no populations with mixed or obviously trans-differentiating characteristics were observed suggesting that trans-differentiation took place immediately after Taf4 loss and was terminated by 5 weeks.

To verify the idea that a sub-population of BCs trans-differentiate into alpha cells, we performed immunostaining with antibodies against Ins and Gcg (Fig. S7). In WT islets, Gcg staining was limited to alpha cells surrounding the BCs within the islet. Strikingly, by 3 weeks after Taf4 inactivation, Gcg stained cells were observed both around and within the islets amongst the Ins-labelled BCs, a phenomenon that was seen at all subsequent stages until 55 weeks after injection. Such cells were not co-stained with Ins, but were labelled uniquely by Gcg. Together with the results of the sc-RNA-seq, these observations support the idea that

the novel population of Gcg labelled cells within the islets correspond to trans-differentiated BCs.

Discussion.

Taf4 is essential for normal beta cell function.

We show that Taf4 inactivation in adult murine BCs led to increased glycaemia due to defective insulin signalling and secretion, a consequence of an immediate and major impact on gene expression in BCs accompanied by trans-differentiation of BCs into alpha cells and at later times apoptosis of part of the BC population. Despite this, the surviving BCs allowed stabilization of glycaemia and longer-term survival of the animals.

Immunostaining showed a rapid and efficient loss of Taf4 expression in BCs by 1 week after Tam injection. At 34 weeks, Taf4-expressing BCs were again detected, but their proportion did not increase by 55 weeks. While it is possible that Taf4-expressing BCs detected at later times arose from rare non-recombined BCs, no Taf4 expressing BCs were observed upon staining of multiple islets at regular intervals from 3-24 weeks. Furthermore, the number of Taf4-expressing BCs did not increase between 34 and 55 weeks arguing against proliferation of non-recombined BCs. Alternatively, Taf4-expressing BCs may arise from trans-differentiation of alpha cells or other islet populations into BCs as has been previously described, for example under conditions of reduced BC-mass [30-33], one of the consequences of Taf4 inactivation. Notwithstanding the mechanisms involved, these data showed that despite reduced BC mass through increased apoptosis and an ensuing elevated glycaemia, the BC population was only partially and slowly replaced by newly generated BCs.

Taf4 inactivation led to a long-term reduction in BC mass and reduced islet size and by 5 weeks, a majority of BCs were filled with pro-insulin granules rather than mature insulin granules. Despite these observations and long-term reduced expression of key BC function

genes (*Slc2a2*, *Ins1*, *G6pc2, Trpm5*), glycaemia was stabilized and even mildly reduced by 10 weeks, well before the first Taf4-expressing BCs were observed at 34 weeks. Indeed, bulk RNA-seq identified sets of genes whose expression was partially corrected at 5 weeks along with up-regulation of *Taf4b* and its heterodimerization partner *Taf12,* a potential compensatory response to Taf4 loss. Moreover, only a sub-population BCs underwent apoptosis, while others displayed long-term survival. This is perhaps related to BC heterogeneity, as WT BCs with high and low levels of ER stress markers were identified consistent with previous reports [34-39]. ER stress is modulated by insulin demand and has been associated with BC differentiation, survival and proliferation. Perhaps differences in ER-stress amongst BC populations contributed to their [39,40]. Furthermore, while *Ins1* expression was diminished, *Ins2* remained strongly expressed in mutant BCs and although insulin maturation was diminished it was not completely lost. In addition, proinsulin is secreted and retains 5-10% of the metabolic activity of mature insulin. Thus, stabilization of an albeit elevated glycaemia compatible with long-term survival of the Taf4$^{b-/-}$ animals can be can be accounted for by persistence of partially functional BCs for many weeks before the appearance of new Taf4 expressing BCs.

A model for beta cell trans-differentiation.

Bulk RNA-seq and ATAC-seq showed that Taf4 inactivation had a potent impact on chromatin accessibility and gene expression with diminished binding of factors important for BC identify. In contrast, increased binding of several transcription factors was predicted from ATAC-seq and some of the corresponding regulons were activated in the abnormal cell populations. Cells with mixed identities displayed regulon activity of BC factors, of Mafb, that in mouse is an alpha cell determinant, and several others such as Myc and Nrf1. These cells did not express *Arx* and therefore displayed an incomplete switch in identity in line with the intermediate levels of *Gcg* compared to fully differentiated alpha cells. Previous studies

on normal mouse or human pancreas by immune-staining or sc-RNA-seq detected only rare cells simultaneously expressing two of the islet hormones [34-38,40]. Such cells have however been detected in pathological situations like type 1 and type 2 diabetes and in mice following loss of Arx and Dnmt1 [34]. The appearance of a sizable population of mixed identity cells was therefore a consequence of Taf4 inactivation and potentially the resulting activation of novel regulons.

We detected two intermediate cell populations. Cells of cluster 1 maintained regulon activity of BC factors, but acquired activity of Mafb, like the mixed identity cells, but showed lower expression of *Gcg*. Cluster 4 cells displayed decreased expression and regulon activity of BC factors while maintaining Mafb activity and higher *Gcg* and *Sst* expression. These two intermediate populations were however set apart from cluster 0 by high activity of the Mafg and Bptf regulons not seen in the other populations. Furthermore, trans-differentiating cells were strongly marked by Foxp1 regulon activity consistent with the reported essential role of Foxp factors in post-natal alpha cells [41]. Inhibition of (s)Maf proteins in BCs led to increased Mafa activity and BC function in agreement with their suppression by activation of Mafg in the trans-differentiating cells [42,43]. Small Mafs also form heterodimers with Nrf1 to regulate anti-oxidant genes [44]. The Nrf1 regulon is activated in cells of clusters 1 and 4 and hence can heterodimerize with Mafg. Nevertheless, the molecular details of Mafg function in regulating Mafa, Mafb or Nrf1 in pancreas remain to be fully addressed.

We propose a model for BC trans-differentiation involving a progressive loss of BC identity associated with gain of regulon activity for Myc, Arnt2, Bhlhe41, and Nrf1. A population of cells further undergo chromatin remodelling and transcriptional reprogramming through the combinatorial action of Bptf and Mafg and finally acquisition of Foxp1 activity promotes trans-differentiation into alpha cells. The proximity of the intermediate cells with the mature delta cell population on the UMAP and the up-regulated expression of *Sst* further

suggest trans-differentiation into delta cells. By 5 weeks, trans-differentiation is completed as no comparable intermediate populations were observed. The appearance of a trans-differentiated alpha cells was further attested by two distinguishable alpha cells populations after 5 weeks and by the appearance of Gcg expressing cells intermixed amongst BCs. The fact that no Taf4 expression was seen in islets at this time, further supports the idea that these Gcg-expressing cells are derived from Taf4-null BCs.

BC trans-differentiation into alpha cells and vice versa has been described in diabetes, in genetically modified mice and after various pathological or chemical treatments. The molecular mechanisms involved particularly at the transcriptional level are however not fully understood [30, 32, 45-47]. In type 2 diabetes, hyperglycaemia seems to induce BC dedifferentiation with diminished expression BC function genes and key BC transcription factors, but expression of BC forbidden genes and it may involve reversion to a Ngn3-expressing precursor state as a prerequisite to trans-differentiation to alpha or delta cells [33, 40, 48]. BC trans-differentiation upon loss of Taf4 occurred rapidly in absence of chronic hyperglycaemia and was characterized by loss of BC markers, but not by up-regulated expression of forbidden genes and it did not require passage via a progenitor state.

Trans-differentiation can also be induced by inflammation in both type 1 and type 2 diabetes [40, 49]. Week 1 after Taf4 inactivation was marked by a potent but transient stress and inflammatory response with strong expression of several Reg-family proteins and multiple chemokines. This inflammatory/stress environment may be responsible for activation of Myc regulon activity and previously implicated in BC de-differentiation [39, 50, 51]. It is therefore possible that this transient response plays a role in triggering BC trans-differentiation also accounting for the absence of trans-differentiating cells at week 5.

Previous studies of trans-differentiation mainly focussed on transcription regulators with known roles in BC and alpha cell identity. However, BC cell populations actively

undergoing trans-differentiation have not been identified and the transcriptional regulatory programs involved only poorly characterized. Here we identify several intermediate and trans-differentiating BC populations and we define novel combinatorial sets of transcription factors associated with the transitions between them providing a model for BC trans-differentiation. While different cues may underlie initiation of BC trans-differentiation in diabetes or following Taf4 inactivation, further studies will determine whether they share common aspects of the transcriptional reprogramming described here.

Methods.

Genetically engineered mice.

Mice with the floxed Taf4 alleles as previously described [15] were crossed with *Ins*-Cre-ER[T2] transgenics and Taf4 inactivated by subcutaneous injection 1 time per day for 3 days with 5 ul of Tamoxifen diluted in oil (final concentration 10mg/ml). Blood glucose and insulin levels were measured over a period of 20 weeks with one measurement per week on the same day at the same time. Glycaemia was measured using a glucometer and insulin levels using the Milliplex Map Kit (Mouse Metabolic Magnetic Bead Panel. MMHMAG-44K, Millipore) Milliplex Map Kit. To assess glucose tolerance, mice were fasted for 12 hours before the first glucose measure and then injection of 100 µl of glucose solution (t= 0, 2g of glucose/ kg mouse weight). Insulin levels were monitored in blood samples taken over a 30-minute period starting 5 minutes before the glucose injection. All animal experiments were performed in accordance with European and national guidelines and policies (2010/63/UE directive and French decree 2013-118) and with approval of the National Ethics Committee.

Immunostaining and counting of islets.

Pancreas were isolated and fixed overnight in 4% paraformaldehyde, washed with PBS, dehydrated, paraffin embedded, and sectioned at 5 µm. For antigen retrieval, the

sections were boiled for 20 min in 10 mM of sodium citrate buffer. Sections were permeabilized with 3 × 5 min 0.1% Triton in PBS, blocked for 1 hr in 5% Neutral Goat Serum (NGS) in PBS, and incubated overnight in 5% NGS with primary antibodies: mouse anti-Glucagon [K79bB10] (ab10988), Guinea Pig anti-Insulin (Dako IR00261-2) and mouse anti-TAF4 (TAF II p135) (Santa Cruz (sc-136093). Sections were washed 3 × 5 min 0.1% Triton in PBS, and incubated with secondary antibodies: Goat anti-Guinea pig IgG H&L Alexa Fluor® 647 (ab150187) and Goat anti-Mouse IgG H&L Alexa Fluor® 488 (ab150117) for 2 hr. Sections were subsequently incubated with 1/2000 Hoechst nuclear stain for 10 min, washed 3 × 5 min in PBS, dried and mounted with Vectashild. Islet counting was performed on hematoxyline/eosin (HE) labeled sections of the pancreas. The entire pancreas was sectioned (5 μm slice) and one slice every 300 μm was scanned and analyzed with NanoZoomer Digital Pathology view annotation software to count and measure the islets.

Isolation of Langerhans islets.

Anesthetized mice (40 μL ketamine (100mg/mL) + 24 μL xylazine (20mg/mL) + 160 μL NaCl intraperitoneal injection per mouse) were injected in the pancreas with 4 ml of collagenase solution (7 ml HBSS, 70 μL HEPES (10 μM) (Gibco), 16,5μL DNAse (10mg/mL) (sigma DN5), 14 mg collagenase (2 mg/mL) (sigma Type V), and for isolation of RNA 2,3 μL of RNAsine (40U/μL). The pancreas was dissected and placed a tube containing 2 ml collagenase solution and incubated at 37 minutes at 38°C before stopping with 30mL Quenching Buffer (QB; 122 mL HBSS + 2,9 mL HEPES (25 μM) + 0,6 g BSA (0,5%) and 2 cycles of centrifugation at 1200 rpm for 2 min at 4°C. The pellet was suspended in 10 ml of Histopaque 1100 (5 ml histopaque 1077; Sigma 10771 and 6 ml histopaque 1119; Sigma 11191), centrifuged for 20 minutes at 1500 rpm at 4°C and the pellet washed 2 times with QB. The supernatant was discarded and pelleted islets transferred to a bacterial culture dish to be collected by pipetting. Purified islets were then cultured for 24 hours in RPMI1640 (HEPES

25mM) + final 2 mM L-glutamine (Gibco 11875-093) + 10% FCS 95 + 1% Penicillin/Streptomycin. Intracellular ATP concentration was measured with the Luminescent ATP Detection Assay Kit (ab113849) on isolated islets that were further digested to obtain a single cell population and FACS sorted to normalize the numbers of cells used.

Bulk RNA-seq from isolated islets

RNA was extracted from isolated islets from 2 mice per genotype with Nucleospin RNA plus XS (Macherey-Nagel). Complementary DNA was generated and linearly amplified from 3 ng total RNA using the Ovation RNA-seq V2 system (NuGEN technologies Inc., Leek, The Netherlands), according to the manufacturer's instructions. The amplified cDNA was then purified using Agencourt AMPure XP beads (Beckman-Coulter, Villepinte, France) in a 1.8:1 bead to sample ratio and fragmented by sonication using a Covaris E220 instrument (with duty cycle: 10%, maximum incident power: 175 watts and cycles/burst: 200 for 120 seconds). The RNA-seq libraries were generated from 100 ng fragmented cDNA using the Ovation Ultralow v2 library system (NuGEN technologies Inc., Leek, The Netherlands) according to the manufacturer's instructions, with only 6 PCR cycles for library amplification. The final libraries were verified for quality and quantified using capillary electrophoresis before sequencing on an Illumina Hi-Seq4000. Reads were preprocessed using cutadapt version 1.10 in order to remove adapter and low-quality sequences (Phred quality score below 20). After this preprocessing, reads shorter than 40 bases were discarded. Reads were mapped to rRNA sequences using bowtie version 2.2.8, and reads mapping to rRNA sequences were removed. Reads were mapped onto the mm9 assembly of Mus musculus genome using STAR version 2.5.3a. Gene expression quantification was performed from uniquely aligned reads using htseq-count version 0.6.1p1, with annotations from Ensembl version 67 and “union” mode. Only non-ambiguously assigned reads were retained. Read counts were normalized across samples with the median-of-ratios method [52]. Comparisons of interest were performed using

the Wald test for differential expression [53] and implemented in the Bioconductor package DESeq2 version 1.16.1. Genes with high Cook's distance were filtered out and independent filtering based on the mean of normalized counts was performed. P-values were adjusted for multiple testing using the Benjamini and Hochberg method [**3**]. Heatmaps were generated with R-package pheatmap v1.0.12. Deregulated genes were defined as genes with log2(fold change) >1 or <-1 and adjusted p-value <0.05. Gene ontology analyses was performed using GSEA (https://www.gsea-msigdb.org/gsea/index.jsp) or David (https://david-d.ncifcrf.gov/).

ATAC-seq from isolated islets.

ATAC-seq was performed from 20 000 cells from isolated islets. Sequenced reads were mapped to the mouse genome assembly mm9 using Bowtie [54] with the following arguments: "-m 1 --strata --best -y -S -l 40 -p 2".

After sequencing, peak detection was performed using the MACS software [55] v2.1.1.20160309. Peaks were annotated with Homer(http://homer.salk.edu/homer/ngs/annotation.html) using the GTF from ENSEMBL v67. Peak intersections were computed using Bedtools [56] Global Clustering was done using seqMINER [57]. In silico footprinting signatures were calculated using TOBIAS [27] (v0.5.1; https://github.com/loosolab/TOBIAS/), differential footprinting scores were plotted with R-package ggplot2 (https://ggplot2.tidyverse.org.).

Single cell RNA-seq from isolated islets.

After islet isolation, further digestion was performed with Accutase (A6964, Sigma) at 27 °C for 5 minutes and the cells were suspended in culture medium. Cells were then FACS sorted to select only live cells. 300 islets from a male WT mouse and 300 islets pooled from 2 male mutant mice 1 week or 5 weeks after Taf4 inactivation were used for scRNA-seq. Cell capture was performed using 10X Genomics Chromium Analyzer. After sequencing, raw reads were processed using CellRanger (v 3.1) to align on the mm10 mouse genome, remove

unexpressed genes and quantify barcodes and UMIs. Data were then analyzed in R (v3.6.3). For the WT sample only, potential doublets were removed by manually filtering out cells expressing high levels of Ins1+Gcg, Ins1+Sst or Gcg+Sst and then by running R-package DoubletFinder [58]. The WT sample was down-scaled to 12000 cells before aggregation with Taf4$^{b-/-}$ 5-week sample and analyzed using Seurat v3.1.4 [59] following the guided clustering tutorial vignette. Only cells with feature count ranging from 200 to 1500 and with percentage of mitochondrial reads <15% were kept for the analysis. Counts are normalized with the "LogNormalize" method and data are scaled to remove unwanted sources of variation. Clustering was done on most variable features using 12 PCs and a resolution of 0.8. Regulome analyses were performed using the SCENIC v1.1.2.2 package.

Acknowledgements

We thank, all the staff of the IGBMC common facilities in particular the IGBMC animal, high throughput sequencing (GenomEast) and electron microscopy facilities, the Mouse Clinical Institute for help with measurement of insulin and glucose levels. This work was supported by institutional grants from the Centre National de la Recherche Scientifique, the Institut National de Santé et de la Recherche Médicale, the Université de Strasbourg, the Ligue Nationale contre le Cancer, Agence Nationale de la Recherche (ANR)-15-CE14-003 project Dysmetaf, the ANR-10-LABX-0030-INRT French state fund under the framwork programme Investissements d'Avenir ANR-10-IDEX-0002-02. The IGBMC high throughput sequencing facility is a member of the "France Génomique" consortium (ANR10-INBS-09-08). ID is an 'équipe labellisée' of the Ligue Nationale contre le Cancer. TK was supported by fellowships from the ANR and the FRM.

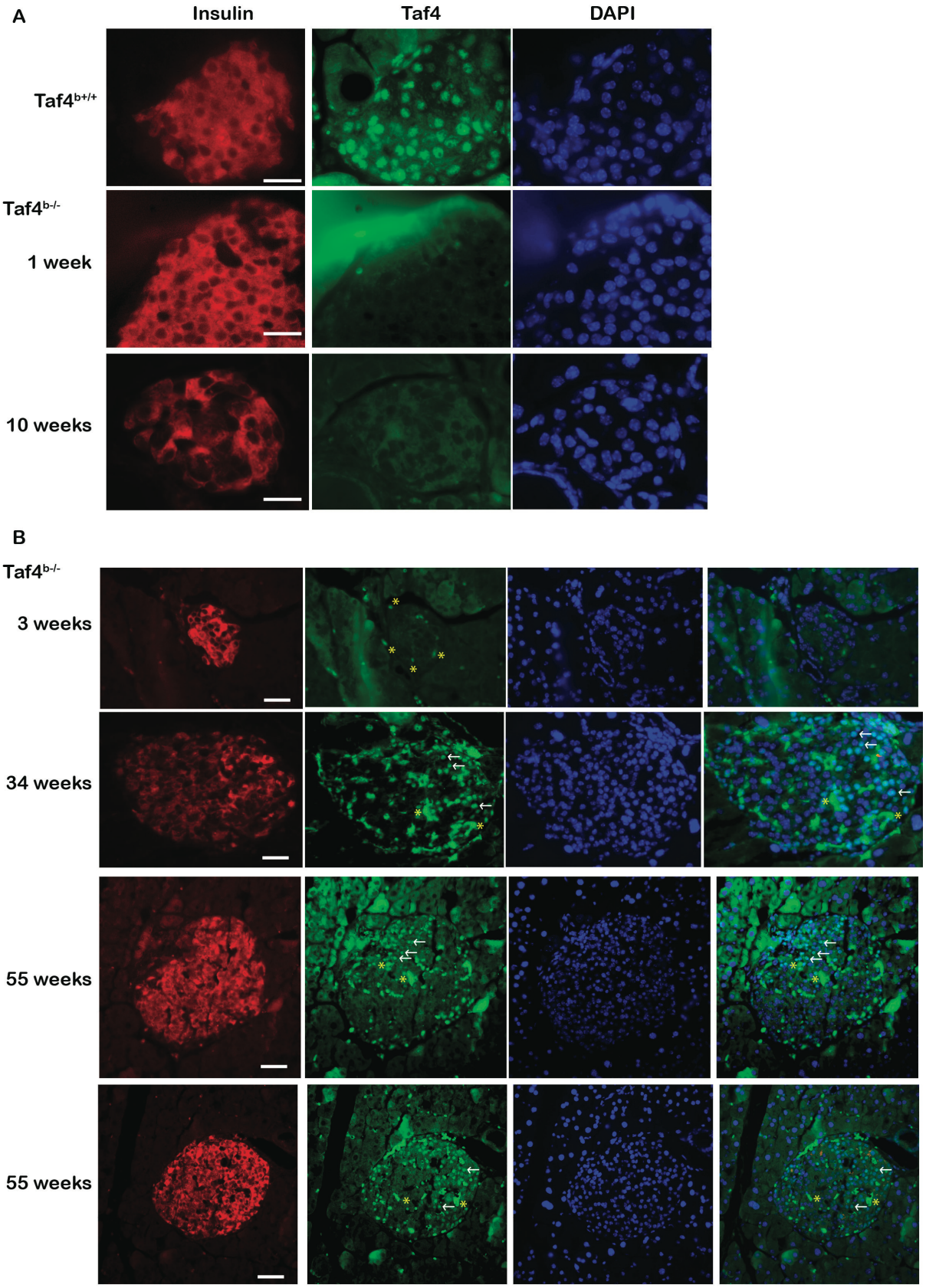
All sequencing data reported here have been submitted to the GEO database.

References.

1. Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424**, 147-151. (2003).
2. Taatjes, D.J., Marr, M.T. & Tjian, R. Regulatory diversity among metazoan co-activator complexes. *Nat Rev Mol Cell Biol* **5**, 403-410 (2004).
3. Bell, B. & Tora, L. Regulation of Gene Expression by Multiple Forms of TFIID and Other Novel TAFII-Containing Complexes. *Exp.Cell Res*. **246**, 11-19 (1999).
4. Malik, S. & Roeder, R.G. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat Rev Genet* **11**, 761-772 (2010).
5. Muller, F. & Tora, L. The multicoloured world of promoter recognition complexes. *Embo J* **23**, 2-8 (2004).
6. Cler, E., Papai, G., Schultz, P. & Davidson, I. Recent advances in understanding the structure and function of general transcription factor TFIID. *Cell Mol Life Sci* **66**, 2123-2134 (2009).
7. Bhuiyan, T. & Timmers, H.T.M. Promoter Recognition: Putting TFIID on the Spot. *Trends Cell Biol* **29**, 752-763 (2019).
8. Patel, A.B., Greber, B.J. & Nogales, E. Recent insights into the structure of TFIID, its assembly, and its binding to core promoter. *Curr Opin Struct Biol* **61**, 17-24 (2019).
9. Patel, A.B. *et al*. Structure of human TFIID and mechanism of TBP loading onto promoter DNA. *Science* **362** (2018).
10. Wright, K.J., Marr, M.T., 2nd & Tjian, R. TAF4 nucleates a core subcomplex of TFIID and mediates activated transcription from a TATA-less promoter. *Proc Natl Acad Sci U S A* **103**, 12347-12352 (2006).
11. Alpern, D. *et al*. TAF4, a subunit of transcription factor II D, directs promoter occupancy of nuclear receptor HNF4A during post-natal hepatocyte differentiation. *Elife* **3**, e03613 (2014).
12. Warfield, L. *et al*. Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID. *Mol Cell* **68**, 118-129 e115 (2017).
13. Bonnet, J. *et al*. The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. *Genes Dev* **28**, 1999-2012 (2014).
14. Timmers, H.T.M. & Tora, L. Transcript Buffering: A Balancing Act between mRNA Synthesis and mRNA Degradation. *Mol Cell* **72**, 10-17 (2018).
15. Mengus, G. *et al*. TAF4 inactivation in embryonic fibroblasts activates TGFbeta signalling and autocrine growth. *Embo J* **24**, 2753-2767 (2005).
16. Langer, D. *et al*. Essential role of the TFIID subunit TAF4 in murine embryogenesis and embryonic stem cell differentiation. *Nat Commun* **7**, 11063 (2016).
17. Fadloun, A. *et al*. The TFIID subunit TAF4 regulates keratinocyte proliferation and has cell-autonomous and non-cell-autonomous tumour suppressor activity in mouse epidermis. *Development* **134**, 2947-2958 (2007).
18. Fadloun, A. *et al*. Retinoic acid induces TGFbeta-dependent autocrine fibroblast growth. *Oncogene* **27**, 477-489 (2008).
19. Martianov, I. *et al*. TAF4 inactivation reveals the 3 dimensional growth promoting activities of collagen 6A3. *PLoS One* **9**, e87365 (2014).
20. Yamagata, K. Roles of HNF1alpha and HNF4alpha in pancreatic beta-cells: lessons from a monogenic form of diabetes (MODY). *Vitam Horm* **95**, 407-423 (2014).
21. Odom, D.T. *et al*. Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303**, 1378-1381 (2004).
22. Han, E.H., Rha, G.B. & Chi, Y.I. MED25 is a mediator component of HNF4alpha-driven transcription leading to insulin secretion in pancreatic beta-cells. *PLoS One* **7**, e44007 (2012).

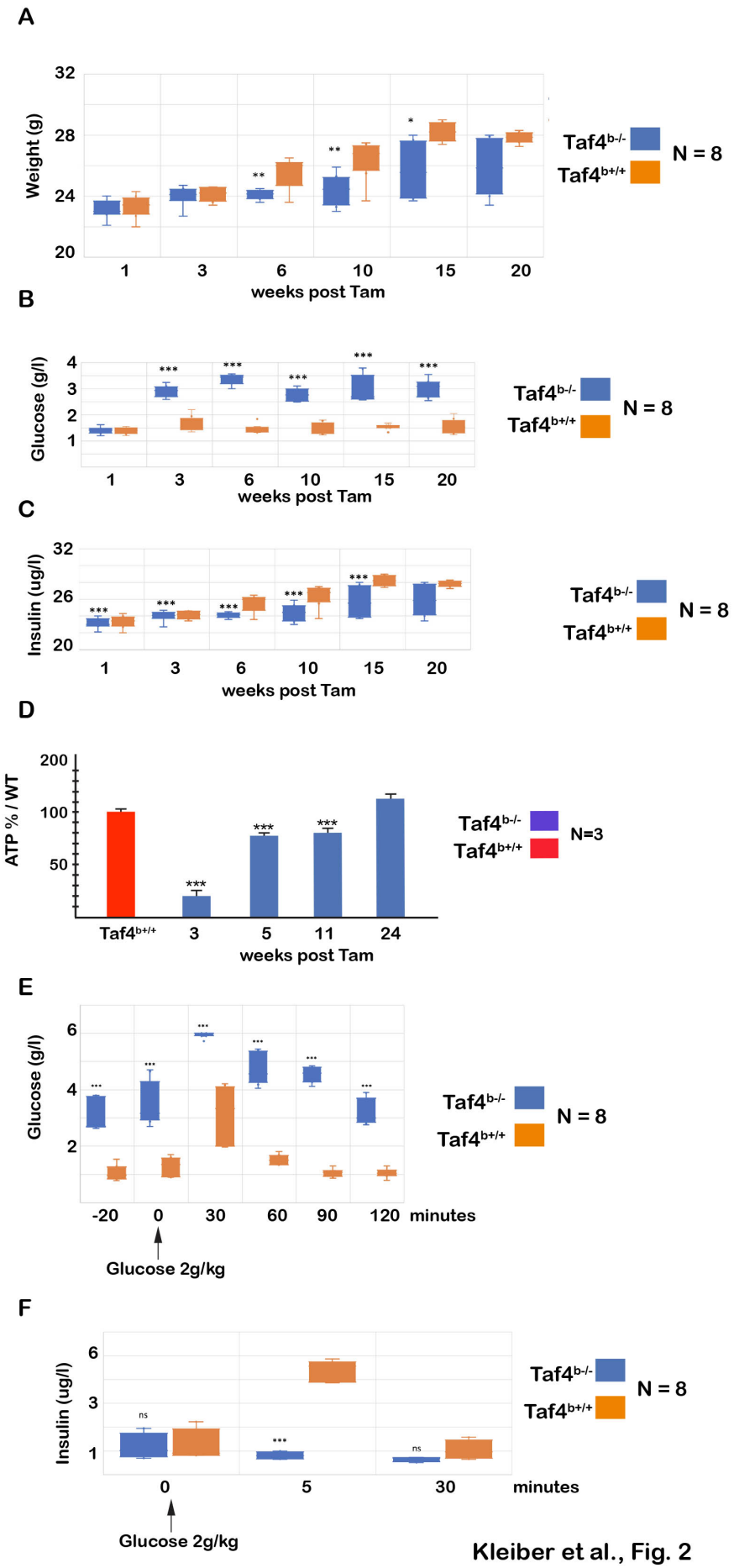
23. Gupta, R.K. *et al*. The MODY1 gene HNF-4alpha regulates selected genes involved in insulin secretion. *J Clin Invest* **115**, 1006-1015 (2005).
24. Miura, A. *et al*. Hepatocyte nuclear factor-4alpha is essential for glucose-stimulated insulin secretion by pancreatic beta-cells. *J Biol Chem* **281**, 5246-5257 (2006).
25. Gupta, R.K. *et al*. Expansion of adult beta-cell mass in response to increased metabolic demand is dependent on HNF-4alpha. *Genes Dev* **21**, 756-769 (2007).
26. Cattin, A.L. *et al*. Hepatocyte nuclear factor 4alpha, a key factor for homeostasis, cell architecture, and barrier function of the adult intestinal epithelium. *Mol Cell Biol* **29**, 6294-6308 (2009).
27. Bentsen M, G.P., Schultheis H, Klee K, PetrovaA, Wiegandt R, Fust A, Preussner J, Kuenne C, Braun T, Kim J, Looso M. Beyond accessibility: ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Biorxiv* (2020).
28. Aibar, S. *et al*. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086 (2017).
29. Juliana, C.A. *et al*. ATF5 regulates beta-cell survival during stress. *Proc Natl Acad Sci U S A* **114**, 1341-1346 (2017).
30. Moin, A.S.M. & Butler, A.E. Alterations in Beta Cell Identity in Type 1 and Type 2 Diabetes. *Curr Diab Rep* **19**, 83 (2019).
31. Zhou, Q. & Melton, D.A. Pancreas regeneration. *Nature* **557**, 351-358 (2018).
32. Thorel, F. *et al*. Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss. *Nature* **464**, 1149-1154 (2010).
33. Hunter, C.S. & Stein, R.W. Evidence for Loss in Identity, De-Differentiation, and Trans-Differentiation of Islet beta-Cells in Type 2 Diabetes. *Front Genet* **8**, 35 (2017).
34. Chakravarthy, H. *et al*. Converting Adult Pancreatic Islet alpha Cells into beta Cells by Targeting Both Dnmt1 and Arx. *Cell Metab* **25**, 622-634 (2017).
35. Baron, M. *et al*. A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst* **3**, 346-360 e344 (2016).
36. Sharma, R.B. *et al*. Insulin demand regulates beta cell number via the unfolded protein response. *J Clin Invest* **125**, 3831-3846 (2015).
37. Muraro, M.J. *et al*. A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst* **3**, 385-394 e383 (2016).
38. Xin, Y. *et al*. RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes. *Cell Metab* **24**, 608-615 (2016).
39. Bensellam, M., Jonas, J.C. & Laybutt, D.R. Mechanisms of beta-cell dedifferentiation in diabetes: recent findings and future research directions. *J Endocrinol* **236**, R109-R143 (2018).
40. Tritschler, S., Theis, F.J., Lickert, H. & Bottcher, A. Systematic single-cell analysis provides new insights into heterogeneity and plasticity of the pancreas. *Mol Metab* **6**, 974-990 (2017).
41. Spaeth, J.M. *et al*. The FOXP1, FOXP2 and FOXP4 transcription factors are required for islet alpha cell proliferation and function in mice. *Diabetologia* **58**, 1836-1844 (2015).
42. Nomoto, H. *et al*. Inhibition of Small Maf Function in Pancreatic beta-Cells Improves Glucose Tolerance Through the Enhancement of Insulin Gene Transcription and Insulin Secretion. *Endocrinology* **156**, 3570-3580 (2015).
43. Katsuoka, F. & Yamamoto, M. Small Maf proteins (MafF, MafG, MafK): History, structure and function. *Gene* **586**, 197-205 (2016).
44. Katsuoka, F., Yamazaki, H. & Yamamoto, M. Small Maf deficiency recapitulates the liver phenotypes of Nrf1- and Nrf2-deficient mice. *Genes Cells* **21**, 1309-1319 (2016).
45. Puri, S., Folias, A.E. & Hebrok, M. Plasticity and dedifferentiation within the pancreas: development, homeostasis, and disease. *Cell Stem Cell* **16**, 18-31 (2015).

46. van der Meulen, T. *et al*. Virgin Beta Cells Persist throughout Life at a Neogenic Niche within Pancreatic Islets. *Cell Metab* **25**, 911-926 e916 (2017).
47. Chera, S. *et al*. Diabetes recovery by age-dependent conversion of pancreatic delta-cells into insulin producers. *Nature* **514**, 503-507 (2014).
48. Talchai, C., Xuan, S., Lin, H.V., Sussel, L. & Accili, D. Pancreatic beta cell dedifferentiation as a mechanism of diabetic beta cell failure. *Cell* **150**, 1223-1234 (2012).
49. Nordmann, T.M. *et al*. The Role of Inflammation in beta-cell Dedifferentiation. *Sci Rep* **7**, 6285 (2017).
50. Laybutt, D.R. *et al*. Overexpression of c-Myc in beta-cells of transgenic mice causes proliferation and apoptosis, downregulation of insulin gene expression, and diabetes. *Diabetes* **51**, 1793-1804 (2002).
51. Pascal, S.M., Guiot, Y., Pelengaris, S., Khan, M. & Jonas, J.C. Effects of c-MYC activation on glucose stimulus-secretion coupling events in mouse pancreatic islets. *Am J Physiol Endocrinol Metab* **295**, E92-102 (2008).
52. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).
53. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
54. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
55. Zhang, Y. *et al*. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
56. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
57. Ye, T. *et al*. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Research* **39**, e35 (2011).
58. McGinnis, C.S., Murrow, L.M. & Gartner, Z.J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337 e324 (2019).
59. Stuart, T. *et al*. Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).



Kleiber et al., Fig. 1

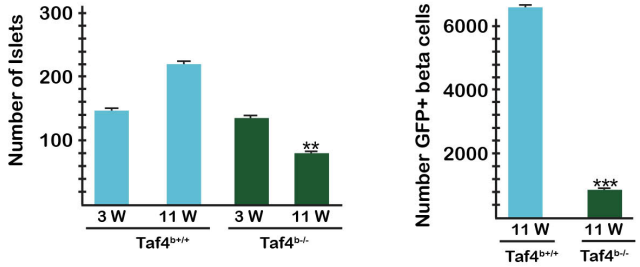
Figure 1. Inactivation of Taf4 in BCs. **A**. Immunostaining of isolated Langerhans islets from mice with the indicated genotypes for Taf4 or insulin as indicated. The number of weeks after Tam injection are indicated. **B**. Immunostaining of isolated Langerhans islets as above but with the addition of the Taf4-DAPI merge. Representative Taf4-positive nuclei at 34 and 55 weeks are indicated by arrows, not to be confounded with the stronger non-specific staining indicated by *. Scale bar = 100 μM.



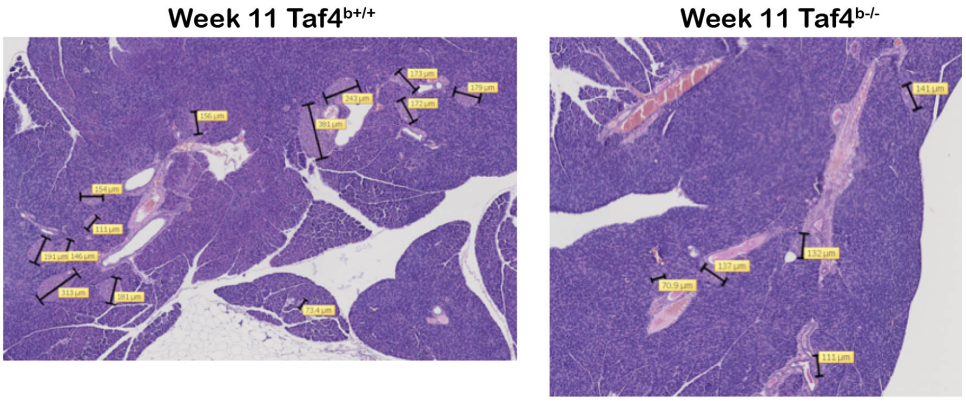
Kleiber et al., Fig. 2

Figure 2. Physiological parameters of *Taf4*$^{b-/-}$ animals. **A.** Weight of *Taf4*$^{+/+}$ and *Taf4*$^{b-/-}$ animals at the indicated number of weeks following Tam injection. **B-C**. Blood glucose and insulin levels of *Taf4*$^{+/+}$ and *Taf4*$^{b-/-}$ animals at the indicated number of weeks following Tam injection. **D.** ATP content of isolated islets from control animals or from *Taf4*$^{b-/-}$ animals at the indicated number of weeks following Tam injection. **E**. Blood glucose levels at the indicated times following injection of glucose in control and *Taf4*$^{b-/-}$ animals. Mice at 14 weeks were injected with Tam and glucose administered 2 weeks following Tam treatment. **F.** Blood insulin levels from the same protocol. Student T-test *** $p<0.001$ from N=8.

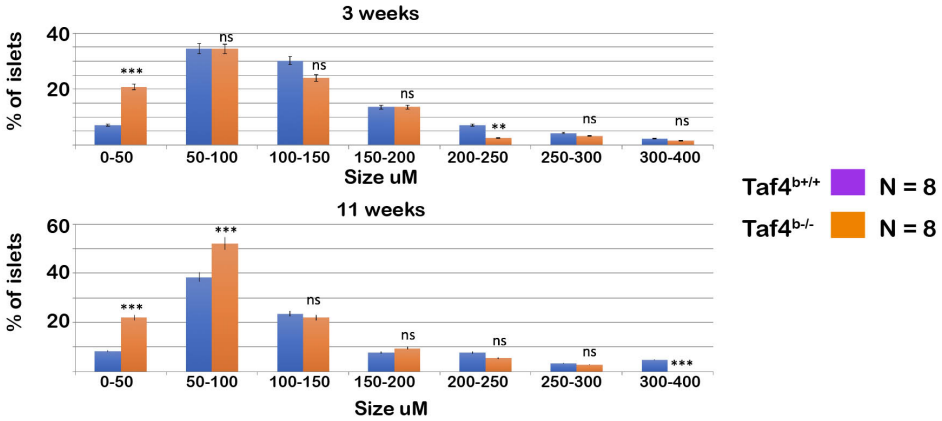
A



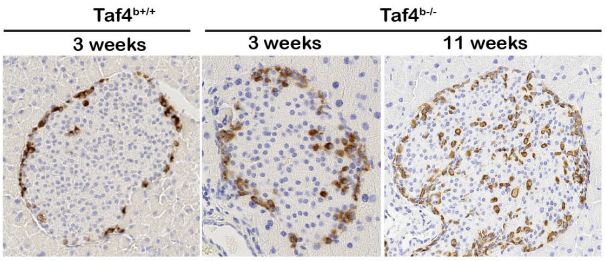
B



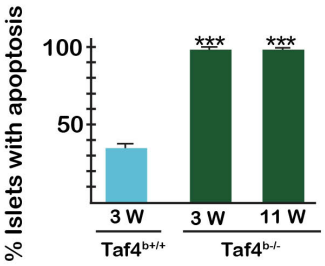
C



D

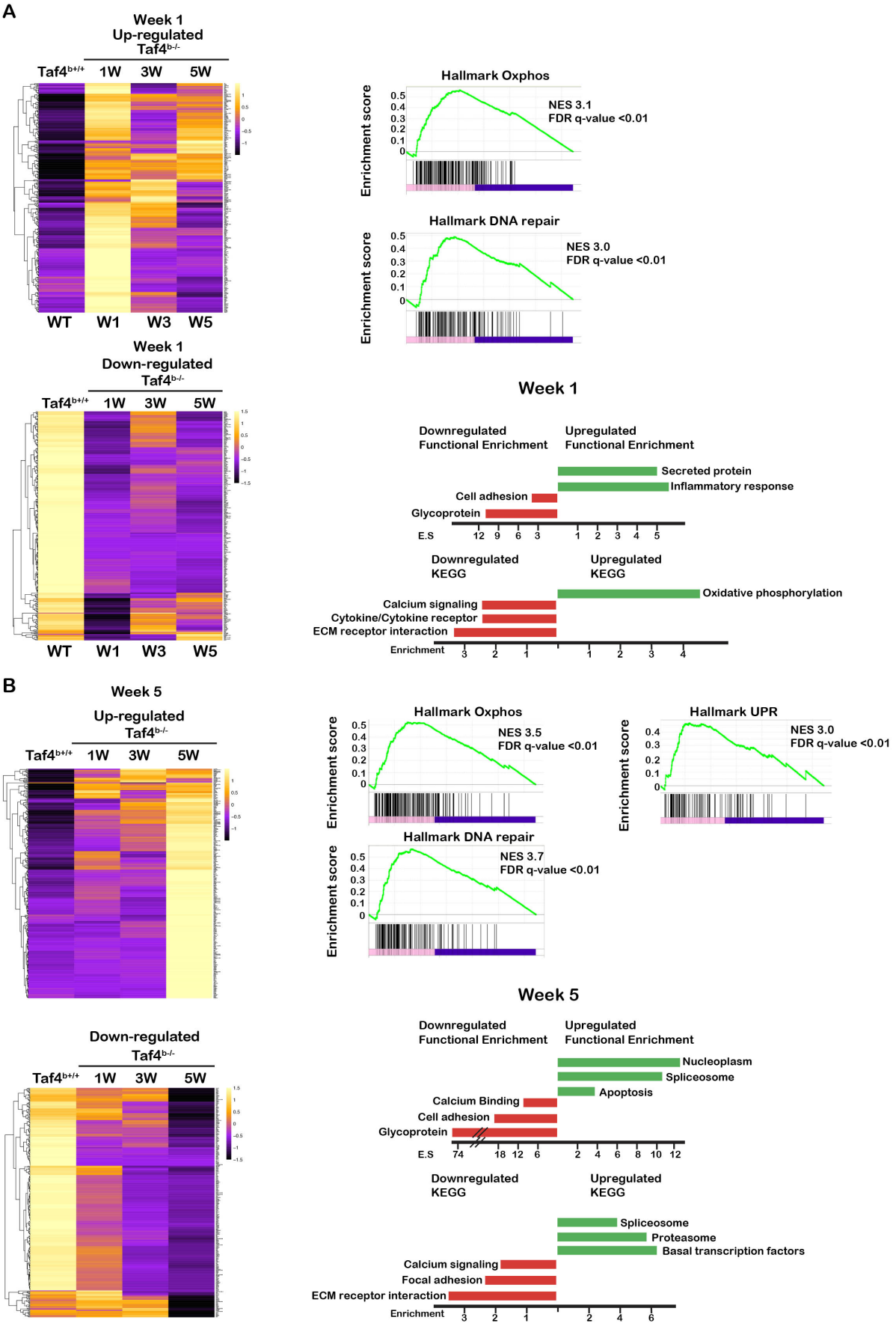


E



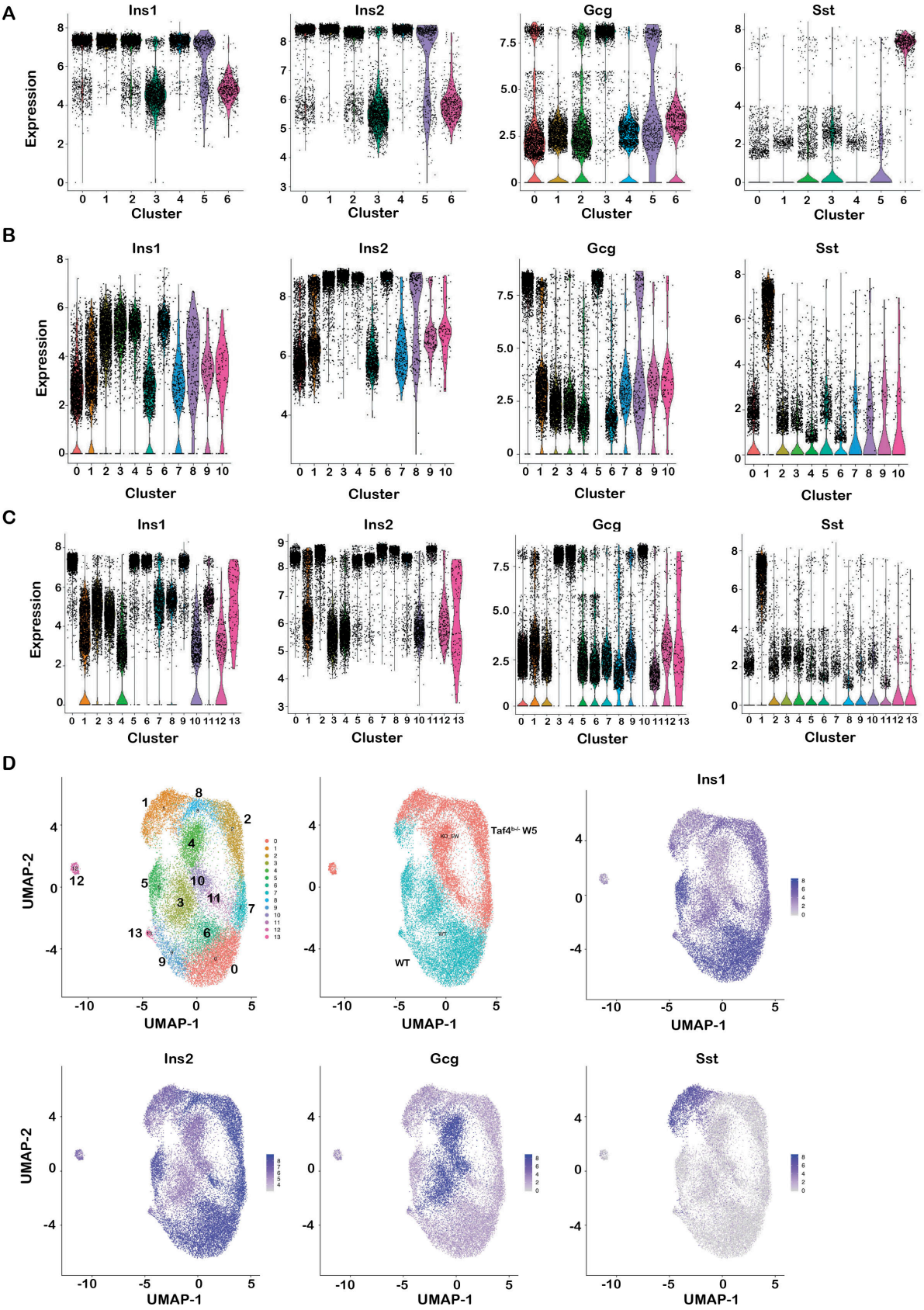
Kleiber et al., Fig. 3

Figure 3. Long term fate of BCs following Taf4 inactivation. **A.** Number of islets in *Taf4*$^{b-/-}$ and control animals. **B.** Histology of pancreas sections used to measure islet size distribution in *Taf4*$^{b-/-}$ and control animals in panel **C**. **D.** Tunel staining of islets from in *Taf4*$^{b-/-}$ and control animals. **E**. % counted of islets with apoptotic BCs. Student T-test ** $p<0.05$, *** $p<0.001$ from N=8.



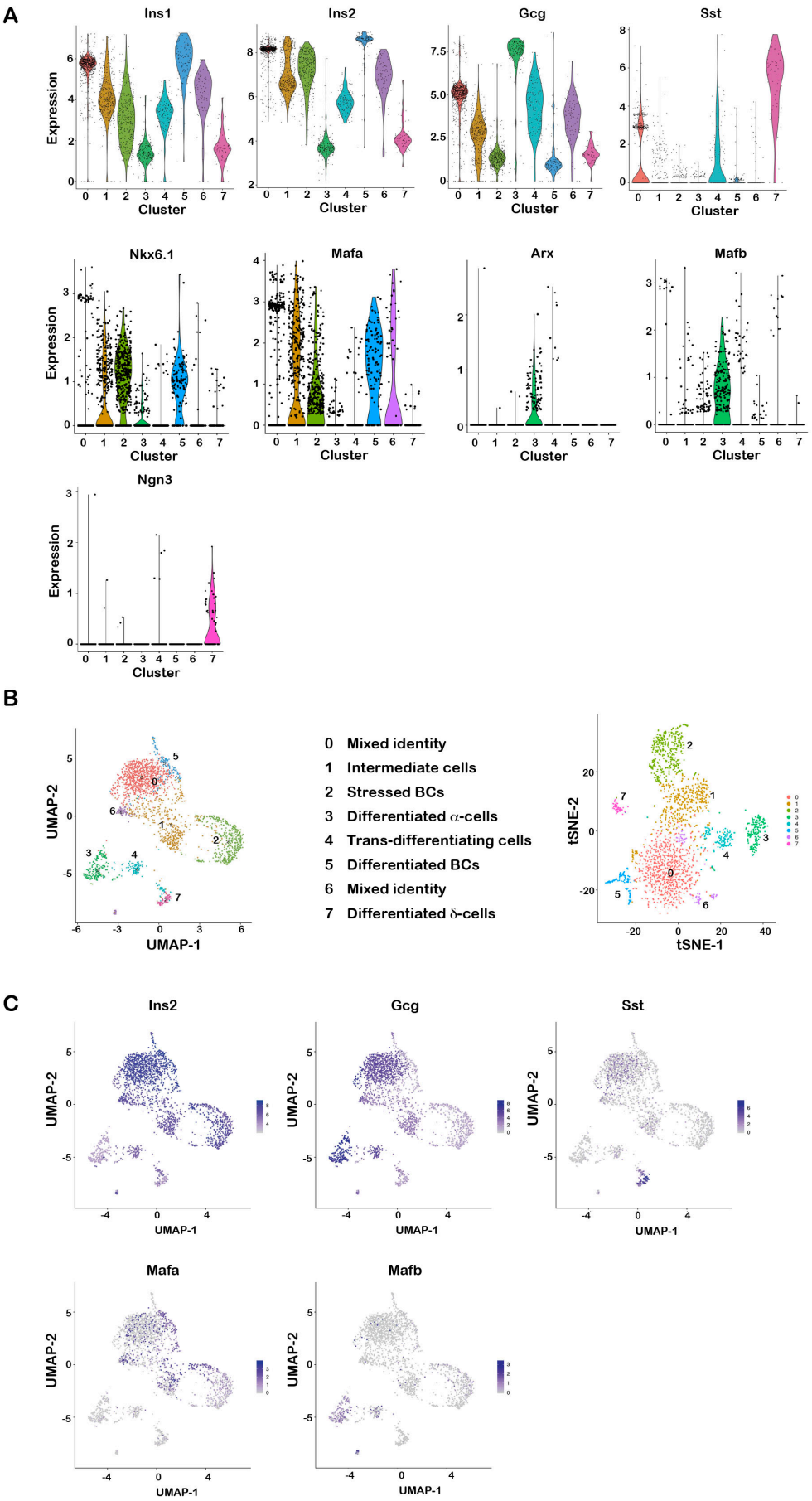
Kleiber et al., Fig. 4

Figure 4. Effects of Taf4 inactivation on gene expression of isolated islets. **A.** Heatmap showing the expression of the 100 genes most up and down-regulated 1 week after Taf4 inactivation. Right hand panels show the GSEA and ontology analyses of the de-regulated genes. **B.** Heatmap showing the expression of the 100 genes most up and down-regulated 5 weeks after Taf4 inactivation. Right hand panels show the GSEA and ontology analyses of the de-regulated genes.



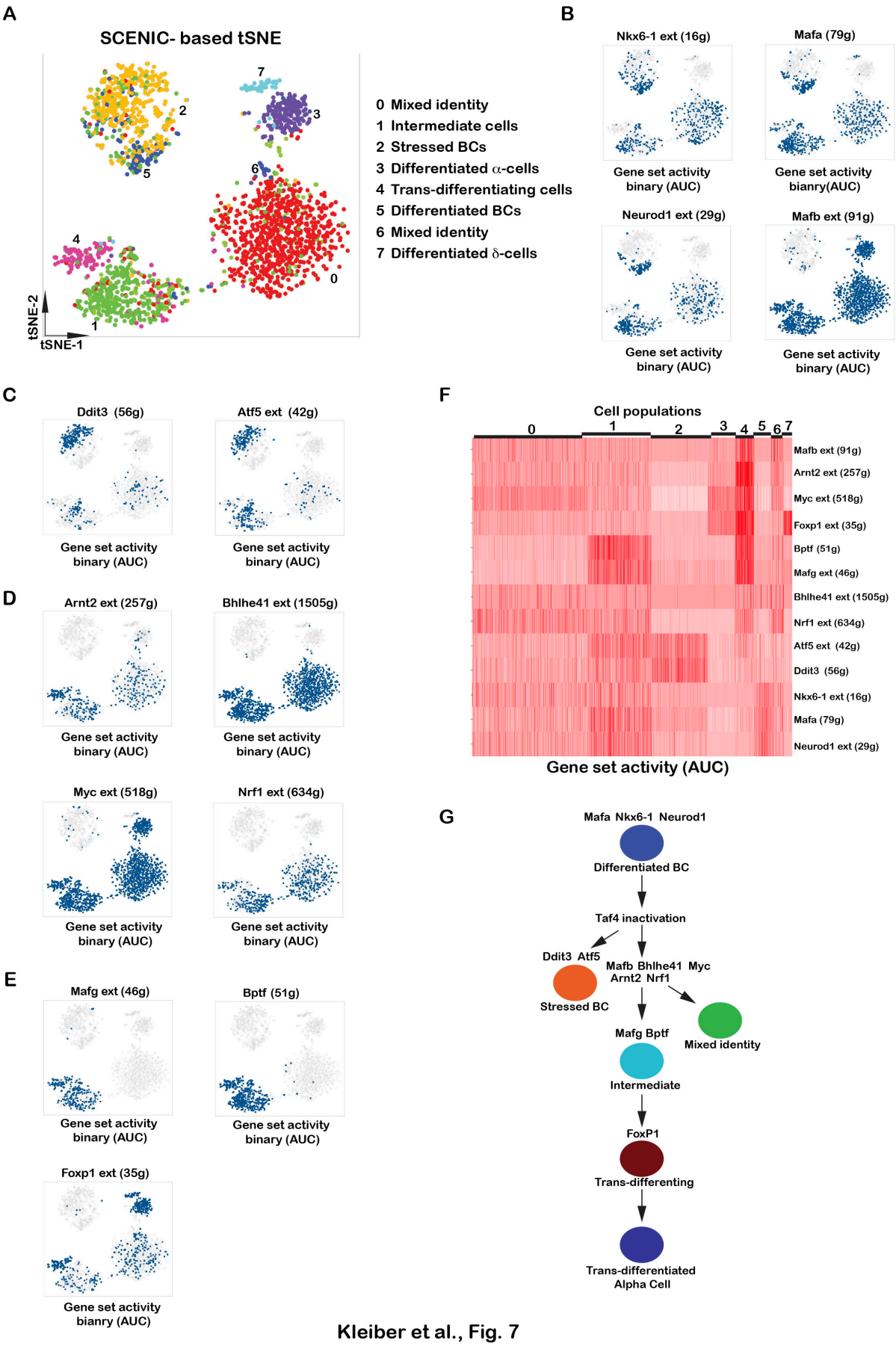
Kleiber et al., Fig. 5

Figure 5. Sc-RNA-seq of WT and week 5 mutant islets. **A-C.** Violin plots of expression of the indicated genes in each cluster from WT islets (A), 5 week mutant islets (B) and the aggregate (C). **D**. UMAP representations of the aggregate data, illustrating origin of cells from WT and mutant animals and the expression of the genes indicated in each panel.



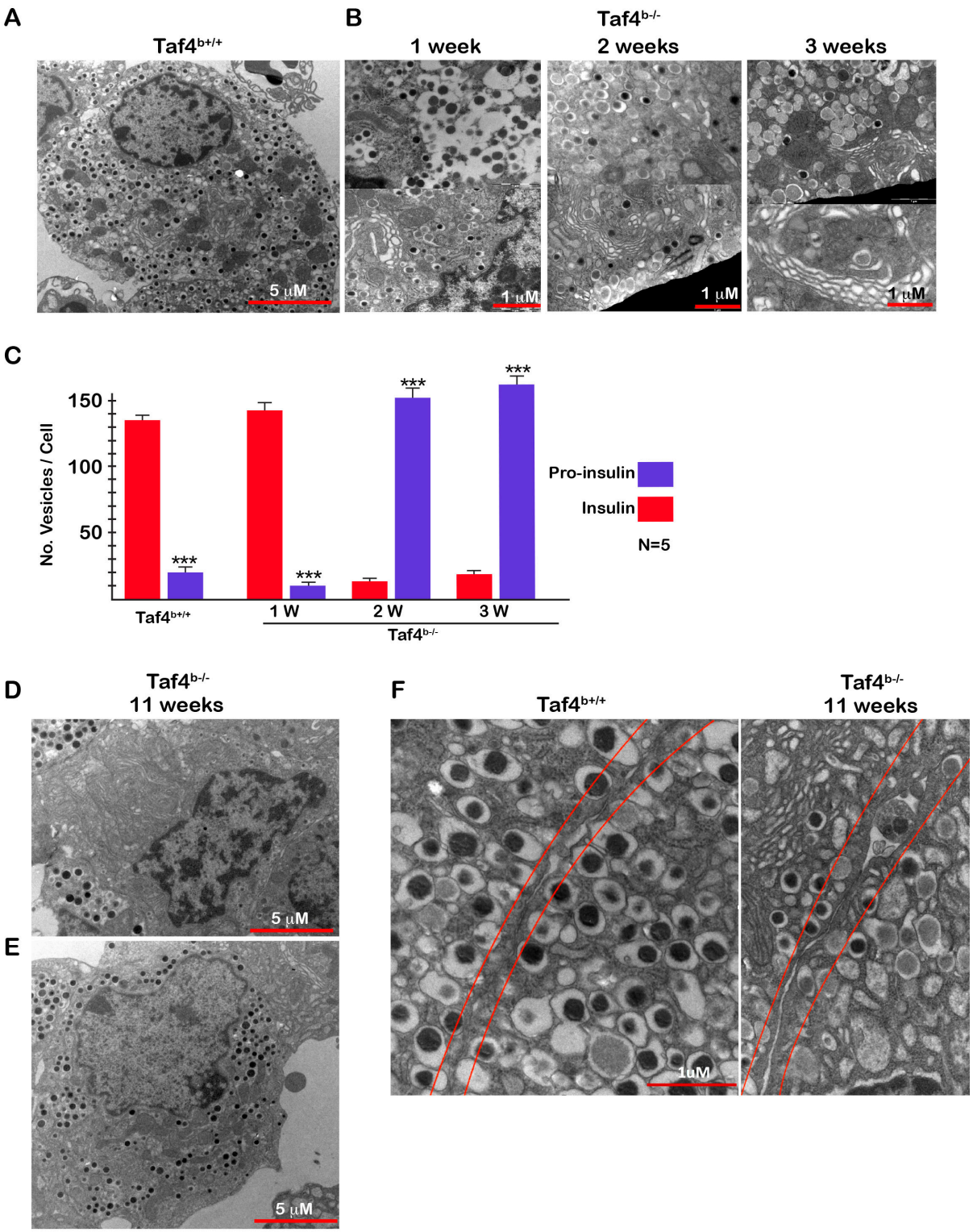
Kleiber et al., Fig. 6

Figure 6. Sc-RNA-seq of week 1 mutant islets. **A.** Violin plots of expression of the indicated genes in each cluster. **B**. UMAP (left) and tSNE (right) representations of the cell populations. **C.** UMAPs illustrating the expression of the genes indicated in each panel.



Kleiber et al., Fig. 7

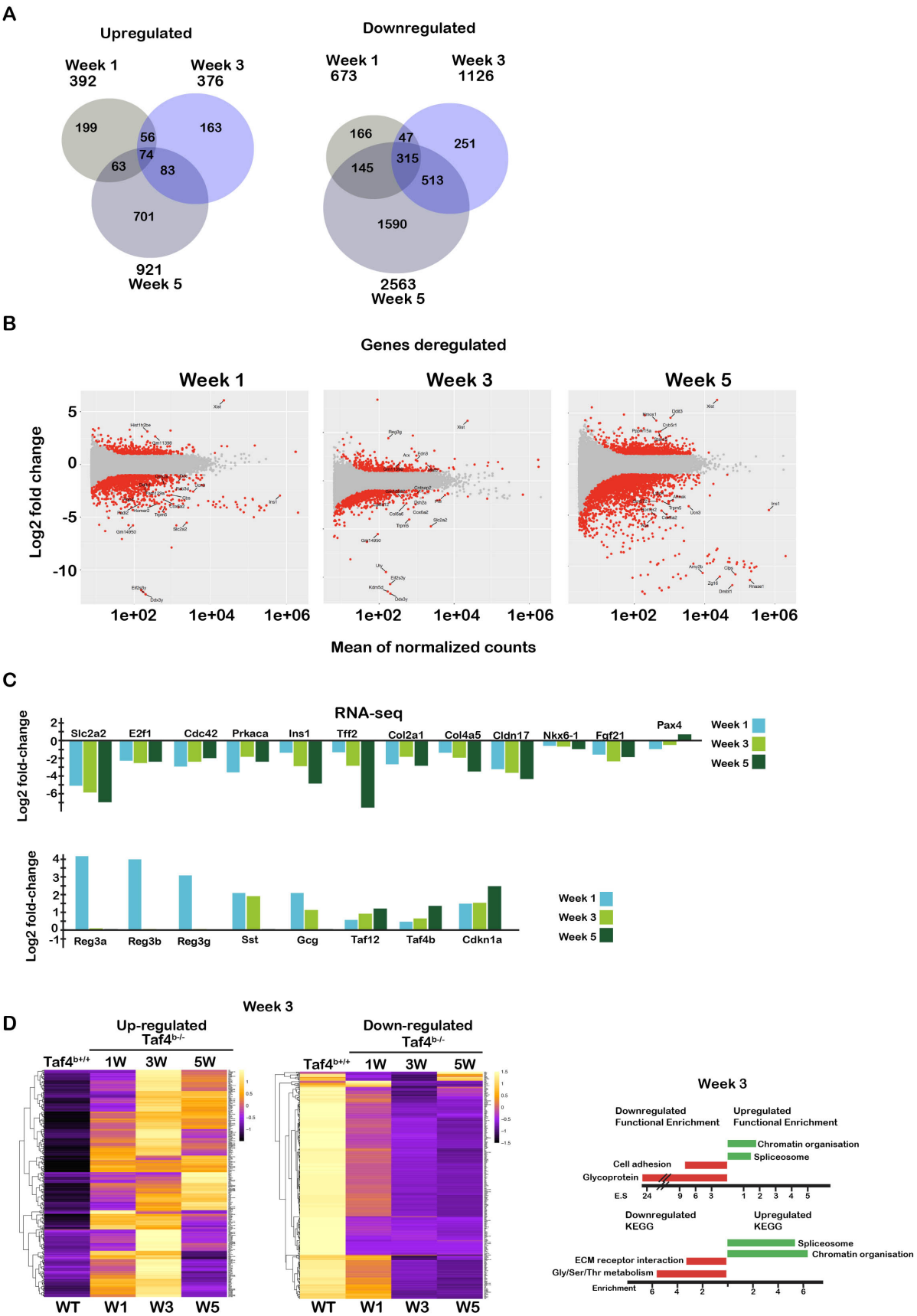
Figure 7. Transcription factors regulating cell state in week 1 mutant islets. **A**. SCENIC-based tSNE representation coloring cells based on the binary activities of the transcription factor regulons. **B-E**. Binary activities of the transcription factor regulons in the different cell populations. **F.** Heatmap of regulon activities in the different cell populations were quantified using AUCell. **G.** A flow-chart of the progression of cell state from differentiated BCs to trans-differentiated alpha cells. The SCENIC regulons marking the transitions of each state are indicated.



Kleiber et al., Fig. S1

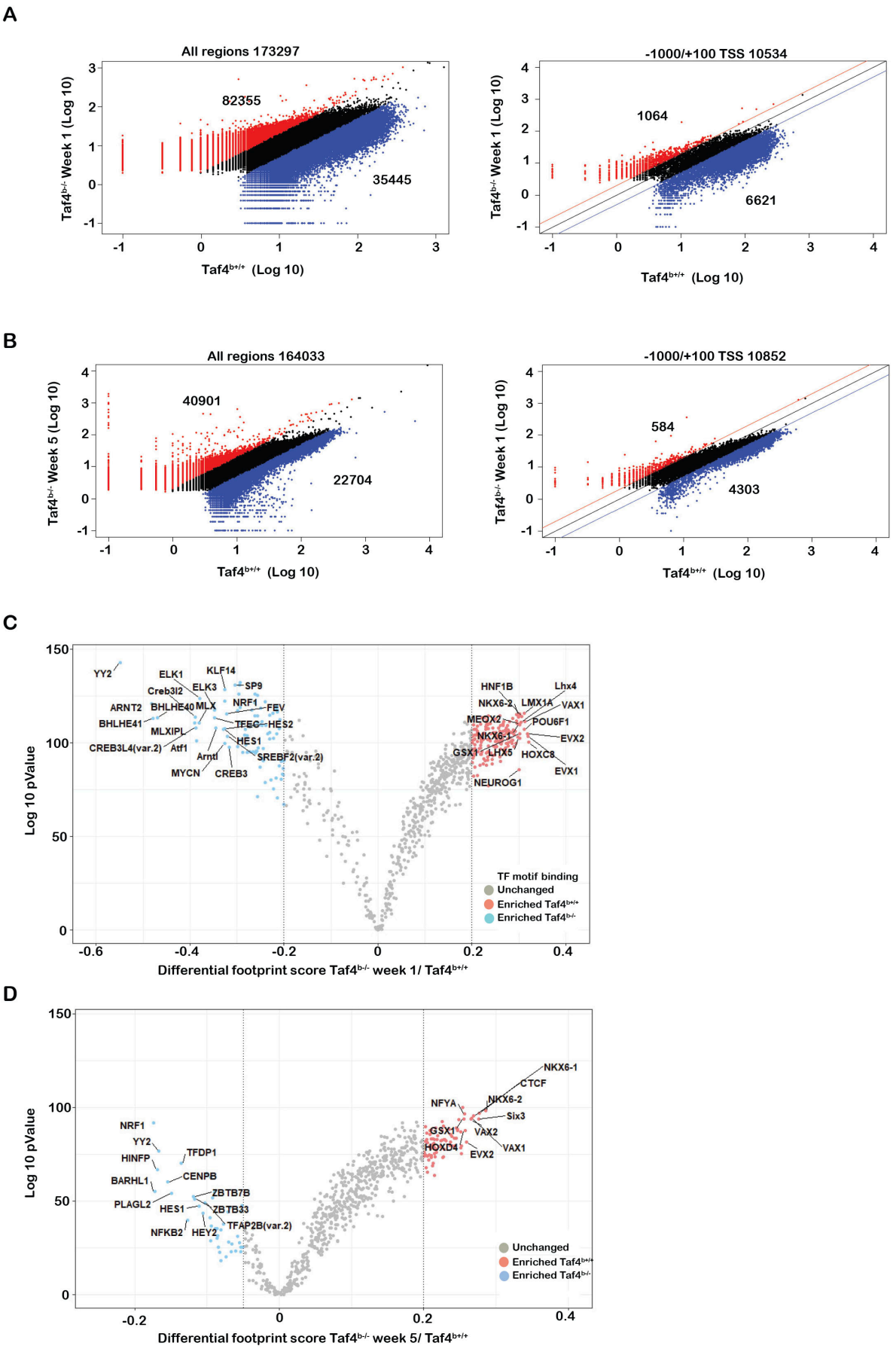
Supplemental Figure 1. Ultrastructure of BCs by electron microscopy. **A**. BC from control animal with densely staining insulin secretory granules in the cytoplasm. **B**. BCs from *Taf4*$^{b-/-}$

animals highlighting accumulation of lighter staining proinsulin secretory granules and expanded Golgi apparatus. **C.** Quantitation of insulin and proinsulin secretory granules. **D-E.** BCs from *Taf4*$^{b-/-}$ animals 11 weeks after Tam injection illustrating cells with few (70%, panel D) or more numerous (30% panel E) insulin secretory granules. **F.** Control and mutant BCs highlighting the cell-cell contacts and the loss of tight contacts between the mutant cells. Student T-test *** $p<0.001$



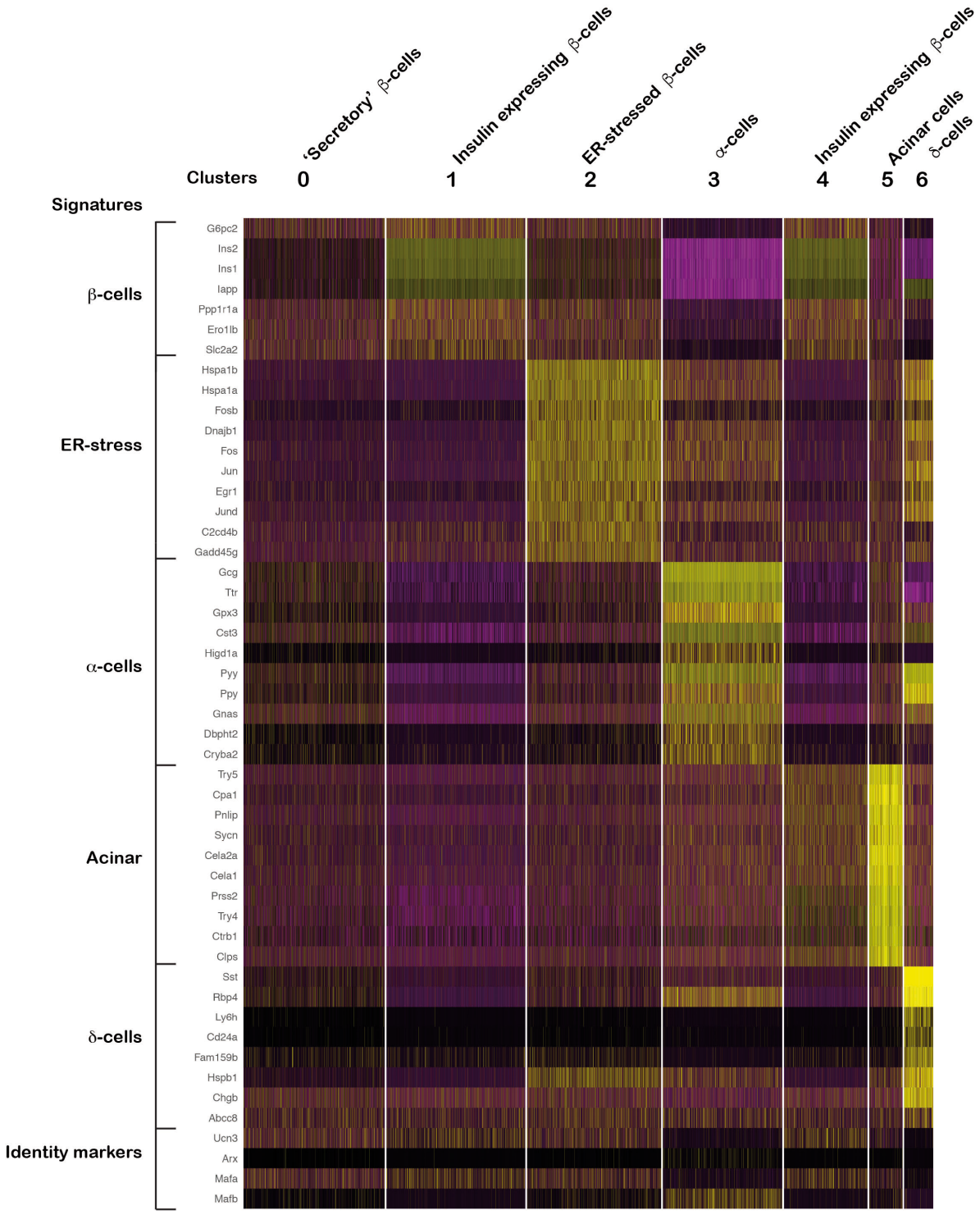
Kleiber et al., Fig. S2

Supplemental Figure 2. Changes in gene expression upon Taf4 inactivation. **A.** Venn diagrams summarizing genes up and down-regulated 1, 3 and 5 weeks after Tam injection. **B**. Volcano plots showing changes in gene expression after 1, 3 and 5 weeks. **C.** Summary of RNA-seq data on the expression of selected up and down-regulated genes critical for BC function, identity, cell-cell contact, stress response and alpha and delta cell markers. **D**. Heatmap showing the expression of the 100 genes most up and down-regulated 3 weeks after Taf4 inactivation. Right hand panel shows ontology analyses of the de-regulated genes.

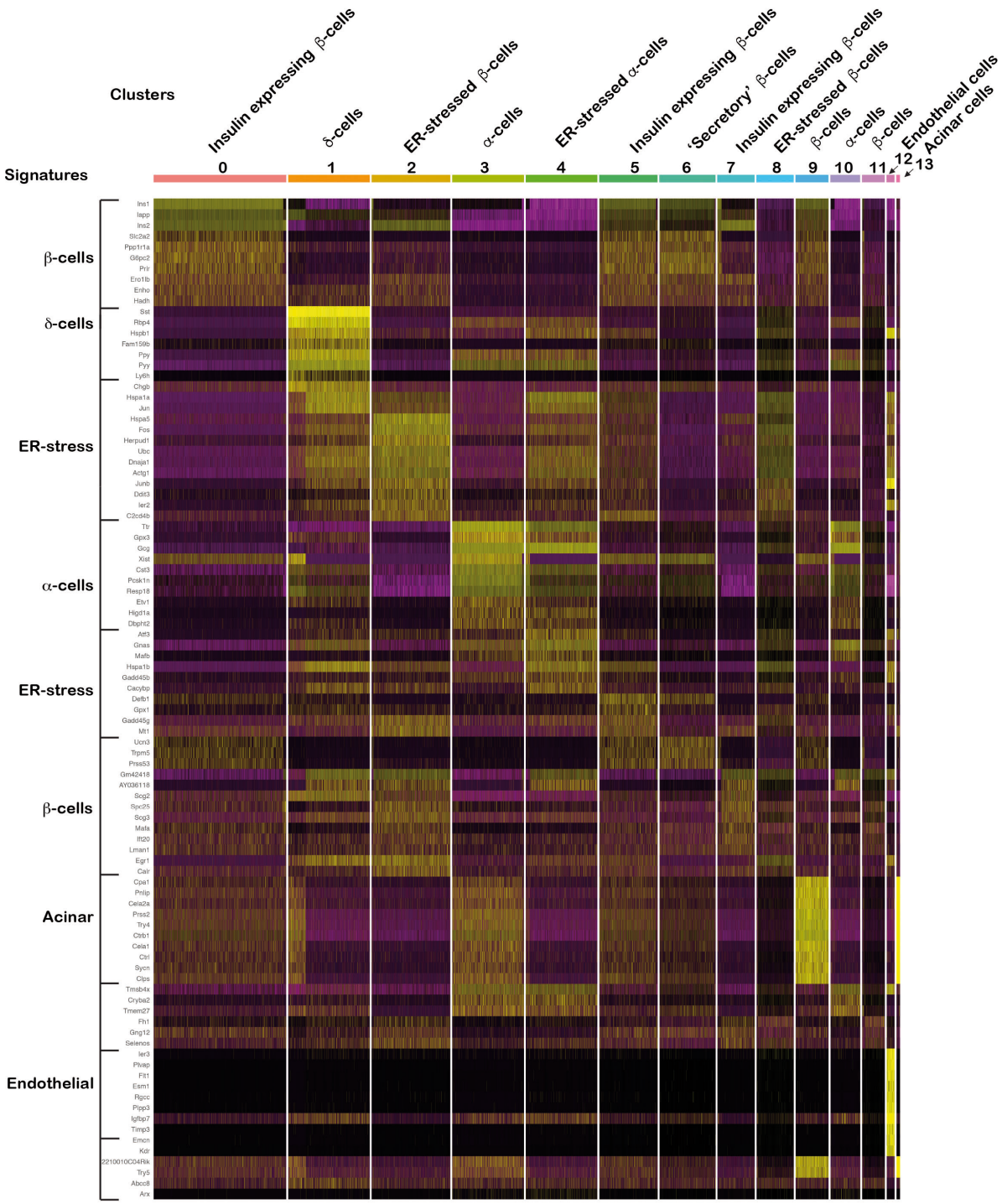


Kleiber et al., Fig. S3

Supplemental Figure 3. Changes in chromatin accessibility in isolated islets 1 and 5 weeks after Taf4 inactivation measures by ATAC-seq. **A-B**. Scatterplots showing Log10 changes in ATAC-seq peaks either globally (left panels) or at the proximal promoter (-1000/+100 base pairs relative the transcription start site, TSS, right panel) at 1 week (A) or 5 weeks (B) following TAF4 inactivation. The number of peaks enriched in each condition are indicated. **C-D**. In silico footprinting using Tobias. Scatter plots of differential footprinting depth of transcription factor motifs in differentially accessible regions between WT and week 1 (C) and week 5 (D). Motifs with more footprint depth in WT are labelled in pink and in Taf4 mutants in light blue.

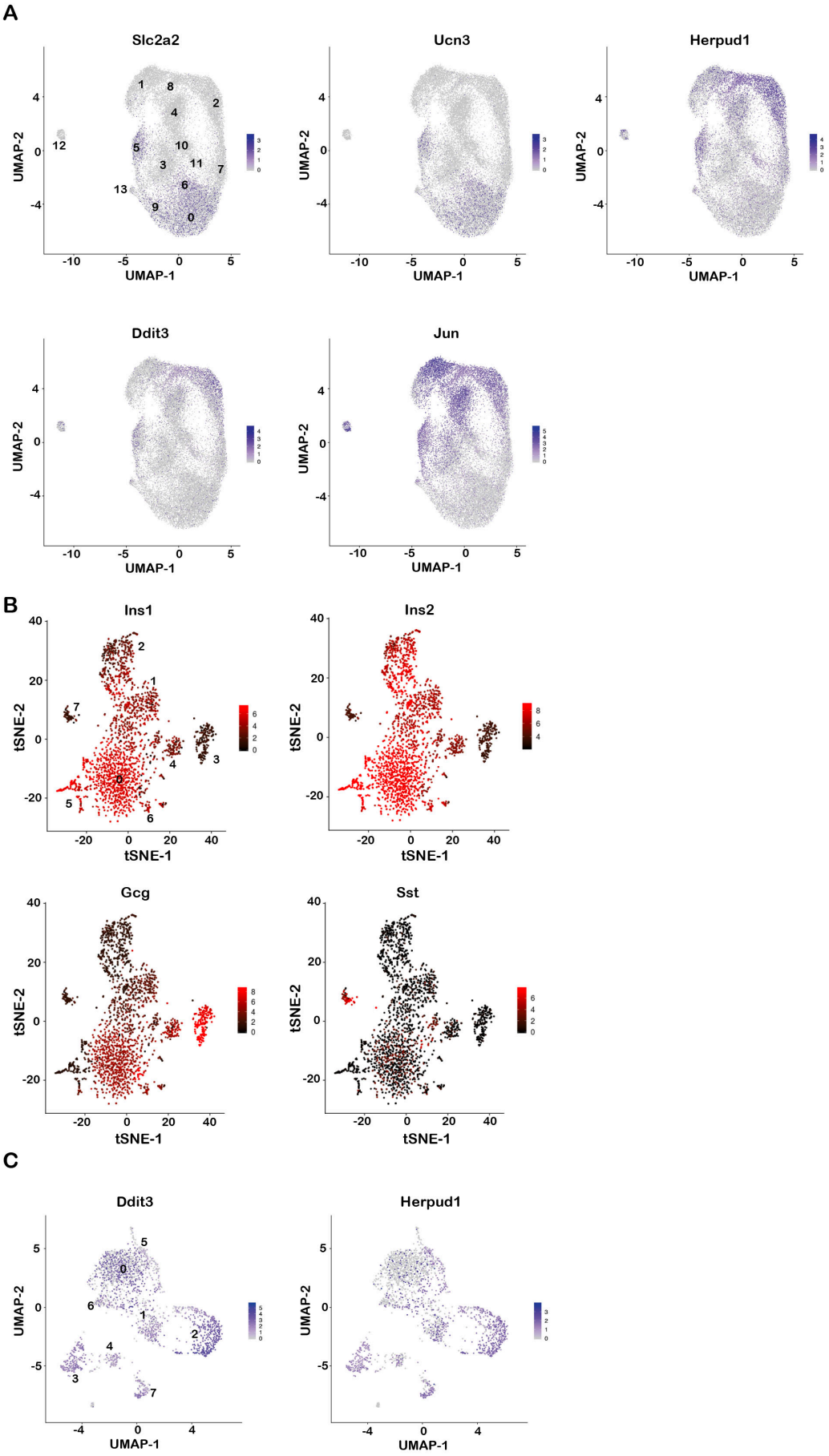


Supplemental Figure 4. Heatmap of differentially expressed genes in the cell clusters from scRNA-seq of WT islets.



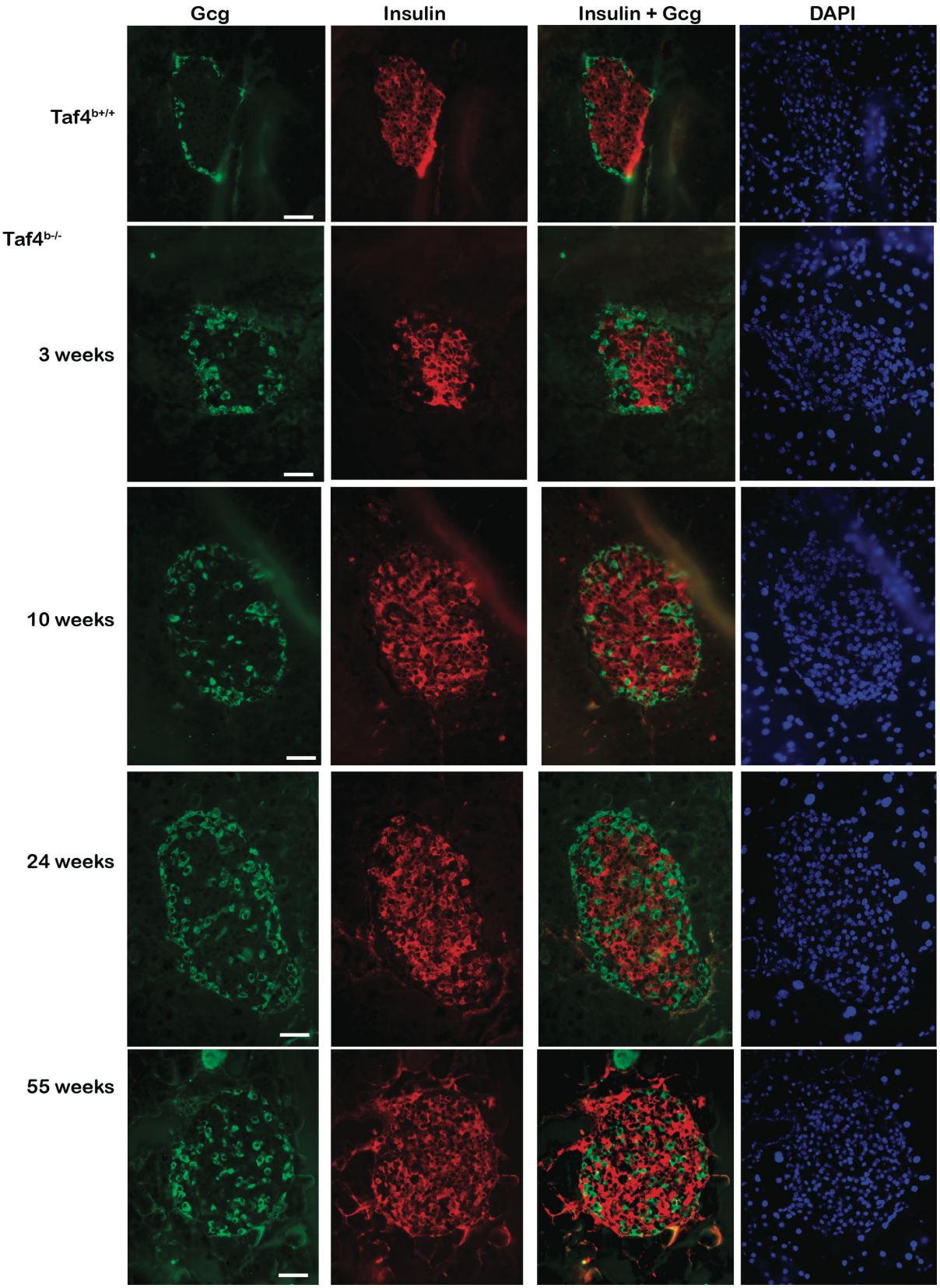
Kleiber et al., Fig. S5

Supplemental Figure 5. Heatmap of differentially expressed genes in the cell clusters from the WT/week 5 aggregate data.



Kleiber et al., Fig. S6

Supplemental Figure 6. A. UMAP representations of the aggregate data illustrating the expression of the genes indicated in each panel. **B** tSNE representations of the week 1 cell populations illustrating the expression of the genes indicated in each panel. **C**. UMAP representations week 1 cell populations illustrating the expression of the genes indicated in each panel.



Kleiber et al., Fig. S7

Supplemental Figure 7. Immunostaining of isolated Langerhans islets from mice with the indicated genotypes for Insulin, Gcg or DAPI as indicated. The number of weeks after Tam injection are indicated. The Insulin-Gcg merge is shown to illustrate the mutually exclusive nature of the labelling. Scale bar = 100 μM