1 **Title: TET1 dioxygenase is required for FOXA2-associated chromatin remodeling in**
2 **pancreatic beta-cell differentiation**
3

4 **Authors:** Jianfang Li[1,2#], Xinwei Wu[1#], Minjung Lee[3], Jie Ke[1], Qingping Lan[1,4], Jia Li[3], Yun
5 Huang[3], De-Qiang Sun[2*], Ruiyu Xie[1,4*]

6 **Affiliations:**

7 [1] Centre of Reproduction, Development, and Aging, Faculty of Health Sciences, University of
8 Macau, Taipa, Macau SAR.

9 [2] Innovation centre for Advanced Interdisciplinary Medicine, The Fifth Affiliated Hospital of
10 Guangzhou Medical University, Guangzhou, 510530, China.

11 [3] Center for Epigenetics & Disease Prevention, Institute of Biosciences and Technology, College
12 of Medicine, Texas A&M University, Houston, TX 77030, USA

13 [4] MoE Frontiers Science Center for Precision Oncology, University of Macau, Taipa, Macau, SAR.

14 # These authors contributed equally to the work.

15 * Correspondence: ruiyuxie@um.edu.mo; deqiangs@zju.edu.cn

Abstract: Existing knowledge of the role of epigenetic modifiers in pancreas development has exponentially increased. However, the function of TET dioxygenases in pancreatic endocrine specification remains obscure. We set out to tackle this issue using a human embryonic stem cell (hESC) differentiation system, in which *TET1*/*TET2*/*TET3* triple knock-out cells display severe defects in pancreatic β-cell specification. Integrative whole-genome analysis identifies unique cell-type-specific hypermethylated regions (hyper-DMRs) displaying reduced chromatin activity and remarkable enrichment of FOXA2, a pioneer transcription factor essential for pancreatic endoderm specification. Intriguingly, TET depletion leads to significant changes in FOXA2 binding at pancreatic progenitor stage, in which gene loci with decreased FOXA2 binding features low levels of active chromatin modifications and enriches for bHLH motifs. Transduction of full-length *TET1* but not the TET1-catalytic-domain in *TET*-deficient cells effectively rescues β-cell differentiation accompanied by restoring *PAX4* hypomethylation. Taking these findings together with the defective generation of functional β-cells upon TET1-inactivation, our study unveils an essential role of TET1-dependent demethylation in establishing β-cell identity. Moreover, we discover a physical interaction between TET1 and FOXA2 in endodermal lineage intermediates, which provides a new mechanistic clue regarding the complex crosstalk between TET dioxygenases and pioneer transcription factors in epigenetic regulation during pancreas specification.

Introduction

During embryonic development, pluripotent human embryonic stem cells (hESCs) differentiate into many diverse lineages that make up the complex tissues and organs of the human body. Pancreatic lineage specification relies on crosstalk between the genome and environmental cues

38 in the progenitor niche. This crosstalk is mediated by cis-regulatory elements that play a prominent
39 role in spatiotemporal gene regulation during embryogenesis. In particular, distal regulatory
40 elements, such as enhancers, serve as information integration hubs that allow binding of multiple
41 regulators, including lineage-specific transcription factors (TFs) as well as epigenetic readers,
42 writers, and erasers to ensure integration of intrinsic and extrinsic environmental cues at these loci
43 [1,2]. It was recently demonstrated that pioneer TFs, such as FOXA1 and FOXA2, are required for
44 proper chromatin opening and establishment of enhancer marks H3K4me1 and H3K27ac during
45 pancreatic fate specification [3,4]. These pioneer TFs bind to nucleosomal DNA to initiate chromatin
46 remodeling associated with DNA demethylation at newly accessible enhancers [5–8].

47 DNA demethylation is mediated by ten-eleven-translocation methylcytosine dioxygenases
48 (TETs), which catalyze sequential oxidation of 5-methylcytosine (5mC) to 5-
49 hydroxymethylcytosine (5hmC), 5-formylcytosine, and 5-carboxylcytosine [9–13]. Distribution of
50 5hmC, a novel epigenetic modification, is dynamically changed by and positively correlated with
51 active gene transcription during early lineage specification [14–16]. Consequently, inhibition of TET
52 family enzymes (TET1, TET2, and TET3) impairs cell fate commitment into neural, hematopoietic,
53 cardiac, and several other lineages [17–19]. We previously demonstrated that 5hmC positively
54 correlates with 'open' chromatin at poised and active enhancers in multiple endodermal lineage
55 intermediates using a stepwise hESC differentiation system toward pancreatic progenitors [20].
56 However, the functional relevance and mechanisms by which TETs regulate pancreatic endocrine
57 specification are currently unclear. While it is recently recognized that pioneer TF FOXA2 plays
58 a critical role in enhancer priming during hESC pancreatic differentiation [3,4,21], there is still scarce
59 understanding of whether and how FOXA2 interacts with TETs to mediate opening of surrounding
60 chromatin.

61

62 **Results**

63 **TET deficiency impairs pancreatic endoderm differentiation.** To determine the biological
64 significance of TET-dependent regulation during pancreas specification, we generated *TET1*,
65 *TET2*, and *TET3* single knock-out (KO), double knock-out (DKO), and triple knock-out (TKO)
66 H1 hESC lines using CRISPR/Cas9 technology (Supplementary Fig. 1a). Mutations resulted in
67 premature termination codons, which were confirmed by Sanger sequencing. Global DNA
68 hydroxymethylation levels from positive clones were assayed by 5hmC dot blot. A significant
69 reduction of 5hmC signals was shown in TKO, TET1KO, TET1/2DKO, and TET1/3DKO cells,
70 whereas TET2KO, TET3KO, and TET2/3DKO cells showed minimal alterations of 5hmC levels
71 (Supplementary Fig. 1b). To avoid functional redundancy, we focused our initial analysis on TKO
72 cells devoid of any TET-mediated active demethylation. Consistent with other *TET* knock-out
73 mESC and hESC lines [17,22], TKO H1-hESCs exhibited no apparent defects in stem cell self-
74 renewal capacity or expression of pluripotent factors (Supplementary Fig. 1e).

75 To determine whether TET proteins affect hESC differentiation toward pancreatic endocrine
76 fate, we used an established stepwise differentiation platform [20] to induce efficient differentiation
77 of hESCs to definitive endoderm (DE), primitive gut tube (GT), pancreatic progenitor (PP), and
78 pancreatic endocrine (PE) (Fig. 1a). Both TET triple-deficient lines (clones 2 and 6) displayed
79 similar differentiation efficiency toward endoderm germ layer as the wild-type (WT) hESC line,
80 in which over 90% of cells were $SOX17^{+}$ by day 3 of differentiation (Supplementary Fig. 1c).
81 Moreover, expression levels of other endoderm markers, such as *FOXA2*, *FOXA1*, and *CXCR4*,
82 were unchanged in TET-depleted cells compared with control cells (Supplementary Fig. 1d, e).

83 This analysis suggests that TET dioxygenases are dispensable for endoderm specification in the
84 context of *in vitro* hESC differentiation. To examine the effects of TET ablation on pancreatic
85 commitment, we subsequently examined the expression of critical pancreatic markers at the PP
86 and PE stages. Using flow cytometry to quantitate the expression of *PDX1* at the PP stage and co-
87 expression of *PDX1* and *NKX6.1* at the PE stage, we observed a substantial decrease in TET knock-
88 out cells (Fig. 1b). These data were confirmed by immunofluorescence staining and RT-qPCR
89 analysis (Fig. 1c, Supplementary Fig. 1e). Consistent with these findings, expression of the
90 pancreatic progenitor markers *SOX9* and *PTF1A* and endocrine hormones insulin, and glucagon
91 were significantly downregulated in TKO lines (Fig. 1c, Supplementary Fig. 1e). Interestingly,
92 *PAX4*, a key determinant for β-cell specification, failed to induce in TET-inactivated cells, whereas
93 expression of the α-cell determinant, *ARX*, was not inhibited (Supplementary Fig. 1e). These
94 results demonstrate that TETs are required for pancreatic β-cell lineage specification.

95 To investigate TET-dependent global transcriptional changes during pancreatic endocrine cell
96 fate commitment, we performed transcriptome analysis of WT and TKO cells at the ES, DE, GT,
97 and PP stages. Principle component analysis of transcriptome data demonstrated that TET-
98 deficient cells were similar at the ES, DE, and GT stages but differed substantially at the PP stage
99 (Fig. 1d). Hierarchical clustering of genes specifically expressed at each stage revealed a
100 remarkable downregulation of PP-specific genes in TKO_PP cells, whereas ES-specific and DE-
101 specific genes showed no obvious changes in TET-deficient cells compared with WT cells
102 (Supplementary Fig. 1f). In particular, 1,173 differentially expressed genes (DEGs) were identified
103 in TKO_PP cells (Fig. 1e, Supplementary Table 1). Among them, there were 921 downregulated
104 DEGs in TET-deficient cells, including many important developmental determinants for
105 pancreatic endocrine specification, such as *PDX1*, *NKX2.2*, *NKX6.1*, *NKX6.2*, *NEUROG3*, and

106 *PAX4*. We also observed 252 genes, particularly HOX family members, which were significantly
107 upregulated upon TET deletion, suggesting that failure to differentiate into pancreatic β-cells is
108 tightly linked to an aberrant TF network. In addition, KEGG pathway analysis of the DEGs showed
109 enrichment for terms associated with maturity-onset diabetes of the young (MODY) and insulin
110 secretion (Fig. 1f), implicating that loss of TETs impairs pancreatic endocrine formation at later
111 stages of endoderm specification.

112 **FOXA2 binding is enriched at distal regulatory elements displaying decreased accessibility**
113 **upon TET deletion**. Given that our previous study demonstrated that TET-mediated cytosine
114 oxidation is strongly associated with open chromatin regions during pancreatic differentiation [20],
115 we used the assay of transposase-accessible chromatin followed by high-throughput sequencing
116 (ATAC-seq) to evaluate chromatin accessibility landscapes in WT and TET-deficient cells. We
117 assessed differential accessible regions (DARs) between WT_PP and TKO_PP cells from a total
118 of 124,322 identified ATAC-seq peaks and found that 4.3% of ATAC-seq peaks significantly
119 changed upon TET deletion (Supplementary Table 2). We found that loss of TET led to an overall
120 reduction in chromatin accessibility, with 5,060 DARs showing reduced accessibility and only 290
121 DARs showing increased accessibility in TKO_PP cells (Supplementary Fig. 1g). Moreover,
122 DARs with decreased accessibility were located primarily at distal regions (> 1 kb transcription
123 start site (TSS)) (Fig. 1g). Decomposition of chromatin accessibility and DEGs in TKO_PP cells
124 revealed that downregulated genes became less accessible (Fig. 1h).

125 To enhance the biological insights of TET depletion-mediated changes in chromatin
126 accessibility, we analyzed the presence of TF binding motifs within DARs and found that regions
127 with reduced accessibility were mostly enriched for motif of pioneer TF FOXA2 essential for
128 pancreas organogenesis and chromatin remodeling [4,21,23,24] (Supplementary Fig. 1h). Since the

129 presence of a binding motif is not necessarily reflected of actual binding, we mapped FOXA2-
130 binding sites by chromatin immunoprecipitation followed by next-generation sequencing (ChIP-
131 seq) in pancreatic progenitors. Consistent with the motif prediction our analysis revealed that
132 FOXA2 peaks coincide with ATAC-seq peaks preferentially decreased upon TET depletion (Fig.
133 1i). Collectively, our data indicate that failure to differentiate to pancreatic β-cells is likely to be
134 linked with the loss of chromatin accessibility at distal regulatory elements enriched for pioneer
135 TF FOXA2.

136 **Differentiation-specific DMRs feature reduced chromatin activity at pioneer TF binding**
137 **sites**. To investigate TET deficiency-mediated alterations in global DNA methylation, we
138 performed whole-genome bisulfite sequencing (WGBS) in WT_PP and TKO_PP cells. Over 86%
139 of sequencing reads were uniquely aligned to hg38 with a high sequencing depth of 24 × on CpG
140 dinucleotides and typical bimodal distribution of methylation ratio in each sample (Supplementary
141 Fig. 2a, b). The methylation ratio (5mC/C) was depleted at TSS but enriched across gene coding
142 regions, especially within TKO_PP cells (Supplementary Fig. 2c). Among 26.6 million CpG sites
143 detected in both WT_PP and TKO_PP cells (depth ≥ 5), a total of 251,658 differentially methylated
144 cytosines (DMCs; credible difference of methylation ratio > 20%) were identified (Fig. 2a).
145 Strikingly, 97.5% of DMCs were hypermethylated (hyper-DMCs), suggesting that TET inhibition
146 results in a pronounced gain of methylation during pancreas differentiation. We observed
147 enrichment of DMCs primarily at intergenic regions and introns (Supplementary Fig. 2d), whereas
148 substantial enrichment of hypo-DMCs was also found in repeat elements, such as long interspersed
149 elements, consistent with a suggested connection between hypomethylation and activation of
150 transposable elements [25,26]. Based on an established link between DNA methylation and
151 transcription regulation, we further calculated changes in methylation levels on cis-regulatory

152 elements previously identified in pancreatic progenitors [3,27]. We found increased methylation at
153 bivalent promoters, active enhancers, and, to a lesser extent, poised enhancers (Fig. 2b). Notably,
154 hypermethylation was preferentially enriched at open chromatin (ATAC-seq) as well as FOXA2,
155 GATA4 (GSE117136) [4], GATA6 (GSE117136) [4], and PDX1 (GSE117136) [4] distal binding regions
156 (Fig. 2b, Supplementary Fig. 2e). This analysis demonstrates that active regulatory elements are
157 hypermethylated upon TET depletion during pancreatic differentiation.

158 Our previous studies reveal that DNA demethylation is correlated with pancreatic endocrine
159 patterning [20]. To gain better insight into the role of TET-mediated methylation changes in
160 pancreatic differentiation, we analyzed differentially methylated regions (DMRs) by connecting at
161 least three consecutive DMCs [28] between TKO_PP and WT_PP cells. We found a total of 16,490
162 hyper-DMRs and classified them into two categories based on their 5mC levels in pancreatic
163 progenitor versus hESC [29] (Fig. 2c). Notably, more than half (n = 10,254) of hyper-DMRs
164 exhibited decreased 5mC in WT_PP cells than in hESCs indicating an active demethylation
165 process during pancreatic differentiation (named differentiation-specific hyper-DMRs), whereas
166 the others (n = 6,236) showed similar methylation levels (non-differentiation hyper-DMRs) (Fig.
167 2c, Supplementary Fig. 2f). Interestingly, we found that pioneer TFs FOXA2, GATA4, GATA6,
168 and PDX1 were predominantly enriched at differentiation-specific hyper-DMRs (Fig. 2d),
169 suggesting that TET-dependent demethylation during pancreatic cell fate commitment occurs at
170 loci primarily associated with lineage-specific TFs. We further compared ATAC-seq peaks at
171 differentiation-specific hyper-DMRs with those at non-differentiation hyper-DMR. We found that
172 chromatin accessibility was markedly reduced in TET-deficient cells at distally located hyper-
173 DMRs in the differentiation-specific group (Fig. 2e, Supplementary Fig. 2g). Consistently,
174 genomic loci displaying decreased accessibility showed a substantial increase in DNA methylation

175 upon depletion of TET (Fig. 2f). For instance, the hyper-DMRs, which were annotated to *PDX1*
176 and its lncRNA *PDX1-AS1* loci, overlapped with pioneer TF-binding sites and showed decreased
177 chromatin activity in a TET-dependent manner (Fig. 2g). Significant hypermethylation was also
178 found at the distal regulatory elements of *NEUROD1* and *NKX2.2* loci (Supplementary Fig. 2h).
179 In summary, these analyses demonstrate that TET-dependent demethylation at TF-enriched distal
180 regulatory elements is essential for chromatin remodeling during pancreas development.

181 **TET deficiency induces a differentiation-specific loss of 5hmC at FOXA2 target sites**. To get
182 additional insight into the TET-mediated hydroxymethylation network governing pancreatic
183 differentiation, we used cytosine-5-methanesulfonate immunoprecipitation (CMS-IP) coupled
184 with high-throughput sequencing to profile 5hmC landscapes in WT_PP and TKO_PP cells. As a
185 consequence of TET deletion, we found genome-wide loss of hydroxymethylation in TET-
186 deficient cells (Supplementary Fig. 3a, b). As expected, 94% of CMS-positive regions showed a
187 significant reduction in TKO_PP cells relative to WT_PP cells (Supplementary Fig. 3c). These
188 differentially hydroxymethylated regions displaying reduced 5hmC (hypo-DHMRs) were
189 primarily located at non-promoter regions (> 1 kb from TSS) (Supplementary Fig. 3d).

190 Reduced DNA hydroxymethylation in TET-deficient cells could be a result of the initial loss
191 of 5hmC at the ES stage or unsuccessful oxidation of 5mC during lineage progression. To provide
192 biological insight into differentiation-specific changes of hydroxymethylation upon TET
193 inactivation, we systematically analyzed hypo-DHMRs overlapped with previously identified
194 regions showing de-novo increased 5hmC during hESC pancreatic differentiation [20]. Specifically,
195 27,113 hypo-DHMRs with higher 5hmC levels in WT_PP than in WT_ES (named 'differentiation-
196 specific hypo-DHMRs') were first isolated from a total of 70,857 TKO hypo-DHMRs (Fig. 3a).
197 We subsequently clustered them into four categories based on the dynamic changes of 5hmC

198 across stages in WT cells (Fig. 3b). In particular, one group gained 5hmC specifically at the PP
199 stage (PP-specific), whereas others gained 5hmC at the earlier DE (DE-to-PP) or GT (GT-to-PP)
200 stages and sustained hydroxymethylation until the PP stage. We found that GATA motifs were
201 mostly associated with the DE-to-PP cluster, in which gain of 5hmC began at the DE stage (Fig.
202 3c, purple).

203 By contrast, the PP-specific cluster displayed a prominent presence of binding sites for HNF6,
204 which is critical for pancreatic endocrine differentiation [30] (Fig. 3c, red). Most notably, FOXA
205 motifs were predominantly associated with the GT-to-PP(h) cluster in which 5hmC levels
206 increased at the GT stage and continued to be elevated until the PP stage (Fig. 3c, blue), whereas
207 both FOXA and GATA motifs were highly associated with the GT(h)-to-PP cluster in which 5hmC
208 peaked at the GT stage and subsequently decreased at the PP stage (Fig. 3c, brown). Collectively,
209 these results suggest a unique temporal binding pattern of GATAs, FOXAs, and HNF6s associated
210 with the dynamic distribution of 5hmC during pancreatic differentiation, which is supported by a
211 recent study demonstrating sequential requirements for these TFs during transitions in pancreas
212 development [4].

213 **5hmC and FOXA2 co-occupy target sites are essential for chromatin opening and enhancer**
214 **activation.** Given that hydroxymethylation is positively associated with chromatin activity [20] and
215 FOXA2 is strongly enriched in regions featured decreased chromatin accessibility in TET-
216 deficient cells (Fig. 1i), we wondered if binding of FOXA2 dynamically correlated with TET-
217 mediated hydroxymethylation. Thus, we mapped FOXA2-binding sites by ChIP-seq at the DE,
218 GT, and PP stages where FOXA2 continually expressed. Consistent with the motif predictions our
219 analysis revealed that FOXA2 deposition corresponded well with the dynamic changes of 5hmC
220 across differentiation stages in the four 'differentiation-specific hypo-DHMRs' clusters (Fig. 3d).

221 Together, these data support that FOXA2 binding is highly dynamic during pancreatic
222 differentiation and strongly prefers differentiation-associated 5hmC deposition sites.

223 It was previously demonstrated that the pioneer TF FOXA2 is essential for enhancer priming
224 during pancreatic differentiation [3,4]. We, thus, speculated that loss of TETs inhibits enhancer
225 activation particularly at hypo-hydroxymethylated regions enriched with FOXA2 binding. We
226 conducted ChIP-seq for enhancer signatures H3K4me1 and H3K27ac in WT_PP and TKO_PP
227 cells and performed integrated analysis. We found a remarkable decrease in ATAC-seq and
228 H3K27ac signal at 'differentiation-specific hypo-DHMRs' overlapped with FOXA2 peaks, where
229 DNA methylation was dramatically increased upon TET depletion (Fig. 3e). Nevertheless, FOXA2
230 and other pioneer TFs expressed at similar levels in TKO and WT cells (Supplementary Fig. 3e),
231 implicating that inhibition of TET did not alter their expression. Collectively, these data reveal that
232 TET deficiency results in a progressive failure of 5mC oxidation at a subset of FOXA2 targets
233 essential for the establishment of active enhancers during differentiation.

234 **FOXA2 transiently binds to differentiation stage-specific genomic loci enriched for bHLH**
235 **motifs.** Being a pioneer TF, FOXA2 is expected to engage its binding sites on nucleosomal DNA
236 to mediate nucleosome depletion and chromatin remodeling [6,31]. Despite its potentially universal
237 binding ability, FOXA2 targeting was highly dynamic between pancreatic differentiation states
238 (Fig. 3d). We, therefore, ask if additional factors can influence the binding of FOXA2. We tackled
239 these questions by mapping FOXA2 binding sites in WT and TKO cells at the DE and PP stages,
240 respectively. Consistent with previous reports [4,32], FOXA2 was primarily located at non-promotor
241 regions (Supplementary Fig. 4a) implicating that FOXA2 is mainly involved in transcription
242 regulation through distal regulatory elements. Upon TET depletion, 99.5% FOXA2 binding sites
243 did not show changes at the DE stage (Fig. 4a), while FOXA2 occupancy was dramatically

244 changed at the PP stage, in which 10% and 6% of FOXA2 target sites showed reduced and greater
245 FOXA2 binding, respectively (Supplementary Fig. 4b, Supplementary Table 3). Since DNA
246 methylation and hydroxymethylation status were significantly altered in TET-deficient cells (Fig.
247 2a, Supplementary Fig. 3c), we wondered whether increases in methylation or decreases in
248 hydroxymethylation contributed to differential recruitment of FOXA2 in TKO_PP cells. We first
249 assessed changes in FOXA2 binding in WT_PP versus TKO_PP cells at the hyper-DMRs or hypo-
250 DHMRs identified in TKO_PP cells. We found that ~75% regions showed no changes in FOXA2
251 binding (Fig. 4b). We then tested if only hypermethylated or hypo-hydroxymethylated FOXA2-
252 bound sites would be affected. We found that similar proportions of FOXA2 target sites displayed
253 reduced/greater of FOXA2 binding regardless of the presence or absence of hypermethylation (Fig.
254 4c). Similarly, hypo-hydroxymethylated FOXA2-bound sites did not show a preference in
255 gain/loss of FOXA2 binding compared to iso-hydroxymethylated FOXA2-bound sites (Fig. 4d).
256 Consistent with the ability of pioneer TFs to engage with inaccessible chromatin, our results
257 strongly demonstrate that DNA methylation/hydroxymethylation states do not interfere in FOXA2
258 binding.

259 To comprehensively characterize temporal patterns of FOXA2 recruitment at the FOXA2-
260 decreased binding sites, we clustered FOXA2 binding signal in WT cells at the DE, GT, and PP
261 stages. Three distinct patterns of FOXA2 occupancy were observed. Cluster I regions (149) were
262 bound by FOXA2 from DE to PP stage, cluster II regions (237) were FOXA2-bound at GT and
263 PP stages, while the most predominant group, cluster III (1888), displayed de novo FOXA2
264 occupancy in pancreatic progenitors (Fig. 4e). We then analyzed the annotations of nearby genes
265 with Genomic Regions Enrichment of Annotations Tool (GREAT) and found that cluster III
266 regions were enriched for terms of neuron and endocrine pancreas development (Fig. 4f). Further

267 examination of ATAC-seq, H3K27ac, and 5mC signals in three FOXA2-decreased clusters
268 revealed high levels of DNA methylation accompanied by low levels of enhancer activity and
269 chromatin accessibility in cluster III regions (Supplementary Fig. 4c). Taken together, our data
270 imply that a subset of FOXA2 transiently binds to differentiation stage-specific genomic loci with
271 low levels of active chromatin modifications at which additional lineage-specific TFs may be
272 required to facilitate FOXA2 binding and subsequent chromatin remodeling.

273 To identify potential TFs associated with differential FOXA2 binding we determined unique
274 DNA binding motifs within FOXA2-decreased, increased and stable sites. Interestingly, loci lost
275 FOXA2 binding, particularly in cluster III, were mostly enriched for basic-helix–loop–helix
276 (bHLH) motifs such as the pancreatic endocrine cell fate determinant NEUROD1 [33] (Fig. 4g,
277 Supplementary Fig. 4d). In contrast, regions gained FOXA2 were mostly enriched with motifs of
278 TEAD and GATA family members, and FOXA2-stable sites most prominently feature forkhead
279 and CTCF motifs (Fig. 4g). Notably, expression of the bHLH TFs including *NEUROD1*, *PTF1A*,
280 and *ASCL1* failed to be induced in TET-knockout cells (Supplementary Fig. 4e) which likely
281 results in a decrease of FOXA2 binding at genomic loci primarily associated with these TFs. In
282 summary, these data suggest that recruitment of FOXA2 to differentiation stage-specific sites is
283 genetically and epigenetically primed with the cooperation of additional lineage-specific TFs.

284 **TET1 is required in pancreatic β-cell specification**. Our findings suggest that TET inactivation-
285 induced aberrant methylation/hydroxymethylation at least in part contributes to defective β-cell
286 specification. To determine the essential role of each TET family member in pancreatic
287 differentiation, we further analyzed mutants with single (*TET1*, *TET2*, or *TET3*) or double (*TET1/2*,
288 *TET2/3*, or *TET1/3*) *TET* knock-out. All mutant lines were able to induce the expression of *PDX1*
289 and *NKX6.1* to levels comparable to WT (Supplementary Fig. 5a, b). However, only those retaining

290 intact *TET1* expression (TET2KO, TET3KO, and TET2/3DKO) displayed proper induction of
291 *PAX4* (Supplementary Fig. 5c). In comparison with *TET2*/*TET3* double-deletion, inhibition of
292 TET1 alone had significant effects on the formation of INS- and C-peptide (CPEP)-expressing β-
293 cells but not GCG-expressing α-cells (Fig. 5a, b).

294 To evaluate global transcriptome changes in response to *TET1* deletion, we performed RNA-
295 seq of TET1KO cells at the PP stage. Integrated analysis of WT_PP, TKO_PP, and TET1KO_PP
296 data revealed a total of 1,590 DEGs among the three lines (Supplementary Table 4). Specifically,
297 555 genes, including most pancreas developmental regulators such as *PDX1*, *SOX9*, *NKX2.2*,
298 *NKX6.1*, *NEUROD1*, and *PAX6*, were downregulated in TKO_PP cells but not in TET1KO_PP
299 cells (Fig. 5c), consistent with effective differentiation of TET1KO-hESCs into PDX1^{+} cells
300 (Supplementary Fig. 5a). However, 536 genes, including the β-cell fate determinants *PAX4*,
301 *NKX6.2*, and *FEV* [34–36], were significantly inhibited in TET1-deficient cells, implicating TET1 as
302 responsible for the activation of a subset of genes essential for β-cell identity. To further analyze
303 the functional consequences of TET1 inhibition, WT-, TET1KO-, and TKO-hESCs were
304 differentiated to the PE stage and subsequently engrafted into SCID-beige mice under the kidney
305 capsules (Fig. 5d). Glucose-stimulated human C-peptide secretion was determined 18 weeks post-
306 transplantation. Notably, mice engrafted with WT_PE cells produced substantial fasting C-peptide
307 in serum (1605 ± 527 pg/ml) and showed statistically significant glucose-stimulated C-peptide
308 secretion (2,358 ± 839 pg/ml). By contrast, mice transplanted with TET1KO_PE cells secreted
309 extremely low amounts of basal C-peptide (163 ± 122 pg/ml) and showed no response to glucose
310 stimulation (149 ± 161 pg/ml). In agreement with these functional results, excised WT_PE cell
311 grafts were highly composed of insulin^{+} β-cells (Fig. 5e). The TET1KO_PE cell grafts displayed
312 much less insulin content, and only the δ-cell hormone somatostatin was detected in TKO_PE cell

313 grafts. Collectively, these data demonstrate that loss of TET1 impairs β-cell specification and
314 maturation.

315 **Full-length TET1 is required for the *PAX4* enhancer to achieve a hypomethylated state**.
316 To further determine whether TET1 is responsible for β-cell specification, we restored the full-
317 length (*TET1FL*) as well as the *TET1*-catalytic-domain (*TET1CD*) expression, respectively, using
318 lentivirus-mediated gene transduction in TKO cells. In contrast to the catalytically inactive *TET1*
319 (*TET1CD*^{mut}) (H1672Y, D1674A), both *TET1FL*- and *TET1CD*-transduced cells showed a global
320 increase in 5hmC levels, confirming the replenishment of hydroxymethylation (Supplementary
321 Fig. 6a). Differentiation of the TKO-TET1FL line toward pancreatic endocrine fate led to robust
322 induction of β-cell determinants, including *PDX1*, *NKX6.1*, and *PAX4*, whereas transduction of
323 *TET1CD* only resulted in a slight increase of *NKX6.1* and *PAX4* (Fig. 6a, b, Supplementary Fig.
324 6c). FACS analysis and immunostaining further demonstrated that TET1CD-transduced cells
325 showed an effective generation of GCG-producing α-cells but not INS/CPEP-expressing β-cells
326 at the PP stage (Supplementary Fig. 6b, c), indicating that full-length TET1 is critical for β-cell
327 specification.

328 As TET1 depletion inhibited the induction of genes essential for β-cell specification (Fig. 5),
329 we reasoned that co-binding of TET1 and FOXA2 at distal regulatory elements modulates focal
330 methylation status and subsequent gene activation. To identify potential interaction between TET1
331 and FOXA2, we performed immunoprecipitation in TKO-TET1FL cells expressing a FLAG-
332 tagged *TET1FL* gene. Anti-FLAG antibodies co-precipitated endogenous FOXA2 in TKO-
333 TET1FL cells differentiated at the DE stage (Fig. 6c). In contrast, co-precipitation of other
334 endoderm proteins, including SOX17, GATA6, and FOXA1, was not observed (Supplementary
335 Fig. 6d). These data illustrate that TET1 physically interacts with FOXA2 and therefore, can be

336 recruited to FOXA2-bound loci to mediated demethylation during lineage induction.

337 Since the induction of *PAX4* was dramatically inhibited in TET1-deficient lines
338 (Supplementary Fig. 5b), we asked if TET1 would cooperate with FOXA2 to mediate focal
339 demethylation at the *PAX4* enhancer. We found that a putative *PAX4* enhancer (~4.0 kb upstream
340 of the TSS), where FOXA2 bound and displayed hypermethylation in TET-deficient cells (Fig. 6d,
341 pink area). We next examined whether the elimination of TET1 alone results in hypermethylation
342 at this site. Locus-specific methylation was determined using a glucosylation and digestion-based
343 method followed by qPCR analysis. Two independent pairs of primers (PAX4-P1 and PAX4-P2)
344 located within the FOXA2-binding site (Fig. 6d) were used to determine the percentage of
345 methylated and non-methylated cytosine in TET1KO, TET2/3DKO, TKO, and WT cells at the PP
346 stage. Notably, the percentage of 5mC was increased to 35–50% for TET1KO and 60–65% for
347 TKO compared with WT but remained unchanged for TET2/3DKO (Supplementary Fig. 6e, left
348 panel). A corresponding decrease in unmethylated cytosine content was observed, with values of
349 45–65% and ~35% of C in TET1KO and TKO samples, respectively (Supplementary Fig. 6e, right
350 panel). As a housekeeping control, we examined methylation content at a nearby region using the
351 PAX4-P3 primer, and as expected, no significant differences in 5mC and C were observed in any
352 samples. We then determined methylation contents at the same *PAX4* locus in TET1-replenished
353 TKO cells. Consistent with the results found in TET2/3DKO cells, no detectable differences in
354 5mC and C were found between TKO-TET1FL and WT samples (Fig. 6e). However, 5mC levels
355 were significantly higher with a value of 30-40% in TKO-TET1CD than TKO-TET1FL samples
356 (Fig 6e, top panel). Together with the insufficient restoration of *PAX4* in TKO-TET1CD cells (Fig.
357 6b), we conclude that full-length TET1 is responsible for the establishment or maintenance of a
358 hypomethylated state at the *PAX4* enhancer, which is essential for transcription activation of *PAX4*

359 and subsequent β-cells generation.

360

361 **Discussion**

362 In the present work, we dissected the roles of TET proteins in pancreatic endocrine commitment
363 based on a stepwise hESC differentiation system. We found that the loss of all three TET family
364 members significantly impaired the differentiation of pancreatic β-cells. Furthermore, we
365 discovered that locus-specific hypermethylation was associated with genes essential for β-cell
366 specification and maturation, such as *PAX4* [34], *PDX1* [37], and *NKX2.2* [38]. The reintroduction of
367 *TET1* in TET-deficient cells effectively reversed hypermethylation and restored the expression of
368 *PAX4*. We further demonstrated that TET1 functions as an upstream epigenetic regulator of *PAX4*
369 through direct recruitment by FOXA2 to a putative *PAX4* enhancer to preserve its unmethylated
370 status, thereby potentiating *PAX4* expression to adopt β-cell fate during endocrine lineage
371 commitment. Consistently, we observed striking increases in methylation at the *PAX4* enhancer in
372 TET1 knock-out cells but not in TET2/TET3 DKO cells, suggesting that TET1 epigenetically
373 regulates induction of the β-cell program in a locus-specific manner. Moreover, despite the
374 successful induction of *PDX1* and *INS* upon deletion of TET1 alone, mice receiving TET1KO cell
375 grafts showed a persistent defective insulin response to glucose, implying that TET1 is also
376 essential for β-cell maturation.

377 Whereas the specification of β-cells was strongly influenced by depletion of *TET1*, we did not
378 observe a significant inhibition on the expression of *ARX*, a critical α-cell fate determinant [39]. In
379 contrast to *PAX4*, no hyper-DMRs were identified within the *ARX* locus, where methylation levels
380 were nearly undetectable in pancreatic progenitors. Previous studies demonstrate that several CpG-
381 rich sites of *Arx*, including one site close to TSS and another site 2 kb upstream of TSS, are heavily

382 methylated in adult β-cells but not α-cells [40]. Moreover, pharmacological inhibition of DNA
383 methyltransferases, *Dnmts*, in pancreatic progenitors promotes α-cell specification [41], whereas
384 deletion of *Dnmt1* in β-cells results in demethylation and depression of *Arx* [40]. We thus speculate
385 that *ARX* is hypomethylated in pancreatic progenitors in a TET-independent manner. The *ARX*
386 locus maintains an unmethylated state when progenitors differentiate into α-cells, whereas it
387 becomes hypermethylated once cells commit to β-cell fate in the presence of DNMTs and other β
388 cell-specific regulators. In line with this hypothesis are findings that Nkx2.2 recruits Dnmt3a to
389 the *Arx* promoter to repress its expression in β-cells [42]. In the future, it will be interesting to explore
390 whether there are differences in the complete epigenetic landscapes of endocrine progenitors,
391 which subsequently differentiate into α- or β-cells.

392 TET dioxygenases are critical for lineage induction in a cell-type-specific manner [43]. How
393 TETs recognize lineage-specific regulatory elements and modulate chromatin remodeling during
394 pancreas development remains unknown. Here, we addressed these questions by performing
395 analyses integrating gene expression with multiple chromatin features, including DNA
396 methylation, hydroxymethylation, chromatin accessibility, and histone modifications of enhancers.
397 We found extensive hypermethylation in TET-deficient cells that differentiated to the pancreatic
398 progenitor stage. A significant portion of hyper-DMRs was distributed in a differentiation-specific
399 manner, in which they were enriched for the binding of FOXA2, a pioneer TF essential for
400 pancreatic endoderm differentiation, and showed remarkable decreases in chromatin activity upon
401 TET inactivation. It is noteworthy that we unveiled a unique TF binding pattern associated with a
402 stepwise increase of 5hmC during pancreatic differentiation. The presence of GATA, FOXA, and
403 HNF6 binding motifs in a temporal manner suggests that TET proteins are recruited to chromatin
404 by a set of lineage-specific TFs sequentially expressed during pancreatic differentiation.

405 During lineage specification, chromatin structure is dynamically changed between 'closed'
406 and 'open' states. Open chromatin regions such as primed and active enhancers are accessible for
407 TF and epigenetic modulator binding to initiate gene transcription. Proper chromatin remodeling
408 is believed to be at least partly triggered by pioneer TFs, which can directly bind to nucleosomal
409 DNA [44]. It was previously suggested that pioneer TF FOXA2 initiates chromatin remodeling and
410 enhancer priming during pancreatic differentiation [3,4]. Interestingly, our analyses revealed that 1)
411 changes in 5hmC mirror the dynamic binding of FOXA2 in cells differentiated from hESCs
412 through defined lineage intermediates toward pancreatic endocrine fate, 2) chromatin activity was
413 markedly decreased at the FOXA2 binding sites associated with hyper-DMRs in TET-deficient
414 cells, and 3) FOXA2 physically interacted with TET1, suggesting that FOXA2 recruits TET1
415 mediate lineage-specific demethylation at a subset of FOXA2 targets to influence chromatin
416 opening.

417 How pioneer TFs selectively recognize differentiation stage-specific target sites and mediate
418 chromatin remodeling is still mostly unknown. Our present study demonstrates that 1) de novo
419 FOXA2 recruitment to genomic loci of low levels of active chromatin modifications is inhibited
420 upon TET depletion, and 2) FOXA2-decreased binding sites are enriched with bHLH TFs, such
421 as NEUROD1 and PTF1A, which fail to be induced in TET-deficient cells. Interestingly, FOXA2
422 stable and continuous binding targets are largely absent of motifs except for the forkhead and
423 CTCF family motifs. Therefore, we speculate that FOXA2 transient binding is facilitated by
424 lineage-specific TFs in a differentiation-stage specific manner during pancreatic cell fate
425 commitment (Fig. 6f). Indeed, it was recently demonstrated that chromatin opening requires
426 cooperative binding of FOXA2 and additional TFs in close vicinity [32]. However, why co-binding
427 of TFs is necessary for initiating chromatin remodeling is unclear. Our findings suggest that de

428 novo binding of FOXA2 to lineage-appropriate sites may require by cell-type-specific TFs, where
429 chromatin modifiers such as TET1 are recruited to mediate subsequent chromatin opening. Further
430 investigation is warranted to determine whether and how lineage-specific TFs facilitate FOXA2
431 binding and synergize the recruitment of multiple chromatin remodeling machineries.

432

433 **Methods**

434 **Cell lines.** The human embryonic stem cell line H1 was obtained from WiCell Research Institute.
435 H1 cells were maintained on Matrigel (Corning) in mTeSR1 (STEMCELL Technologies) and
436 passaged every for 4-5 days using 0.5 mM EDTA and 3 mM NaCl. Mycoplasma detection was
437 regularly performed by the Stem Cell Core Facility at the University of Macau.

438 **Pancreatic differentiation.** All differentiation experiments were repeated at least three times with
439 three individual clones of the same phenotype. Pancreatic differentiation was performed as
440 previously described with slight modifications [45]. In brief, hESCs were dissociated with Accutase
441 (eBioscience) and seeded at a density of 19,000 cells/cm^2 on Matrigel in mTeSR1 supplemented
442 with 10 μM Rho-associated protein kinase inhibitor Y-27632 (Miltenyi Biotec). Upon reaching 95%
443 confluence, cells were exposed to differentiation medium with daily media feeding. Alternatively,
444 a suspension-based differentiation format was used as previously described [27]. Briefly, 5.5×10^6
445 cells were suspended in 5.5 ml mTeSR1 per well of 6-well ultra-low attachment plates (Coring
446 Costar) and cultured overnight in an orbital shaking CO_2 incubator (Eppendorf New Brunswick)
447 at 95 rpm. Cell aggregates were supplied with differentiation medium and continually rotated at
448 95 rpm with daily media feeding.

449 ES-to-DE (3 d): hESCs were exposed to 100 ng/ml Activin A (PeproTech) and 25 ng/ml Wnt-
450 3a (PeproTech) in basal medium-I containing MCDB 131 medium (Gibco), 1.5 g/l $NaHCO_3$, 1×
451 GlutaMAX (Gibco), 10 mM glucose, and 0.5% BSA for 1 day. For 2 additional days, cells were
452 cultured in basal medium-I further supplemented with 100 ng/ml Activin A.
453 DE-to-GT (3 d): DE intermediates were incubated in basal medium-I supplemented with 50
454 ng/ml FGF7 (PeproTech) for 3 days.
455 GT-to-PP (5 d): GT intermediates were cultured for 3 days in basal medium-II containing
456 MCDB 131 medium, 2.5 g/l $NaHCO_3$, 1× GlutaMAX, 10 mM glucose, 2% BSA, and 1:200 ITS-
457 X (Gibco), which was further supplemented with 50 ng/ml FGF7, 0.25 μM hedgehog inhibitor
458 SANT-1 (Sigma), 1 μM retinoic acid (Sigma), 100 nM BMP inhibitor LDN193189 (Stemgent),
459 and 200 nM PKC activator TPB (Millipore). After 3 days of culture, cells were treated for 2 days
460 with 2 ng/ml FGF7, 0.25 μM SANT-1, 0.1 μM retinoic acid, 200 nM LDN193189, and 100 nM
461 TPB in basal medium-II.
462 PP-to-PE (3 d): PP intermediates were differentiated in basal medium-III containing MCDB
463 131 medium, 1.5 g/l $NaHCO_3$, 1× GlutaMAX, 20 mM glucose, 2% BSA, and 1:200 ITS-X, which
464 was further supplemented with 0.25 μM SANT-1, 0.05 μM retinoic acid, 100 nM LDN193189, 1
465 μM T3 (3,3',5-Triiodo-L-thyronine sodium salt, Sigma), 10 μM ALK5 inhibitor II (Enzo Life
466 Sciences), 10 μM $ZnSO_4$, and 10 μg/ml heparin (Sigma).
467 **Animal experiments.** All animal experiments were approved by the University of Macau Animal
468 Ethics Committee. Immunocompromised SCID-beige mice were obtained from Charles River and
469 maintained under a 12-h light/dark cycle with free access to standard mouse diet. For
470 transplantation, d14 cell aggregates were further incubated in basal medium-III supplemented with
471 100 nM LDN193189, 1 μM T3, 10 μM ALK5 inhibitor II, 10 μM $ZnSO_4$, and 100 nM γ-secretase

472 inhibitor XX (Millipore) for 1 day. Cell aggregates (~5 × 10^6 cells) were then transplanted into 6-
473 week-old SCID-beige mice under the kidney capsule as previously described [46].
474 Glucose-stimulated human C-peptide secretion was assessed with mice 18 weeks post-
475 transplantation. Blood samples were collected after overnight fasting and 30 min following an
476 intraperitoneal injection of glucose (2 g/kg body weight). Human C-peptide levels in isolated
477 plasma were quantified using the STELLUX Chemi Human C-peptide ELISA kit (ALPCO
478 Diagnostics) according to the manufacturer's instructions.
479 **Generation of *TET* knock-out lines**. *TET* knock-out hESCs were generated using CRISPR/Cas9
480 technology. gRNAs (Supplementary Table 5) were designed to target the sequences encoding exon
481 7 of *TET1*, exon 3 of *TET2*, or exon 3 of *TET3* and cloned into pX330 vector (Addgene #42230)
482 as previously described [47]. Constructs containing validated gRNAs were electroporated together
483 with a vector expressing puromycin into hESCs using the P3 Primary Cell 4D-Nucleofector X kit
484 (Lonza) following the manufacturer's instructions. The electroporated cells were plated at ~2,000
485 cells/cm^2 and cultured on Matrigel in mTeSR1 supplemented with 10 μM Y-27632 for 2 days.
486 Successfully transfected cells were selected with 1 μg/ml puromycin in mTeSR^{TM}1 and allowed
487 to expand to form visible colonies from a single cell. Subsequently, clonal colonies were manually
488 picked and reseeded individually into 24-well plates. The amplified colonies were analyzed by
489 Sanger sequencing at targeted loci for the presence of Indel mutations.
490 **TET1 overexpression cell lines**. To generate overexpression constructs for TET1, the C-terminal
491 FLAG-tagged human full-length TET1 (TET1FL), TET1 catalytic domain (TET1CD), and TET1
492 catalytic inactive mutant (TET1CD^{mut}) were amplified from FH-TET1-pEF (Addgene #49792) or
493 pIRES-hrGFP II-mTET1 (Addgene #83569), and cloned into the *NotI* digested PCDH-CAG-
494 MCS-P2A-Puro vector (a kind gift from Dr. R. Xu, University of Macau) [48]. Lentivirus particles

495 were prepared as previously described [49]. For viral transduction, TKO-hESCs were grown on
496 Matrigel in mTeSR1 using 24-well plates and treated with 6 μg/ml polybrene for 15 min at 37°C.
497 Concentrated lentivirus particles (10 μl, 1×10^6 TU/ml) were added to cell culture and incubated
498 overnight at 37°C. On the next day, viral infection was repeated (30 μl, 1×10^6 TU/ml) to increase
499 transduction efficiency. Infected cells were cultured in mTeSR1 for 2 days and then exposed to 1
500 μg/ml puromycin for 10 days. The TET1 overexpression cell lines TKO-TET1FL, TKO-TET1CD,
501 TKO-TET1CD^{mut}, were amplified and frozen down.

502 **Co-immunoprecipitation.** Cells were lysed in ice-cold lysis buffer (20 mM Tris pH 7.5, 150
503 mM NaCl, 1 mM EDTA, 1% Triton X-100, 5mM 4-nitrophenyl phosphate di(Tris) salt, 2mM
504 Na2VO4, 0.5% sodium deoxylcholate, protease inhibitors), sonicated on Diagenode Bioruptor for
505 5 min (30 s ON and 30 s OFF), and then centrifuged at 12,000 rpm for 10 min at 4 °C. Supernatants
506 were incubated with pre-balanced ANTI-FLAG® M2 Affinity Gel (Sigma-Aldrich, A2220)
507 overnight at 4 °C. The beads were washed three times with wash buffer I (20 mM Tris pH 8.0, 0.3
508 M KCl, 10% glycerol, 1 mM EDTA, 1 mM DTT, 0.1% NP-40 and proteinase inhibitors) and two
509 times with wash buffer II (20 mM Tris pH 8.0, 0.1 M KCl, 10% glycerol, 1 mM EDTA, 1 mM
510 DTT and proteinase inhibitors). 4 × Laemmli sample buffer with β-mercaptoethanol was added to
511 washed beads, and then boiled for 10 min at 95 °C for western blotting analyses.

512 **Locus-specific detection of 5mC**. Detection of 5mC content at particular CCGG sites was
513 performed using the Epimark 5hmC and 5mC analysis kit (New England Biolabs) following the
514 manufacturer's instructions. Briefly, genomic DNA was extracted using the DNeasy Blood and
515 Tissue kit (Qiagen) followed by RNase treatment. RNase-treated DNA samples were incubated
516 with T4 β-glucosyltransferase at 37°C for 16 h. The glycosylated DNA was subsequently digested
517 with *MspI* or *HpaII* for 8 h at 37°C. Samples were treated with Proteinase K at 40°C for 30 min

518 and then at 95°C for 10 min to inactivate the enzymes. Site-specific methylation contents were
519 examined by RT-qPCR using the primers listed in Supplementary Table 6. The percentage of 5mC
520 and unmodified cytosine were calculated using the comparative C_t method.

521 **5hmC dot blot assay**. Genomic DNA was extracted using the DNeasy Blood and Tissue kit
522 (Qiagen) according to the manufacturer's instructions. DNA was denatured in 1 M NaOH
523 supplemented with 25 mM EDTA at 95°C for 10 min and then neutralized with 2 M ice-cold
524 ammonium acetate for 10 min. Two-fold serial dilutions of the DNA samples were spotted onto
525 nitrocellulose membrane. The air-dried membrane was fixed with UV irradiation (CL-1000 UV
526 crosslinker, Ultra-Violet Products), blocked with 5% non-fat milk, and incubated with a rabbit
527 anti-5hmC antibody (1:10,000, Active Motif) followed by an HRP-conjugated anti-rabbit antibody
528 (1:5,000, Jackson ImmunoResearch). Signal was visualized with SuperSignal West Pico PLUS
529 chemiluminescent substrate (Thermo Fisher Scientific). The same membrane was subsequently
530 stained with 0.02% methylene blue in 0.3 M sodium acetate to ensure equal loading of input DNA.

531 **RNA isolation for real-time quantitative PCR**. Total RNA was extracted using the RNeasy Plus
532 Mini kit (Qiagen) according to the manufacturer's instructions. cDNA was synthesized using the
533 PrimeScript RT reagent kit (Takara). Real-time quantitative PCR was performed in triplicate using
534 the SYBR Premix Ex Taq (Tli RNase H Plus) kit (Takara). The expression of *TBP* was used for
535 the normalization of mRNA expression. All primers used for RT-qPCR were listed in
536 Supplementary Table 6.

537 **Immunocytochemical analysis**. Differentiated cells were fixed in 4% paraformaldehyde for 30
538 min at room temperature, washed, and then permeabilized with 0.15% Triton X-100 at room
539 temperature for 1 h. Following blocking with 5% normal donkey serum (Jackson Immuno
540 Research Laboratories), samples were incubated with primary antibodies at 4°C overnight and then

541 appropriate secondary antibodies for 1 h at room temperature. Images were acquired using the
542 Zeiss Axio Observer microscope. For transplant grafts, tissues were fixed with 4%
543 paraformaldehyde overnight at 4°C, washed with PBS, and subsequently exposed to 30% sucrose
544 overnight at 4°C. Samples were mounted with Optimal Cutting Temperature Compound (Tissue-
545 Tek) and sectioned at 10 μm. Immunofluorescent staining was performed on cryosections as
546 described above. Antibodies used for immunofluorescent staining are listed in Supplementary
547 Table 7.

548 **Flow cytometry analysis.** Cells were enzymatically dissociated into single cells and fixed with 4%
549 paraformaldehyde for 20 min at 4°C. Fixed cells were permeabilized with Perm/Wash Buffer (BD
550 Biosciences) and stained with primary antibodies (Supplementary Table 7) diluted in Perm/Wash
551 Buffer overnight at 4°C. Cells were subsequently washed, stained with appropriate secondary
552 antibodies for 1 h at room temperature, and assessed using an Accuri C6 flow cytometer (BD
553 Biosciences). Data were analyzed using FlowJo software (TreeStar).

554 **RNA-seq library construction and data analysis.** The integrity of extracted RNA was
555 determined using an Agilent 2100 Bioanalyzer (Agilent Technologies). Subsequently, polyA-
556 tailed RNA was selected using Dynabeads oligo (dT) (Thermo Fisher Scientific), and libraries
557 were prepared using the NEBNext Ultra RNA Library Prep kit for Illumina (New England Biolabs).
558 Libraries were subjected to high-throughput sequencing on an Illumina HiSeq 2500 system (150
559 cycles, paired-end) at Novogene (Tianjin, China). To process the sequencing data, low-quality
560 bases and adaptor were trimmed using TrimGalore v.0.5.0 (options: --quality 20 and --length
561 50, https://github.com/FelixKrueger/TrimGalore). Clean reads were aligned to the hg38 genome
562 reference using STAR v.2.5.3 [50] with default parameters, and only uniquely mapped reads were
563 used for downstream analysis. A count matrix for each gene was generated using htseq-count

564 (HTSeq package [51]). DESeq2 [52] was used to identify significant DEGs in knock-out samples
565 compared with WT samples at different differentiation stages (|fold change| ≥ 2; False Discovery
566 Rate (FDR) < 0.05). Hierarchical cluster analysis of the union DEGs was used to determine stage-
567 specific signature genes. The 'ClusterProfilter' package in R was used for the functional
568 enrichment analysis of DEGs in KEGG pathways. BAM files were converted to bigwig files by
569 Bam2wig.py in RSeQC [53–55].

570 **ATAC-seq library preparation and data analysis**. ATAC-seq libraries were prepared as
571 previously described [56]. In brief, cells were enzymatically dissociated and lysed in lysis buffer (10
572 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPAL CA-630). Immediately after
573 centrifugation, transposition reactions were carried out by adding Tn5 transposes from the Illumina
574 Nextera DNA library preparation kit to the isolated nuclei and incubation at 37°C for 30 min. DNA
575 fragments were purified using the MinElute PCR Purification kit (Qiagen) and amplified using the
576 KAPA real-time library amplification kit (Roche). Libraries were purified using PCRClean DX
577 beads (Aline Biosciences) and subsequently subjected to high-throughput sequencing on an
578 Illumina NextSeq 500 instrument (75 cycles, paired-end).

579 Adaptor trimming of raw reads was performed using TrimGalore v0.5.0 (options: --quality 20
580 and --length 20), and high-quality (Q ≥ 20) reads were uniquely aligned to the hg38 genome
581 reference using Bowtie2 with the '--very-sensitive' option. Reads mapped to mitochondrial DNA
582 and PCR duplicate reads were removed, and uniquely mapped reads were extracted for
583 downstream analysis. Genrich v.0.5 (https://github.com/jsh58/Genrich) with ATAC-seq mode
584 (option: -j, -q 0.01) was applied for each sample (with two biological replicates) to call ATAC-seq
585 peaks. In total, 52,817 and 38,697 peaks were detected in WT_PP and TKO_PP cells, respectively.
586 BEDTools [57] intersect was used to generate a total of 124,322 non-overlapping peak regions, and

587 DeepTools multiBamSummary was applied to count the reads falling into peak regions. The counts
588 were used in DESeq2 for normalization and identification of significant differential accessible
589 regions (DARs) between WT_PP and TKO_PP cells by the criteria (|Fold change| ≥ 2; FDR <
590 0.05). Volcano plots were generated using the R package ggplot2 [58]. Motif annotation of DARs
591 was performed using HOMER software [59].

592 **CMS-IP-seq library preparation and data analysis.** CMS-IP-seq libraries were performed as
593 previously described with minor modifications (Huang et al., 2012). Purified genomic DNA (with
594 5% mouse DNA and 0.5% lambda DNA spike-in) was sheared to 200 - 500 bp fragments using a
595 M220 Focused-ultrasonicator (Covaris). Bisulfite-converted DNA libraries with methylated
596 adapters were enriched using an in-house anti-CMS antibody bound to protein A/G Dynabeads.
597 Amplified libraries were purified by AmpuXP beads (Beckman Coulter) and then sequenced using
598 the Illumina NextSeq (75 and 40 cycles, single-end) system.

599 Analysis of CMS-IP-seq data was performed by in-house software 'HaMiP' v0.1.1
600 (https://github.com/lijinbio/HaMiP). Briefly, raw reads were mapped to hg38 and spike-in mm10
601 genome references using bsmap v.2.89 (options: -n 1 -q 3 -r 0) [60]. After removing PCR duplicates
602 and reads mapped to both human and spike-in mouse genome, species-specific reads were used to
603 perform normalization for each sample according to the spike-in size factors. The whole human
604 genome was divided into 200-bp windows and the normalized mean wigsum in each window was
605 calculated to call the hydroxymethylation-enriched peak regions for each sample against input
606 control. In total, 75,324 and 503 peaks were detected in WT_PP and TKO_PP samples,
607 respectively. Differentially DHMRs between WT_PP and TKO_PP were identified (*G*-test; FDR
608 < 0.05). GREAT analysis with single-nearest genes option was used to perform functional
609 annotation of hypo-DHMRs.

610 CMS-IP-seq datasets of hESCs and multiple pancreatic lineage intermediates (DE, GT, and
611 PP) were downloaded from GSE97992 [20] and mapped to the hg38 genome reference using bsmap
612 with the same parameters as in the previous analysis. Bam2wig.py was used to transform the BAM
613 file to normalized bigwig files (option: -t 2000000000). The average 5hmC signals within each
614 hypo-DHMR for all stages were calculated using DeepTools [61]. Compared with hESCs, increased
615 5hmC peaks were defined as peaks in PP cells with a ≥ 1.5-fold increase.

616 **WGBS library preparation and data analysis**. Genomic DNA was isolated using the DNeasy
617 Blood & Tissue kit (Qiagen). Library preparation and high-throughput sequencing were conducted
618 by BGI (Shengzhen, China). In brief, purified genomic DNA (with 1% unmethylated lambda DNA
619 spike-in, Promega) was sheared to a fragment size of 100 - 700 bp (primary size 250 bp). Sheared
620 DNA was ligated with methylated adaptors (MGIEasy WGBS Adapters-16, MGI) and subjected
621 to bisulfite conversion using the EZ DNA Methylation-gold kit (Zymo Research). Bisulfite-
622 converted DNA was amplified with 13 PCR cycles and purified by AMPure XP beads (Beckman
623 Coulter). All libraries were sequenced on an MGISEQ-2000 system (100 cycles, paired-end).

624 For data analysis, raw reads were mapped against hg38 genome reference using bsmap v.2.89
625 [60] with paired mode (options: -n 1 -q 3 -r 0), and only uniquely mapped reads were retained. More
626 than 26.6 million CpG sites with coverage of ≥ 5 reads were detected in both WT_PP and TKO_PP
627 samples, which were used for downstream analyses. BSeQC [62] and the Mcall module in MOABS
628 [28] was applied to perform quality control and calculate the methylation ratio for each CpG site
629 (options: --trimWGBSEndRepairPE2Seq 40). Bisulfite conversion efficiencies were estimated
630 using spike-in unmethylated lambda phage DNA. The Mcomp module was used to call significant
631 differentially methylated CpG sites and regions (DMCs and DMRs, absolute credible difference
632 of DNA methylation ratio > 20% and FDR < 0.05). WGBS data of hESC was downloaded from the

633 ENCODE database [29], and the average 5mC level of each hyper-DMR in hESC was also estimated.
634 Detected hyper-DMRs (up- and down- 1 kb) were divided into 10-bp bins, and the average
635 methylation level of each bin was calculated for hESC, WT_PP and TKO_PP samples,
636 independently. The output bedGraph files from Mcall included single-base resolution DNA
637 methylation ratios, which were transformed into bigwig file format. Motif enrichment analysis of
638 DMRs was performed using HOMER software, functional annotation was performed using
639 GREAT with default settings, and many plots related to WGBS data was performed using in-house
640 software Mmint (https://github.com/lijiacd985/Mmint).

641 **ChIP-seq library preparation and data analysis**. ChIP-seq was performed for FOXA2 (WT_DE,
642 WT_GT, WT_PP, TKO_DE, and TKO_PP cells), H3K4me1 (WT_PP and TKO_PP cells), and
643 H3K27ac (WT_PP and TKO_PP cells). Chromatin immunoprecipitation was performed as
644 previously described [27]. In brief, ~1 × 10^7 cells were crosslinked, washed, and lysed in nuclear
645 lysis buffer (50 mM Trish HCl, pH 8.0, 5 mM EDTA, 1% SDS, and 1× protease inhibitor cocktail).
646 Chromatin was sheared to 200–500-bp fragments using Bioruptor (Diagenode), and DNA
647 fragments were precipitated with appropriate antibodies. ChIP-seq libraries were prepared using
648 the NEB Next Ultra II DNA Library Prep kit (New England Biolabs) following the manufacturer's
649 instruction and subjected to high-throughput sequencing on an Illumina HiSeq 2500 system (150
650 cycles, paired-end) at Novogene (Tianjin, China).

651 Quality control and alignment of raw reads of ChIP-seq data were performed similarly to
652 the ATAC-seq data analysis described above. TrimGalore (options: --quality 20 and --length 50)
653 was used to remove the adaptor. Bowtie2 with the '--very-sensitive' option was used for alignment,
654 and only uniquely mapped reads were retained. Bam2wig.py was used to transform the BAM file
655 to normalized bigwig files for visualization (option: -t 2000000000). For FOXA2 ChIP-seq, Macs2

656 [63] was used to call ChIP-seq enriched peak regions with default parameters for each replicate. The
657 BEDTools intersect was used to obtain the highly confident FOXA2 bound regions in two
658 biological replicates for each sample. Then BEDTools merge was used to merge the confident
659 peaks from WT and TKO samples to create the consensus FOXA2 bound regions at the DE and
660 PP stage, independently. The reads number was counted for consensus peaks in each replicate and
661 DESeq2 was used to call significant differential FOXA2 bound regions (FC≥ 2, FDR < 0.05).
662 Similar to ATAC-seq, motif analysis of FOXA2 bound regions was performed using HOMER
663 software and functional annotation was performed using GREAT.

664 **Integration of analyses.** RNA-seq, ATAC-seq, WGBS, CMS-IP-seq and ChIP-seq library
665 preparations were performed as previously described. Detail experimental procedures can be found
666 in Supplementary Materials. The number of mapped reads, mapped ratios and other statistic
667 information are listed in Supplementary Table 8. To compare DNA methylation between TKO_PP
668 and WT_PP cells at annotated genomic features, ChIP-seq datasets of GATA4, GATA6, and
669 PDX1 were downloaded from GSE117136. Bivalent promoters and poised and active enhancers
670 in hESC-derived pancreatic progenitors were downloaded from EMTAB1086 and GSE54471,
671 respectively. Data were analyzed as described above. The methylation difference between WT_PP
672 and TKO_PP across each genomic feature was performed by deepTools and Mmint software.

673 For integrated data visualization, bigwig files from RNA-seq, ATAC-seq, WGBS, CMS-
674 IP-seq and ChIP-seq experiments were uploaded to the University of California, Santa Cruz
675 (UCSC) genome browser. R package ggplot2 was used to plot volcano, bar, boxplot, point,
676 heatmap and scatter plots [58].

677 **Quantification and statistical analysis**. All statistical analyses were performed using GraphPad
678 Prism software. Student's unpaired two-tailed *t*-tests were used for qRT-PCR and flow cytometry

679 experiments. Student's paired one-tailed *t*-tests were used for the glucose-stimulated human C-
680 peptide secretion experiment. Quantification data are presented as mean ± SD. For all statistical
681 analyses, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$. Statistical analyses for RNA-
682 seq, WGBS, CMS-IP-seq, and ChIP-seq data are described in the corresponding sections.
683

684 **Data availability**

685 WGBS, CMS-IP, RNA-seq, ATAC-seq, and ChIP-seq data from this study were submitted to the
686 NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession
687 number GSE146486. All relevant data supporting the critical findings of this study are available
688 within the article and its supplementary information files or from the corresponding author upon
689 reasonable request.
690

691 **Acknowledgments**

692 We thank the University of Macau, Faculty of Health Sciences, Animal Research Facility for
693 animal housing. This work was supported by National Natural Science Foundation of China
694 (NSFC 31701276) and The Science and Technology Development Fund, Macau SAR
695 (0022/2019/AMJ) to R.X. The authors also thank the members of Macau Society for Stem Cell
696 Research (MSSCR) for inspiring discussion.
697

698 **Author contributions**

699 R.X. conceived the project. R.X. and D.S. directed and oversaw the project. X.W. and J. K.
700 performed experiments and collected data. J.F.L. performed computational analysis for WGBS,
701 RNA-seq, CMS-IP-seq, ATAC-seq, and ChIP-seq. M.L. prepared the CMS-IP library and

702 performed sequencing. Q.L. assisted with the animal studies. J.L. provided expert advice on
703 bioinformatic analyses. Y.H. critically reviewed the manuscript. R.X., J.F.L., X.W., and D.S.
704 wrote the manuscript; all other authors provided editorial advice.
705
706 **Competing interests**
707 The authors declare that no conflicts of interest exist.
708

709 **References**

710 1. Buecker, C. & Wysocka, J. Enhancers as information integration hubs in development:
711 Lessons from genomics. *Trends Genet.* **28**, 276–284 (2012).
712 2. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional
713 Enhancers in Development and Evolution. *Cell* **167**, 1170–1187 (2016).
714 3. Wang, A. *et al.* Epigenetic Priming of Enhancers Predicts Developmental Competence of
715 hESC-Derived Endodermal Lineage Intermediates. *Cell Stem Cell* **16**, 386–399 (2015).
716 4. Lee, K. *et al.* FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation.
717 *Cell Rep.* **28**, 382–393 (2019).
718 5. Cirillo, L. A. & Zaret, K. S. An Early Developmental Transcription Factor Complex that
719 Is More Stable on Nucleosome Core Particles Than on Free DNA. *Mol. Cell* **4**, 961–969
720 (1999).
721 6. Iwafuchi-Doi, M. *et al.* The Pioneer Transcription Factor FoxA Maintains an Accessible
722 Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Mol. Cell*
723 **62**, 79–91 (2016).
724 7. Donaghey, J. *et al.* Genetic determinants and epigenetic effects of pioneer-factor
725 occupancy. *Nat. Genet.* **50**, 250–258 (2018).
726 8. Geusz, R. J. *et al.* FoxA-dependent demethylation of DNA initiates epigenetic memory of
727 cellular identity. *bioRxiv* (2020). doi:10.1016/j.devcel.2021.02.005
728 9. He, Y. F. *et al.* Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG
729 in Mammalian DNA. *Science* **333**, 1303–1307 (2011).
730 10. Ito, S. *et al.* Role of tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and
731 inner cell mass specification. *Nature* **466**, 1129–1133 (2010).
732 11. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-
733 carboxylcytosine. *Science* **333**, 1300–1303 (2011).
734 12. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in
735 mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
736 13. Zhang, H. *et al.* TET1 is a DNA-binding protein that modulates DNA methylation and
737 gene transcription via hydroxylation of 5-methylcytosine. *Cell Res.* **20**, 1390–1393 (2010).
738 14. Tsagaratou, A. *et al.* Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-
739 cell development and differentiation. *Proc. Natl. Acad. Sci. USA* **111**, E3306–E3315
740 (2014).
741 15. Tan, L. *et al.* Genome-wide comparison of DNA hydroxymethylation in mouse embryonic
742 stem cells and neural progenitor cells by a new comparative hMeDIP-seq method. *Nucleic*
743 *Acids Res.* **41**, e84 (2013).
744 16. Kim, M. *et al.* Dynamic changes in DNA methylation and hydroxymethylation when hES
745 cells undergo differentiation toward a neuronal lineage. *Hum. Mol. Genet.* **23**, 657–667
746 (2014).
747 17. Verma, N. *et al.* TET proteins safeguard bivalent promoters from de novo methylation in
748 human embryonic stem cells. *Nat. Genet.* **50**, 83–95 (2018).
749 18. Ko, M. *et al.* Ten-eleven-translocation 2 (TET2) negatively regulates homeostasis and
750 differentiation of hematopoietic stem cells in mice. *Proc. Natl. Acad. Sci. USA* **108**,
751 14566–14571 (2011).
752 19. Fang, S. *et al.* Tet inactivation disrupts YY1 binding and long-range chromatin
753 interactions during embryonic heart development. *Nat. Commun.* **10**, 1–18 (2019).
754 20. Li, J. *et al.* Decoding the dynamic DNA methylation and hydroxymethylation landscapes

755 in endodermal lineage intermediates during pancreatic differentiation of hESC. *Nucleic*
756 *Acids Res.* **46**, 2883–2900 (2018).
757 21. Geusz, R. J. *et al.* A dual mechanism of enhancer activation by FOXA pioneer factors
758 induces endodermal organ fates. *bioRxiv* (2020).
759 22. Lu, F., Liu, Y., Jiang, L., Yamaguchi, S. & Zhang, Y. Role of Tet proteins in enhancer
760 activity and telomere elongation. *Genes Dev.* **28**, 2103–2119 (2014).
761 23. Rodríguez-Seguí, S., Akerman, I. & Ferrer, J. GATA believe it: New essential regulators
762 of pancreas development. *J. Clin. Invest.* **122**, 3469–3471 (2012).
763 24. Kaestner, K. H. The FoxA factors in organogenesis and differentiation. *Curr. Opin. Genet.*
764 *Dev.* **20**, 527–532 (2010).
765 25. López-Moyado, I. F. *et al.* Paradoxical association of TET loss of function with genome-
766 wide DNA hypomethylation. *Proc. Natl. Acad. Sci. USA* **116**, 16933–16942 (2019).
767 26. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses
768 is constrained by cytosine methylation. *Nat. Genet.* **20**, 116–117 (1998).
769 27. Xie, R. *et al.* Dynamic chromatin remodeling mediated by polycomb proteins orchestrates
770 pancreatic differentiation of human embryonic stem cells. *Cell Stem Cell* **12**, 224–237
771 (2013).
772 28. Sun, D. *et al.* MOABS: model based analysis of bisulfite sequencing data. *Genome Biol.*
773 **15**, 1–12 (2014).
774 29. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome.
775 *Nature* **489**, 57–74 (2012).
776 30. Zhang, H. *et al.* Multiple, temporal-specific roles for HNF6 in pancreatic endocrine and
777 ductal differentiation. *Mech. Dev.* **126**, 958–973 (2009).
778 31. Li, Z. *et al.* Foxa2 and H2A.Z Mediate Nucleosome Depletion during Embryonic Stem
779 Cell Differentiation. *Cell* **151**, 1608–1616 (2012).
780 32. Cernilogar, F. M. *et al.* Pre-marked chromatin and transcription factor co-binding shape
781 the pioneering activity of Foxa2. *Nucleic Acids Res.* **47**, 9069–9086 (2019).
782 33. Naya, F. J. *et al.* Diabetes, defective pancreatic morphogenesis, and abnormal
783 enteroendocrine differentiation in BETA2/NeuroD-deficient mice. *Genes Dev.* **11**, 2323–
784 2334 (1997).
785 34. Sosa-pineda, B., Chowdhury, K., Torres, M., Oliver, G. & Gruss, P. The Pax4 gene is
786 essential for differentiation of insulin-producing beta cells in the mammalian pancreas.
787 *Nature* **386**, 399–402 (1997).
788 35. Nelson, S. B., Schaffer, A. E. & Sander, M. The transcription factors Nkx6.1 and Nkx6.2
789 possess equivalent activities in promoting beta-cell fate specification in Pdx1+ pancreatic
790 progenitor cells. *Development* **134**, 2491–2500 (2007).
791 36. Ohta, Y. *et al.* Convergence of the insulin and serotonin programs in the pancreatic β-cell.
792 *Diabetes* **60**, 3208–3216 (2011).
793 37. Ahlgren, U., Jonsson, J., Jonsson, L., Simu, K. & Edlund, H. β-cell-specific inactivation
794 of the mouse Ipf1/Pdx1 gene results in loss of the β-cell phenotype and maturity onset
795 diabetes. *Genes Dev.* **12**, 1763–1768 (1998).
796 38. Sussel, L. *et al.* Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes
797 due to arrested differentiation of pancreatic β cells. *Development* **125**, 2213–2221 (1998).
798 39. Collombat, P. *et al.* Opposing actions of Arx and Pax4 in endocrine pancreas
799 development. *Genes Dev.* **17**, 2591–2603 (2003).
800 40. Dhawan, S., Georgia, S., Tschen, S. I., Fan, G. & Bhushan, A. Pancreatic β Cell Identity

801 Is Maintained by DNA Methylation-Mediated Repression of Arx. *Dev. Cell* **20**, 419–429
802 (2011).
803 41. Liu, J. *et al.* Neurog3-Independent Methylation Is the Earliest Detectable Mark
804 Distinguishing Pancreatic Progenitor Identity. *Dev. Cell* **48**, 49–63 (2019).
805 42. Papizan, J. B. *et al.* Nkx2.2 repressor complex regulates islet β-cell specification and
806 prevents β-to-α-cell reprogramming. *Genes Dev.* **25**, 2291–2305 (2011).
807 43. Wu, X., Li, G. & Xie, R. Decoding the role of TET family dioxygenases in lineage
808 specification. *Epigenetics and Chromatin* **11**, 1–10 (2018).
809 44. Cirillo, L. A. *et al.* Opening of compacted chromatin by early developmental transcription
810 factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–89 (2002).
811 45. Rezania, A. *et al.* Reversal of diabetes with insulin-producing cells derived in vitro from
812 human pluripotent stem cells. *Nat. Biotechnol.* **32**, 1121–1133 (2014).
813 46. Zmuda, E. J., Powell, C. A. & Hai, T. A method for murine islet isolation and subcapsular
814 kidney transplantation. *J. Vis. Exp.* **50**, e2096 (2011).
815 47. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**,
816 819–824 (2013).
817 48. Wang, X. *et al.* Noninvasive application of mesenchymal stem cell spheres derived from
818 hESC accelerates wound healing in a CXCL12-CXCR4 axis-dependent manner.
819 *Theranostics* **9**, 6112–6128 (2019).
820 49. Tu, S. *et al.* Takusan: A Large Gene Family that Regulates Synaptic Activity. *Neuron* **55**,
821 69–85 (2007).
822 50. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
823 (2013).
824 51. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-
825 throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
826 52. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
827 for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
828 53. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in
829 multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
830 54. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing
831 biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
832 55. Wang, L., Wang, S. & Li, W. RSeQC: Quality control of RNA-seq experiments.
833 *Bioinformatics* **28**, 2184–2185 (2012).
834 56. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for
835 Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **109**, 21–29
836 (2015).
837 57. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing
838 genomic features. *Bioinformatics* **26**, 841–842 (2010).
839 58. Villanueva, R. A. M. & Chen, Z. J. ggplot2: Elegant Graphics for Data Analysis. *Meas.*
840 *Interdiscip. Res. Perspect.* **17**, 160–167 (2019).
841 59. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime
842 cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**,
843 576–589 (2010).
844 60. Xi, Y. & Li, W. BSMAP: Whole genome bisulfite sequence MAPping program. *BMC*
845 *Bioinformatics* **10**, 1–9 (2009).
846 61. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. DeepTools: A flexible

847 platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, 187–191 (2014).
848 62. Lin, X. *et al.* BSeQC: Quality control of bisulfite sequencing experiments. *Bioinformatics*
849 **29**, 3227–3229 (2013).
850 63. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using
851 MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).
852

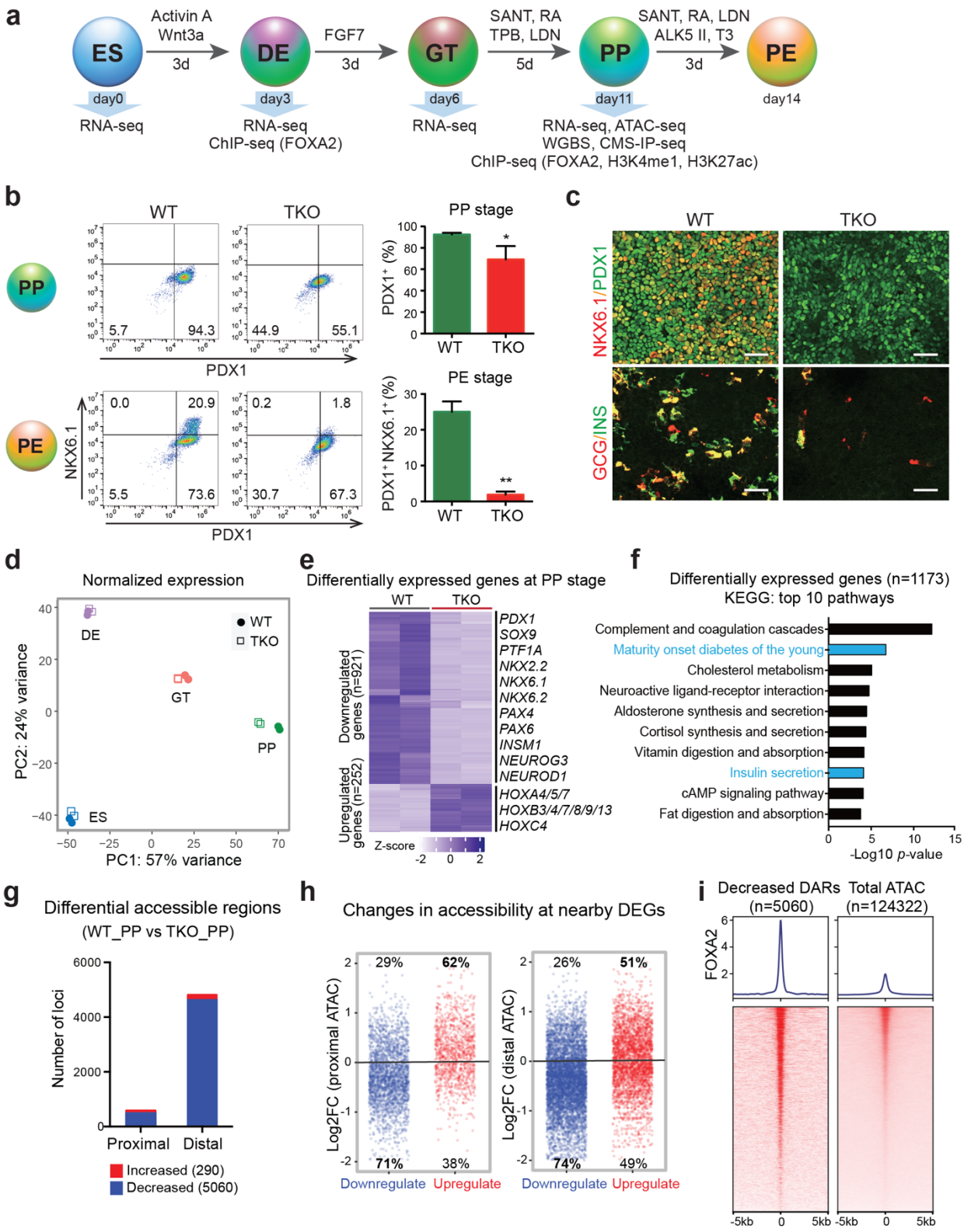
a
ES
Activin A
Wnt3a
3d
DE
FGF7
3d
GT
SANT, RA
TPB, LDN
5d
PP
SANT, RA, LDN
ALK5 II, T3
3d
PE
day0
RNA-seq
day3
RNA-seq
ChIP-seq (FOXA2)
day6
RNA-seq
day11
RNA-seq, ATAC-seq
WGBS, CMS-IP-seq
ChIP-seq (FOXA2, H3K4me1, H3K27ac)
day14
b
WT
TKO
PP
PE
PDX1
NKX6.1
PP stage
PE stage
PDX1+ (%)
PDX1+NKX6.1+ (%)
c
NKX6.1/PDX1
GCG/INS
d
Normalized expression
PC2: 24% variance
PC1: 57% variance
DE
GT
PP
ES
e
Differentially expressed genes at PP stage
Downregulated genes (n=921)
Upregulated genes (n=252)
PDX1
SOX9
PTF1A
NKX2.2
NKX6.1
NKX6.2
PAX4
PAX6
INSM1
NEUROG3
NEUROD1
HOXA4/5/7
HOXB3/4/7/8/9/13
HOXC4
Z-score
f
Differentially expressed genes (n=1173)
KEGG: top 10 pathways
Complement and coagulation cascades
Maturity onset diabetes of the young
Cholesterol metabolism
Neuroactive ligand-receptor interaction
Aldosterone synthesis and secretion
Cortisol synthesis and secretion
Vitamin digestion and absorption
Insulin secretion
cAMP signaling pathway
Fat digestion and absorption
-Log10 p-value
g
Differential accessible regions
(WT_PP vs TKO_PP)
Number of loci
Proximal
Distal
Increased (290)
Decreased (5060)
h
Changes in accessibility at nearby DEGs
Log2FC (proximal ATAC)
Log2FC (distal ATAC)
29%
62%
71%
38%
26%
51%
74%
49%
Downregulate
Upregulate
i
Decreased DARs (n=5060)
Total ATAC (n=124322)
FOXA2
-5kb
0
5kb

Fig. 1

854 **Fig. 1 TET hydroxylases are required for pancreatic endocrine lineage specification.**

855 **a** Stepwise differentiation of hESCs to pancreatic endocrine cells. Wild-type (WT) and
856 *TET1*/*TET2*/*TET3* triple knock-out (TKO) hESC lines were differentiated and analyzed by RNA-
857 seq, WGBS, CMS-IP-seq, ATAC-seq, and ChIP-seq at the indicated stages (ES: embryonic stem
858 cell; DE: definitive endoderm; GT: primitive gut tube; PP: pancreatic progenitor; PE: pancreatic
859 endocrine). **b** Representative flow cytometry plots for PDX1 and NKX6.1 in WT or TKO cells at
860 the PP and PE stages. Quantification of the percentage of $PDX1^+$ $NKX6.1^+$ cells is shown in the
861 right panel. n = 3 independent differentiation. **c** Immunostaining of PDX1/NKX6.1 at the PP stage
862 and insulin (INS)/glucagon (GCG) at the PE stage are shown in WT and TKO cells. Scale bar =
863 50 μm. **d** Principal component analysis showing variance in normalized transcriptome between
864 WT and TKO cells at the ES, DE, GT, and PP stage. Each plotted point represents one biological
865 replicate. **e** Genes with significant changes in expression in TKO cells relative to WT cells at the
866 PP stage. Differentially expressed genes (DEGs) are classified into down- and upregulated groups.
867 Each row corresponds to one individual gene and each column to a different biological replicate.
868 The color scale from white to blue represents Z-score normalized gene expression levels from low
869 to high ($|\text{fold change}| \geq 2$; FDR < 0.05). **f** Functional analysis of DEGs in TKO_PP cells showing
870 the top 10 KEGG pathways. Benjamini-Hochberg corrected *p*-values were used. **g** Number of
871 regions with significant changes in chromatin accessibility upon TET depletion at proximal (≤ 1
872 kb from TSS) and distal (> 1 kb from TSS) regions (FDR < 0.05). **h** Dot plots depicting the ratios
873 (WT_PP over TKO_PP) of ATAC-seq signals (y-axis) at proximal (left) or distal (right) regions
874 of DEGs that were downregulated (blue) or upregulated (red) in TKO_PP cells compared with
875 WT_PP cells. Values indicate the fraction of increased or decreased chromatin accessibility
876 regions in down- and upregulated genes, respectively. **i** Average density plots (top) and heatmaps

877 (bottom) of FOXA2 binding at decreased accessible regions (DARs) and total identified ATAC-
878 seq peaks (± 5 kb). Heatmaps are ranked by decreased FOXA2 binding. The color scale from white
879 to red represents the normalized signal from low to high.

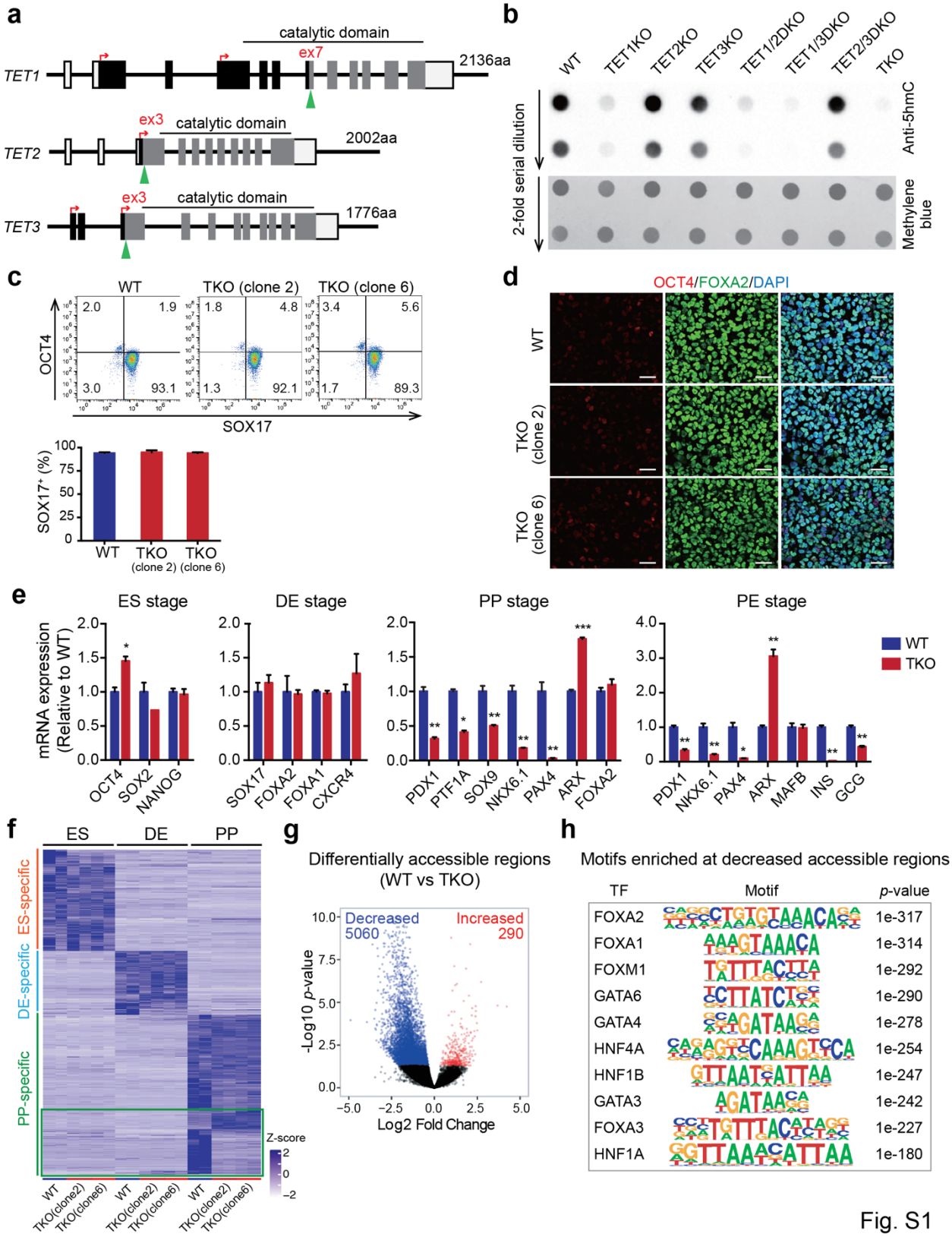
a
catalytic domain
ex7
TET1
2136aa
ex3
catalytic domain
TET2
2002aa
ex3
catalytic domain
TET3
1776aa
b
WT
TET1KO
TET2KO
TET3KO
TET1/2DKO
TET1/3DKO
TET2/3DKO
TKO
2-fold serial dilution
Anti-5hmC
Methylene blue
c
WT
TKO (clone 2)
TKO (clone 6)
OCT4
SOX17
2.0
1.9
3.0
93.1
1.8
4.8
1.3
92.1
3.4
5.6
1.7
89.3
SOX17+ (%)
WT
TKO (clone 2)
TKO (clone 6)
d
OCT4/FOXA2/DAPI
WT
TKO (clone 2)
TKO (clone 6)
e
ES stage
DE stage
PP stage
PE stage
mRNA expression (Relative to WT)
WT
TKO
OCT4
SOX2
NANOG
SOX17
FOXA2
FOXA1
CXCR4
PDX1
PTF1A
SOX9
NKX6.1
PAX4
ARX
FOXA2
PDX1
NKX6.1
PAX4
ARX
MAFB
INS
GCG
f
ES
DE
PP
ES-specific
DE-specific
PP-specific
Z-score
WT
TKO(clone2)
TKO(clone6)
g
Differentially accessible regions (WT vs TKO)
Decreased 5060
Increased 290
-Log10 p-value
Log2 Fold Change
h
Motifs enriched at decreased accessible regions
TF
Motif
p-value
FOXA2 1e-317
FOXA1 1e-314
FOXM1 1e-292
GATA6 1e-290
GATA4 1e-278
HNF4A 1e-254
HNF1B 1e-247
GATA3 1e-242
FOXA3 1e-227
HNF1A 1e-180

Fig. S1

881 **Supplementary Fig. 1 Generation and characterization of TET-deficient cell lines.**

882 **a** TET knock-out mutants were generated using CRISPR gRNAs (green arrowheads) targeting
883 exon 7 of *TET1,* exon 3 of *TET2*, and exon 3 of *TET3*. Translation start sites are indicated by red
884 arrows. **b** Analysis of global 5-hydroxymethylcytosine (5hmC) levels (top) in WT, *TET1* knock-
885 out (TET1KO), *TET2* knock-out (TET2KO), *TET3* knock-out (TET3KO), *TET1/TET2* double
886 knock-out (TET1/2DKO), *TET1/TET3* double knock-out (TET1/3DKO), *TET2/TET3* double
887 knock-out (TET2/3DKO), and *TET1/TET2/TET3* triple knock-out (TKO) hESCs by 5hmC dot blot
888 analysis. The bottom panel shows methylene blue staining using the total amount of input DNA
889 as the loading control. **c** Representative plots of flow cytometry for the expression of pluripotency
890 marker (OCT4) and endoderm marker (SOX17) at the DE stage are shown in the top panel.
891 Quantification of the percentage of $SOX17^{+}$ cells is shown in the bottom panel. n = 3 independent
892 differentiation. **d** Immunostaining of OCT4 and FOXA2 in WT and TKO cells at the DE stage.
893 Scale bar = 50 μm. **e** Expression analysis by RT-qPCR for specific genes in WT and TKO cells at
894 the ES, DE, PP, and PE stages. RT-qPCR validation was performed with three independent batches
895 of samples. **f** Heatmap showing stage-specific DEGs in TKO compared with WT cells at the ES,
896 DE, and PP stages ($|\text{fold change}| \geq 2$; FDR < 0.05). Two mutant lines were used for TKO (clones
897 2 and 6). Each column represents one biological replicate for each cell line. Genes located in green
898 rectangle are DEGs in TKO_PP compared with WT_PP cells. **g** Volcano plot of ATAC-seq data
899 illustrating differentially accessible regions in TKO_PP cells (FDR < 0.05). **h** Transcription factor
900 (TF) motif enrichment analysis of genomic regions showed significantly decreased ATAC-seq
901 signals in TKO_PP cells. The top 10 significant motifs are shown after removing redundant motifs.

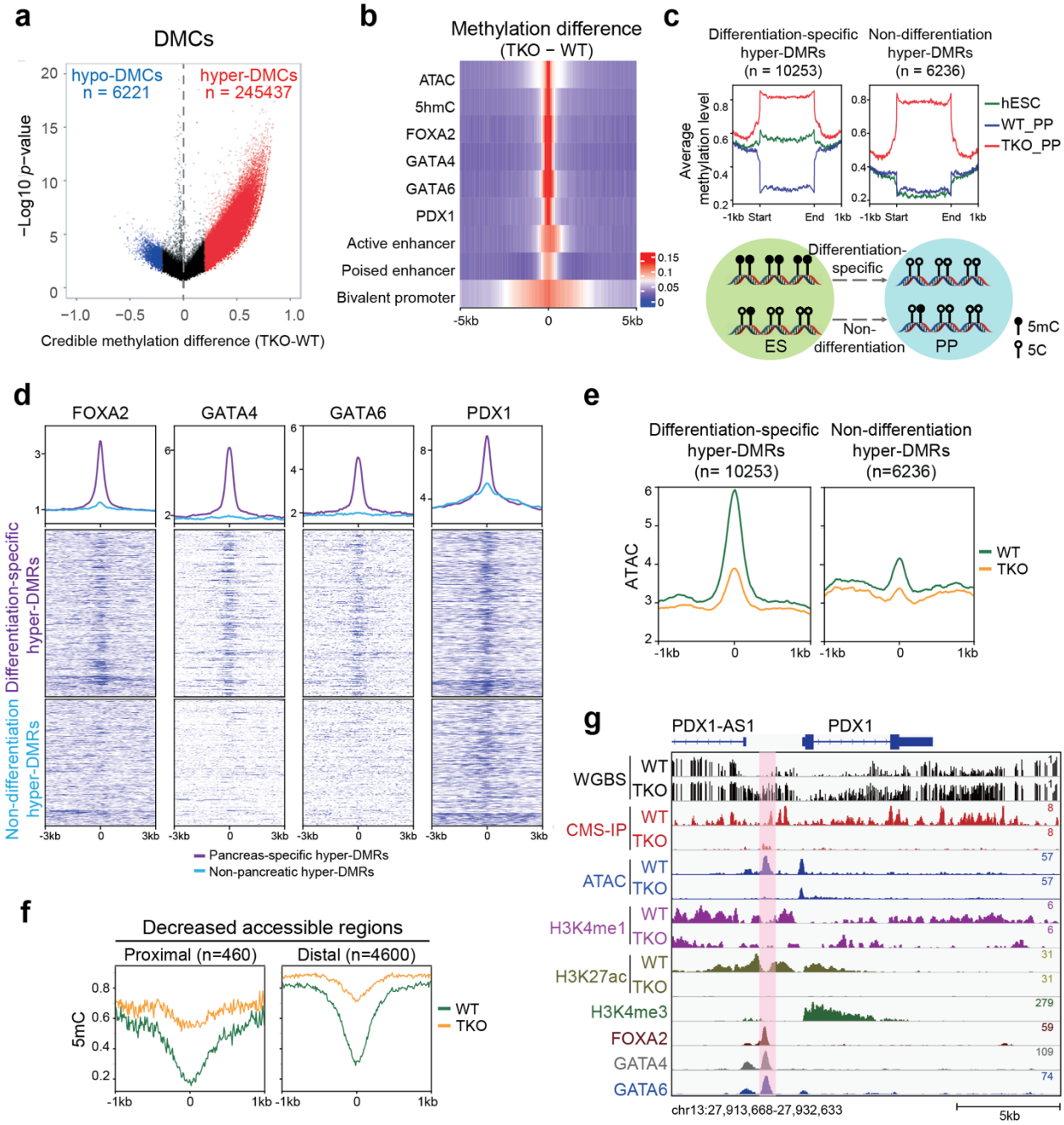
a
DMCs
hypo-DMCs
n = 6221
hyper-DMCs
n = 245437
−Log10 p−value
Credible methylation difference (TKO-WT)
b
Methylation difference
(TKO − WT)
ATAC
5hmC
FOXA2
GATA4
GATA6
PDX1
Active enhancer
Poised enhancer
Bivalent promoter
-5kb
0
5kb
c
Differentiation-specific hyper-DMRs (n = 10253)
Non-differentiation hyper-DMRs (n = 6236)
Average methylation level
hESC
WT_PP
TKO_PP
Differentiation-specific
Non-differentiation
ES
PP
5mC
5C
d
FOXA2
GATA4
GATA6
PDX1
Differentiation-specific hyper-DMRs
Non-differentiation hyper-DMRs
Pancreas-specific hyper-DMRs
Non-pancreatic hyper-DMRs
e
Differentiation-specific hyper-DMRs (n= 10253)
Non-differentiation hyper-DMRs (n=6236)
ATAC
WT
TKO
f
Decreased accessible regions
Proximal (n=460)
Distal (n=4600)
5mC
WT
TKO
g
PDX1-AS1
PDX1
WGBS
CMS-IP
ATAC
H3K4me1
H3K27ac
H3K4me3
FOXA2
GATA4
GATA6
chr13:27,913,668-27,932,633
5kb

Fig. 2

903 **Fig. 2 Differentiation-specific hyper-DMRs show reduced chromatin activity at pioneer TF**
904 **binding sites.**
905 **a** Volcano plot of WGBS data illustrating differentially methylated CpGs (DMCs) identified in
906 TKO_PP cells compared with WT_PP cells. Red and blue represent increased and decreased 5mC
907 in TKO_PP cells, respectively (credible methylation difference > 0.2). **b** Heatmap illustrating
908 methylation difference between TKO_PP and WT_PP cells at centers of annotated genomic
909 features (± 5 kb) for chromatin accessibility (ATAC), hydroxylation (5hmC), TF binding (FOXA2,
910 GATA4, GATA6, and PDX1), bivalent promoters, poised enhancers, and active enhancers.
911 Average 5mC signals of every 100-bp bin were calculated. **c** Classification of TKO hyper-DMRs
912 based on 5mC levels in hESCs (green), WT_PP cells (blue), and TKO_PP cells (red). **d** Average
913 density plots and heatmaps of FOXA2, GATA4, GATA6, and PDX1 signals at differentiation-
914 specific hyper-DMRs or non-differentiation hyper-DMRs in pancreatic progenitors. **e** Average
915 density plots of ATAC-seq at differentiation-specific hyper-DMRs or non-differentiation hyper-
916 DMRs in WT_PP (green) or TKO_PP (orange) cells. **f** Enrichment profile of methylation ratio
917 (5mC/C) at proximal (≤ 1 kb from TSS) and distal (> 1 kb from TSS) decreased accessible regions
918 in WT_PP (green) and TKO_PP (orange) cells. **g** Genome-browser view of the *PDX1*/*PDX1-AS1*
919 locus. A specific TKO hyper-DMR showing decreased 5hmC, ATAC-seq, and H3K27ac signals
920 is highlighted in pink.

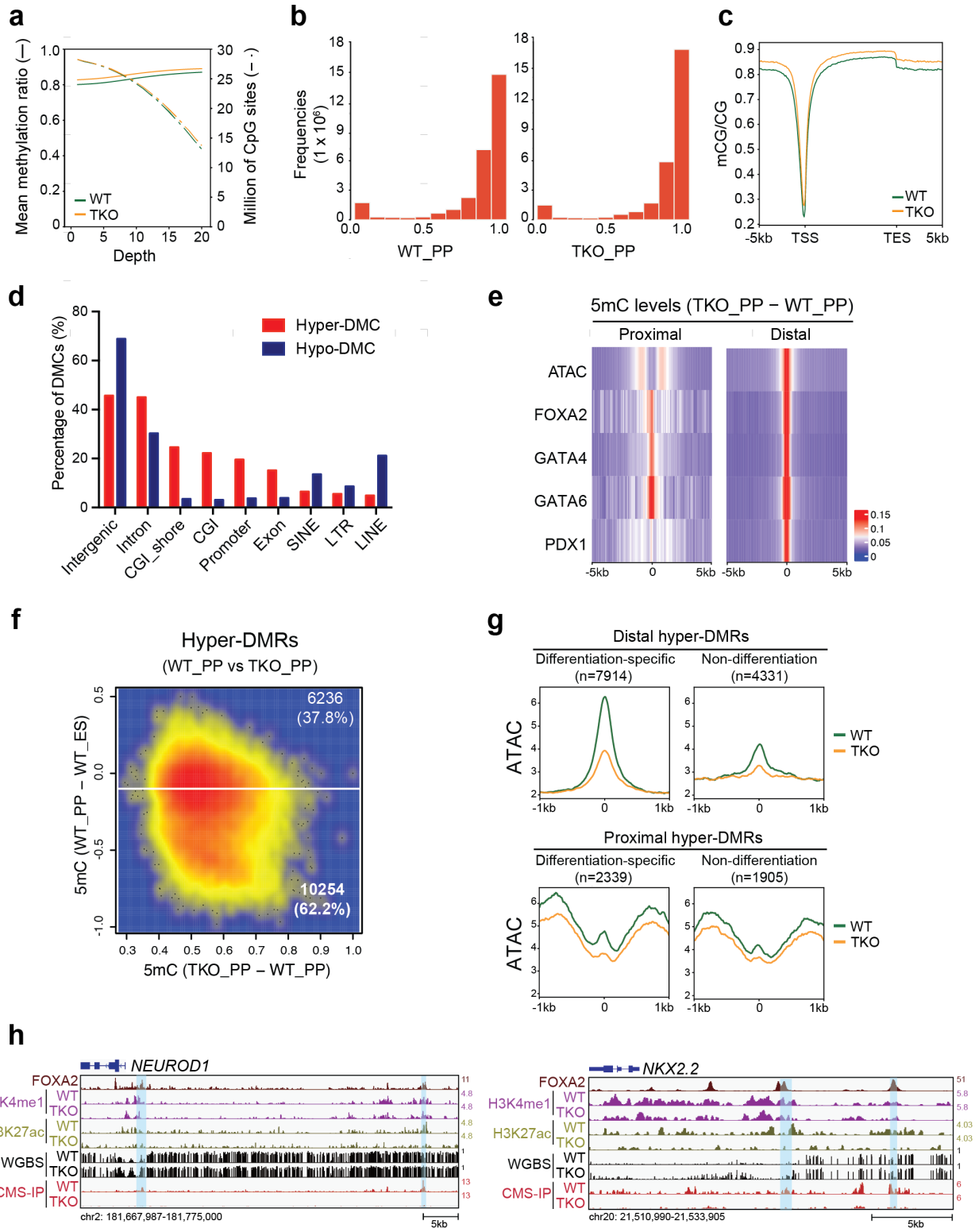
a
Mean methylation ratio (—)
Million of CpG sites (- - -)
Depth
WT
TKO
b
Frequencies (1 x 10^6)
WT_PP
TKO_PP
c
mCG/CG
-5kb
TSS
TES
5kb
WT
TKO
d
Percentage of DMCs (%)
Hyper-DMC
Hypo-DMC
Intergenic
Intron
CGI_shore
CGI
Promoter
Exon
SINE
LTR
LINE
e
5mC levels (TKO_PP − WT_PP)
Proximal
Distal
ATAC
FOXA2
GATA4
GATA6
PDX1
f
Hyper-DMRs
(WT_PP vs TKO_PP)
6236
(37.8%)
10254
(62.2%)
5mC (WT_PP − WT_ES)
5mC (TKO_PP − WT_PP)
g
Distal hyper-DMRs
Differentiation-specific
(n=7914)
Non-differentiation
(n=4331)
Proximal hyper-DMRs
Differentiation-specific
(n=2339)
Non-differentiation
(n=1905)
ATAC
WT
TKO
h
NEUROD1
FOXA2
H3K4me1
H3K27ac
WGBS
CMS-IP
chr2: 181,667,987-181,775,000
5kb
NKX2.2
chr20: 21,510,990-21,533,905

Fig. S2

922 **Supplementary Fig. 2 Loss of TET leads to extensive hypermethylation of pancreatic**
923 **regulatory elements**.
924 **a** CpG coverage in WT_PP and TKO_PP cells. The x-axis represents the sequencing coverage for
925 CpG sites. The solid line represents the average methylation ratio (5mC/C) at the corresponding
926 coverage for each sample (left y-axis). The dashed line represents the number of CpGs over the
927 corresponding coverage for each sample (right y-axis). **b** Bar plots showing the distribution of
928 methylation in WT_PP and TKO_PP cells. **c** Mean distribution of the average methylation ratio
929 (5mC/C) across gene bodies in WT_PP and TKO_PP cells. **d** Distribution of hyper-DMCs (red)
930 and hypo-DMCs (blue) at various genomic features. **e** Heatmap illustrating methylation difference
931 between TKO_PP and WT_PP cells at centers of annotated genomic features (± 5 kb) for
932 chromatin accessibility (ATAC) and TF binding (FOXA2, GATA4, GATA6, and PDX1) at
933 proximal (≤ 1 kb from TSS) and distal (> 1 kb from TSS) regions. **f** Comparison of global
934 methylation at hyper-DMRs revealed that TKO_PP cells show differentiation-specific hyper-
935 DMRs (n = 10,253) with decreased methylation in pancreatic progenitors (WT_PP) compared with
936 hESCs (WT_ES) (methylation change < -0.1). **g** Average density plots of ATAC-seq signals at
937 distal and proximal differentiation-specific or non-differentiation hyper-DMRs in WT_PP (green)
938 and TKO_PP (orange) cells. **h** Genome-browser views of *NEUROD1* and *NKX2.2* loci.
939 Representative hyper-DMRs showing decreased 5hmC and H3K27ac signals are highlighted in
940 blue.

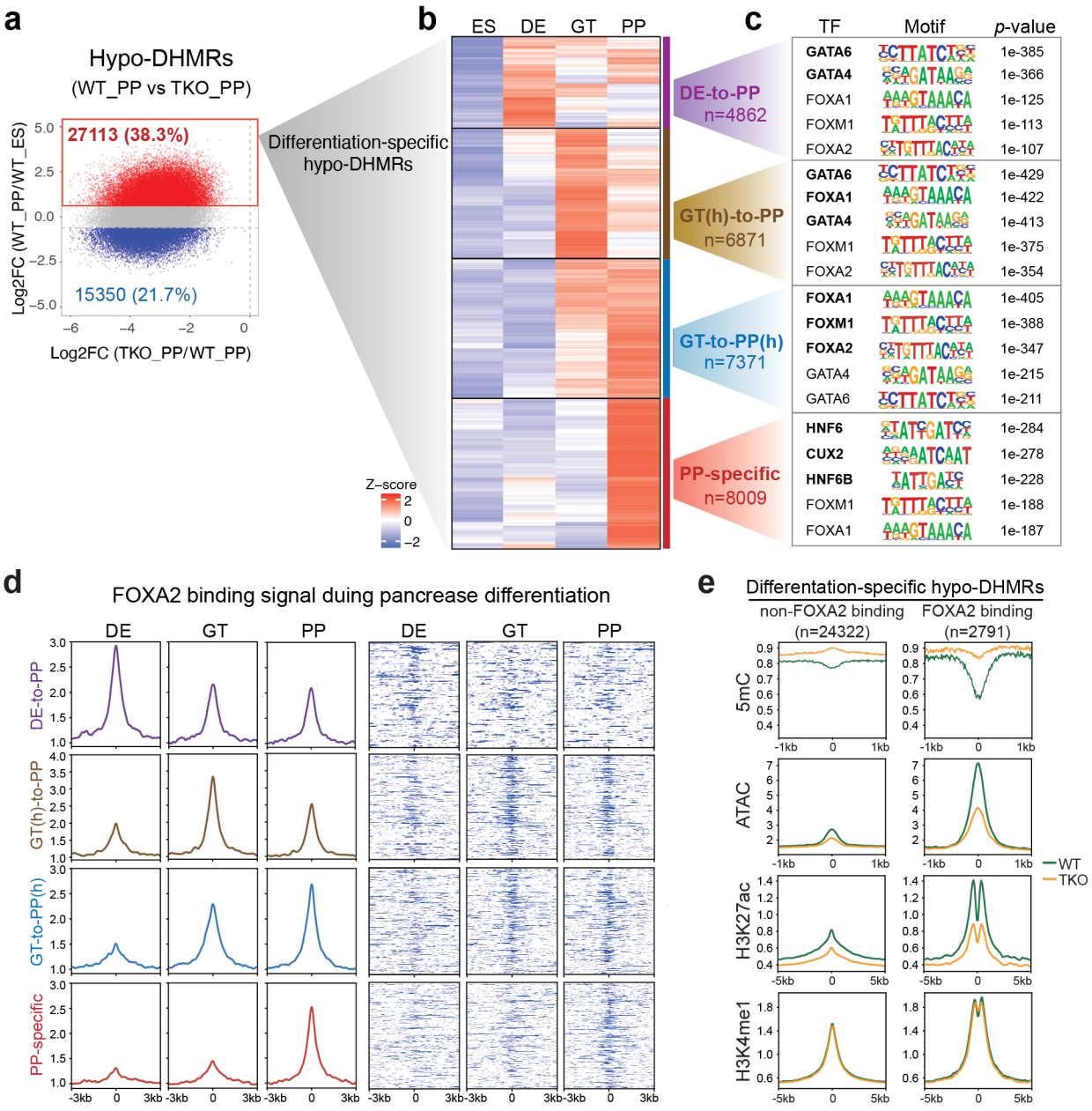
a
Hypo-DHMRs
(WT_PP vs TKO_PP)
27113 (38.3%)
15350 (21.7%)
Log2FC (WT_PP/WT_ES)
Log2FC (TKO_PP/WT_PP)
Differentiation-specific hypo-DHMRs
b
ES DE GT PP
Z–score
c
TF Motif p-value
DE-to-PP n=4862
GATA6 1e-385
GATA4 1e-366
FOXA1 1e-125
FOXM1 1e-113
FOXA2 1e-107
GT(h)-to-PP n=6871
GATA6 1e-429
FOXA1 1e-422
GATA4 1e-413
FOXM1 1e-375
FOXA2 1e-354
GT-to-PP(h) n=7371
FOXA1 1e-405
FOXM1 1e-388
FOXA2 1e-347
GATA4 1e-215
GATA6 1e-211
PP-specific n=8009
HNF6 1e-284
CUX2 1e-278
HNF6B 1e-228
FOXM1 1e-188
FOXA1 1e-187
d
FOXA2 binding signal duing pancrease differentiation
DE GT PP DE GT PP
DE-to-PP
GT(h)-to-PP
GT-to-PP(h)
PP-specific
e
Differentation-specific hypo-DHMRs
non-FOXA2 binding (n=24322)
FOXA2 binding (n=2791)
5mC
ATAC
H3K27ac
H3K4me1
WT
TKO

Fig. 3

941

942 **Fig. 3 Loss of TET disrupts epigenetic dynamics at differentiation-specific hypo-DHMRs**
943 **associated with distinct pioneer TFs motifs.**
944 **a** Scatterplot showing $\log_2$-fold change of 5hmC in pancreatic progenitors (WT_PP) versus hESCs
945 (WT_ES) within hypo-DHMRs identified in TKO_PP cells. Red indicates increased 5hmC and
946 blue shows decreased 5hmC in pancreatic progenitors relative to hESCs (|fold change| ≥ 1.5, FDR
947 < 0.05). **b** Clustering of 5hmC signals in WT cells at ES, DE, GT, and PP stage. Each row
948 represents one hypo-DHMR showing increased 5hmC in pancreatic progenitors compared with
949 hESCs. Red represents high 5hmC; blue indicates low 5hmC. **c** Motif enrichment analysis among
950 the four clusters defined in **b**. The top five known motifs are shown. **d** Average density plots and
951 heatmaps showing dynamic changes of FOXA2 binding in DE, GT, and PP lineage intermediates
952 across the four clusters defined in **b**. The color scale from white to blue in heatmaps represents
953 normalized FOXA2 signal from low to high. **e** Enrichment profile of methylation ratio (5mC/C),
954 chromatin accessibility (ATAC), H3K27ac, and H3K4me1 at 'differentiation-specific hypo-
955 DHMRs' alone (left column) and 'differentiation-specific hypo-DHMRs' overlapped with
956 FOXA2-bound sites (right column). The WT signal is marked in green and the TKO signal in
957 orange.

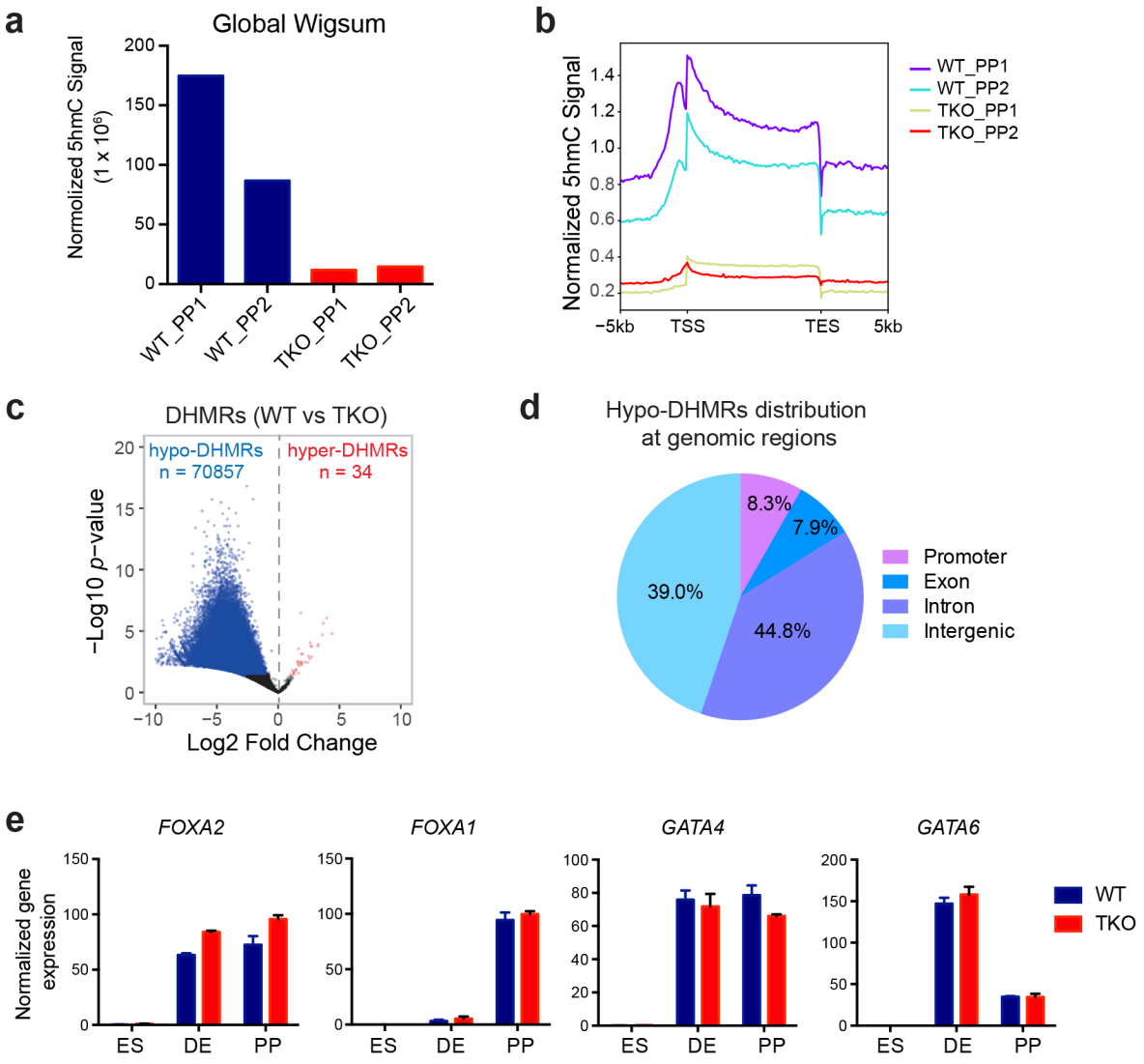
a
Global Wigsum
Normolized 5hmC Signal (1 x 10^6)
200
150
100
50
0
WT_PP1
WT_PP2
TKO_PP1
TKO_PP2
b
Normalized 5hmC Signal
1.4
1.2
1.0
0.8
0.6
0.4
0.2
−5kb
TSS
TES
5kb
WT_PP1
WT_PP2
TKO_PP1
TKO_PP2
c
DHMRs (WT vs TKO)
hypo-DHMRs
n = 70857
hyper-DHMRs
n = 34
−Log10 p−value
20
15
10
5
0
−10
−5
0
5
10
Log2 Fold Change
d
Hypo-DHMRs distribution
at genomic regions
8.3%
7.9%
39.0%
44.8%
Promoter
Exon
Intron
Intergenic
e
FOXA2
FOXA1
GATA4
GATA6
Normalized gene expression
ES
DE
PP
WT
TKO

Fig. S3

959 **Supplementary Fig. 3 Hypo-hydroxymethylation at cis-regulatory elements upon TET**
960 **inactivation.**

961 **a** Normalized 5hmC signal according to the spike-in analysis for WT_PP and TKO_PP samples
962 with two independent replicates. **b** Distribution of 5hmC signal across gene bodies, from the
963 transcription start site (TSS) to transcription end site (TES), in WT_PP and TKO_PP cells, with
964 two independent replicates (± 5 kb). **c** Volcano plot of CMS-IP-seq signals illustrating differential
965 hydroxymethylation regions (DHMRs) in TKO_PP cells relative to WT_PP cells (FDR < 0.05). **d**
966 Diagram illustrating the overall distribution of hypo-DHMRs in promoter (within ± 1 kb from
967 TSS), exon, intron, and intergenic regions. **e** Normalized gene expression of *FOXA1*, *FOXA2*,
968 *GATA4*, and *GATA6* in WT and TKO cells at the ES, DE, and PP stages by RNA-seq.

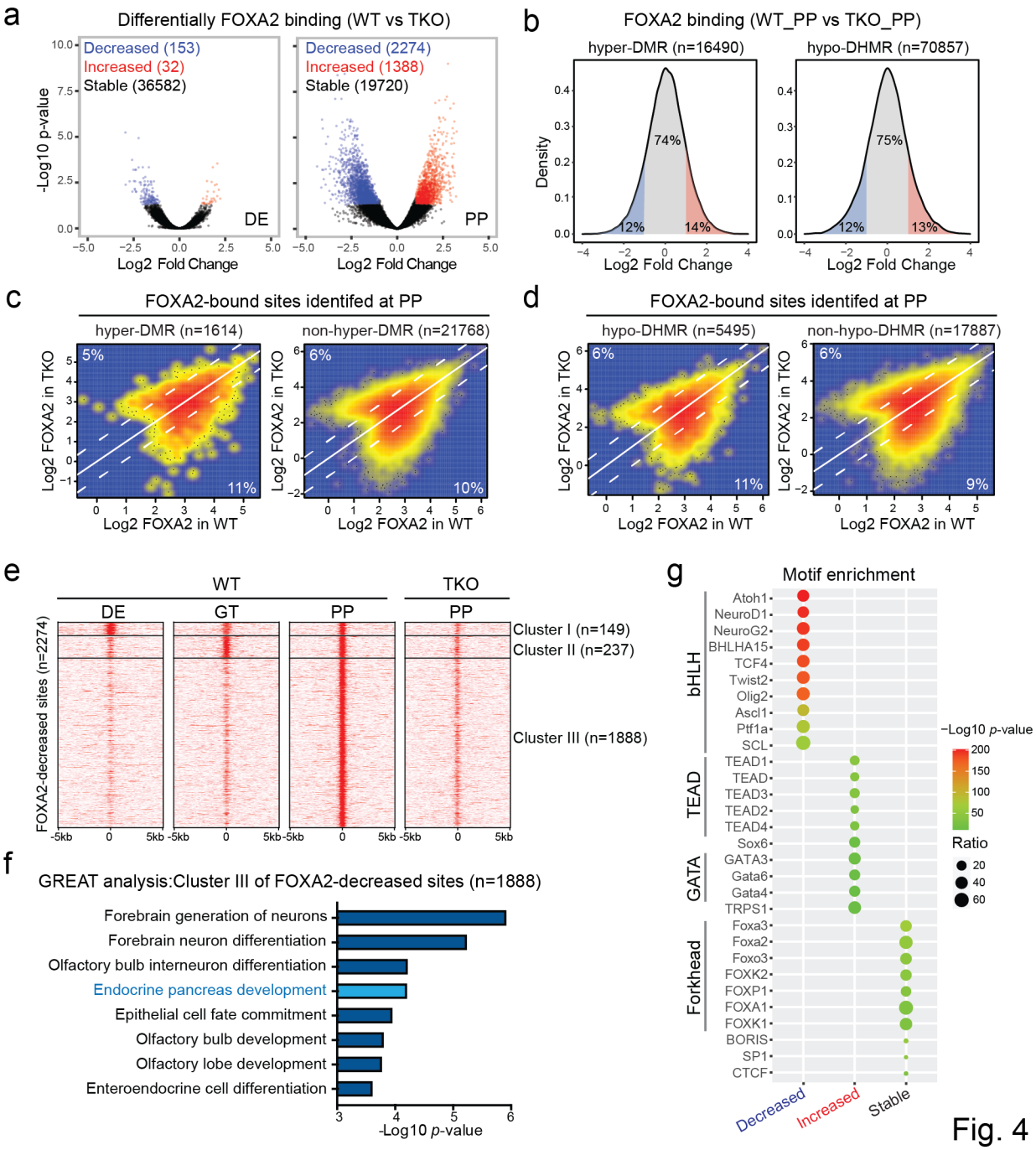
a
Differentially FOXA2 binding (WT vs TKO)
Decreased (153)
Increased (32)
Stable (36582)
Decreased (2274)
Increased (1388)
Stable (19720)
-Log10 p-value
DE
PP
Log2 Fold Change
b
FOXA2 binding (WT_PP vs TKO_PP)
hyper-DMR (n=16490)
hypo-DHMR (n=70857)
Density
74%
12%
14%
75%
12%
13%
Log2 Fold Change
c
FOXA2-bound sites identifed at PP
hyper-DMR (n=1614)
non-hyper-DMR (n=21768)
Log2 FOXA2 in TKO
Log2 FOXA2 in WT
5%
11%
6%
10%
d
FOXA2-bound sites identifed at PP
hypo-DHMR (n=5495)
non-hypo-DHMR (n=17887)
6%
11%
6%
9%
e
WT
TKO
DE
GT
PP
PP
FOXA2-decreased sites (n=2274)
Cluster I (n=149)
Cluster II (n=237)
Cluster III (n=1888)
-5kb
0
5kb
f
GREAT analysis:Cluster III of FOXA2-decreased sites (n=1888)
Forebrain generation of neurons
Forebrain neuron differentiation
Olfactory bulb interneuron differentiation
Endocrine pancreas development
Epithelial cell fate commitment
Olfactory bulb development
Olfactory lobe development
Enteroendocrine cell differentiation
-Log10 p-value
g
Motif enrichment
bHLH
Atoh1
NeuroD1
NeuroG2
BHLHA15
TCF4
Twist2
Olig2
Ascl1
Ptf1a
SCL
TEAD
TEAD1
TEAD
TEAD3
TEAD2
TEAD4
GATA
Sox6
GATA3
Gata6
Gata4
TRPS1
Forkhead
Foxa3
Foxa2
Foxo3
FOXK2
FOXP1
FOXA1
FOXK1
BORIS
SP1
CTCF
Decreased
Increased
Stable
−Log10 p-value
Ratio

Fig. 4

969

970 **Fig. 4 FOXA2 transiently binds to differentiation stage-specific genomic loci enriched for**
971 **bHLH motifs.**
972 **a** Volcano plots of FOXA2 ChIP-seq data illustrating differential FOXA2 binding sties identified
973 in TKO cells compared to WT cells at the DE (left) and PP stage (right). Red and blue represent
974 increased and decreased FOXA2 signal in TKO cells, respectively (|fold change| ≥ 2; FDR < 0.05).
975 **b** Density plots showing the percentages of differential FOXA2 binding in TKO_PP cells
976 compared to WT cells at hyper-DMRs (left) and hypo-DHMRs (right). The color of red, blue, and
977 gray represents increased, decreased, and non-changed FOXA2 signals in TKO_PP cells,
978 respectively (|fold change| ≥ 2). **c** Scatterplots presenting FOXA2 signals in WT_PP cells versus
979 TKO_PP cells at FOXA2 binding sites overlapped with (left) or without (right) hyper-DMRs.
980 Percentages of differential FOXA2 binding are indicated (|fold change| ≥ 2). **d** Scatterplots
981 presenting FOXA2 signals in WT_PP cells versus TKO_PP cells at FOXA2 binding sites
982 overlapped with (left) or without (right) hypo-DHMRs. Percentages of differential FOXA2
983 binding are indicated (|fold change| ≥ 2). **e** Clustering of FOXA2 binding signals at the DE, GT,
984 and PP stage in regions showed lost FOXA2 in TKO_PP cells. **f** Pathway enrichment annotations
985 from GREAT for cluster III decreased FOXA2-bound sites. Benjamini-Hochberg corrected *p*-
986 values were used. **g** Transcription factor motif enrichment analysis of genomic regions showed
987 decreased, increased, and stable FOXA2 binding in TKO_PP cells. The top 10 known motifs are
988 shown. The color representing *p*-values and the size of circle representing proportion of peaks
989 containing a TF motif for each group.

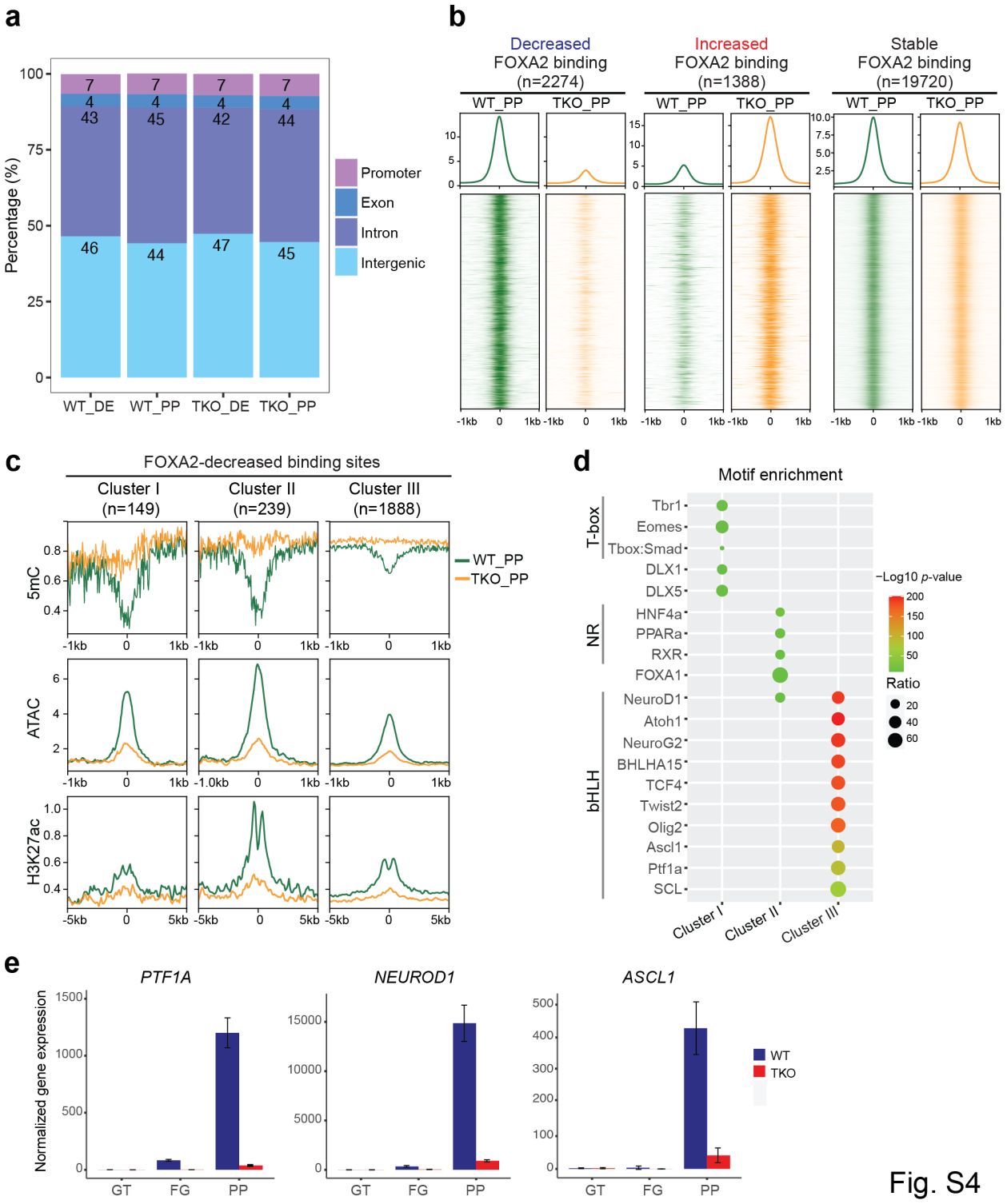
a
Percentage (%)
Promoter
Exon
Intron
Intergenic
WT_DE
WT_PP
TKO_DE
TKO_PP
b
Decreased FOXA2 binding (n=2274)
Increased FOXA2 binding (n=1388)
Stable FOXA2 binding (n=19720)
WT_PP
TKO_PP
c
FOXA2-decreased binding sites
Cluster I (n=149)
Cluster II (n=239)
Cluster III (n=1888)
5mC
ATAC
H3K27ac
WT_PP
TKO_PP
d
Motif enrichment
T-box
NR
bHLH
Tbr1
Eomes
Tbox:Smad
DLX1
DLX5
HNF4a
PPARa
RXR
FOXA1
NeuroD1
Atoh1
NeuroG2
BHLHA15
TCF4
Twist2
Olig2
Ascl1
Ptf1a
SCL
Cluster I
Cluster II
Cluster III
−Log10 p-value
Ratio
e
PTF1A
NEUROD1
ASCL1
Normalized gene expression
GT
FG
PP
WT
TKO

Fig. S4

990

991 **Supplementary Fig. 4 De novo FOXA2 binding at pancreas-specific loci features low active**
992 **chromatin modifications.**
993 **a** Bar graph illustrating the percentage of FOXA2 binding sites associated with different genomic
994 features in WT_DE, WT_PP, TKO_DE, and TKO_PP cells. **b** Average density plots and heatmaps
995 of FOXA2 signals in WT_PP (green) and TKO_PP (orange) cells across decreased, increased, and
996 stable FOXA2-bound sites. **c** Average density plot of methylation ratio (5mC/C), chromatin
997 accessibility (ATAC), and H3K27ac at cluster III decreased FOXA2-bound regions in WT_PP
998 (green) and TKO_PP (orange) cells. **d** Transcription factor motif enrichment analysis of genomic
999 regions identified in cluster I, cluster II, and cluster III of decreased FOXA2 binding in TKO_PP
1000 cells. Only the top scoring motifs are shown. The color representing *p*-values and the size of circle
1001 representing proportion of peaks containing a TF motif in each group. **e** Normalized gene
1002 expression of *PTF1A*, *NEUROD1*, and *ASCL1* in WT and TKO cells at the GT, FG, and PP stages
1003 by RNA-seq.

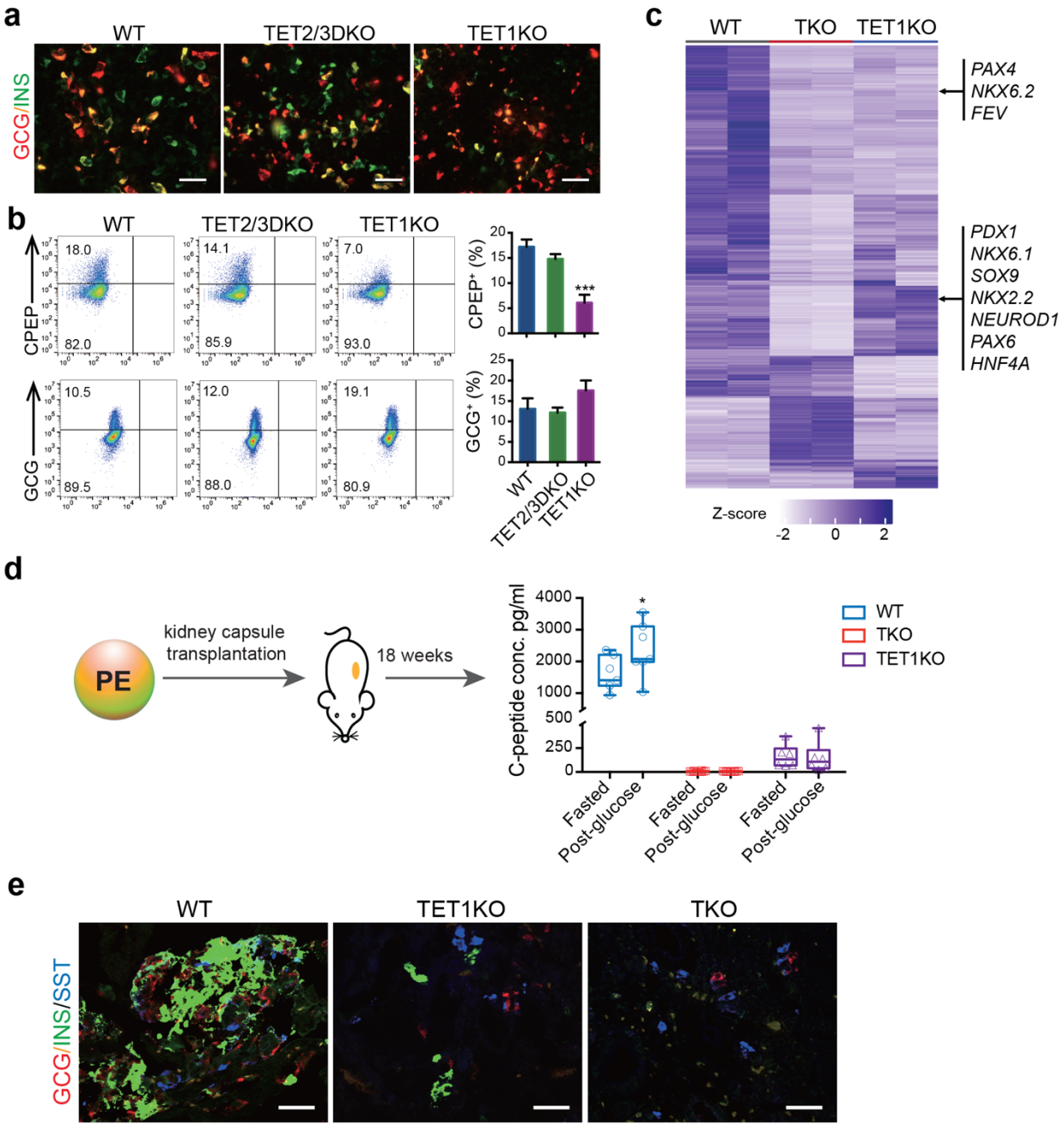
a
WT
TET2/3DKO
TET1KO
GCG/INS
b
WT
TET2/3DKO
TET1KO
CPEP
GCG
18.0
82.0
14.1
85.9
7.0
93.0
10.5
89.5
12.0
88.0
19.1
80.9
CPEP+ (%)
GCG+ (%)
***
c
WT
TKO
TET1KO
PAX4
NKX6.2
FEV
PDX1
NKX6.1
SOX9
NKX2.2
NEUROD1
PAX6
HNF4A
Z-score
-2
0
2
d
PE
kidney capsule transplantation
18 weeks
C-peptide conc. pg/ml
WT
TKO
TET1KO
Fasted
Post-glucose
e
WT
TET1KO
TKO
GCG/INS/SST

Fig. 5

1004

1005 **Fig. 5 TET1 knock-out cells show impaired differentiation into functional β-cells.**

1006 **a** Immunostaining of INS, and GCG in WT, TET2/3DKO, and TET1KO cells at the PE stage.
1007 Scale bar = 50 μm. **b** Representative plots of flow cytometry of human C-peptide (CPEP) and
1008 glucagon (GCG) in WT, TET2/3DKO, and TET1KO cells at the PE stage. Quantifications of the
1009 percentage of $CPEP^+$ or GCG^+ cells are shown in the right panel. n = 3 independent differentiation.
1010 **c** Heatmap showing the hierarchical clustering of DEGs among WT, TKO, and TET1KO cells at
1011 the PP stage. Each row represents one DEG, and each column represents one biological replicate.
1012 The color scale from white to blue represents normalized gene expression levels from low to high
1013 ($|\text{fold change}| \geq 2$; FDR < 0.05). **d** WT, TKO, and TET1KO cells were differentiated to the PE
1014 stage and transplanted under the kidney capsule of nondiabetic female SCID-beige mice. Eighteen
1015 weeks post-implantation, human C-peptide levels were measured after an overnight fast and 30
1016 min following an i.p. glucose injection. C-peptide levels from individual mice are shown in box
1017 and whisker plots. **e** Immunostaining of insulin (INS), glucagon (GCG), and somatostatin (SST)
1018 in cell grafts from WT_PE, TET1KO_PE, and TKO_PE transplanted mice. Scale bar = 50 μm.

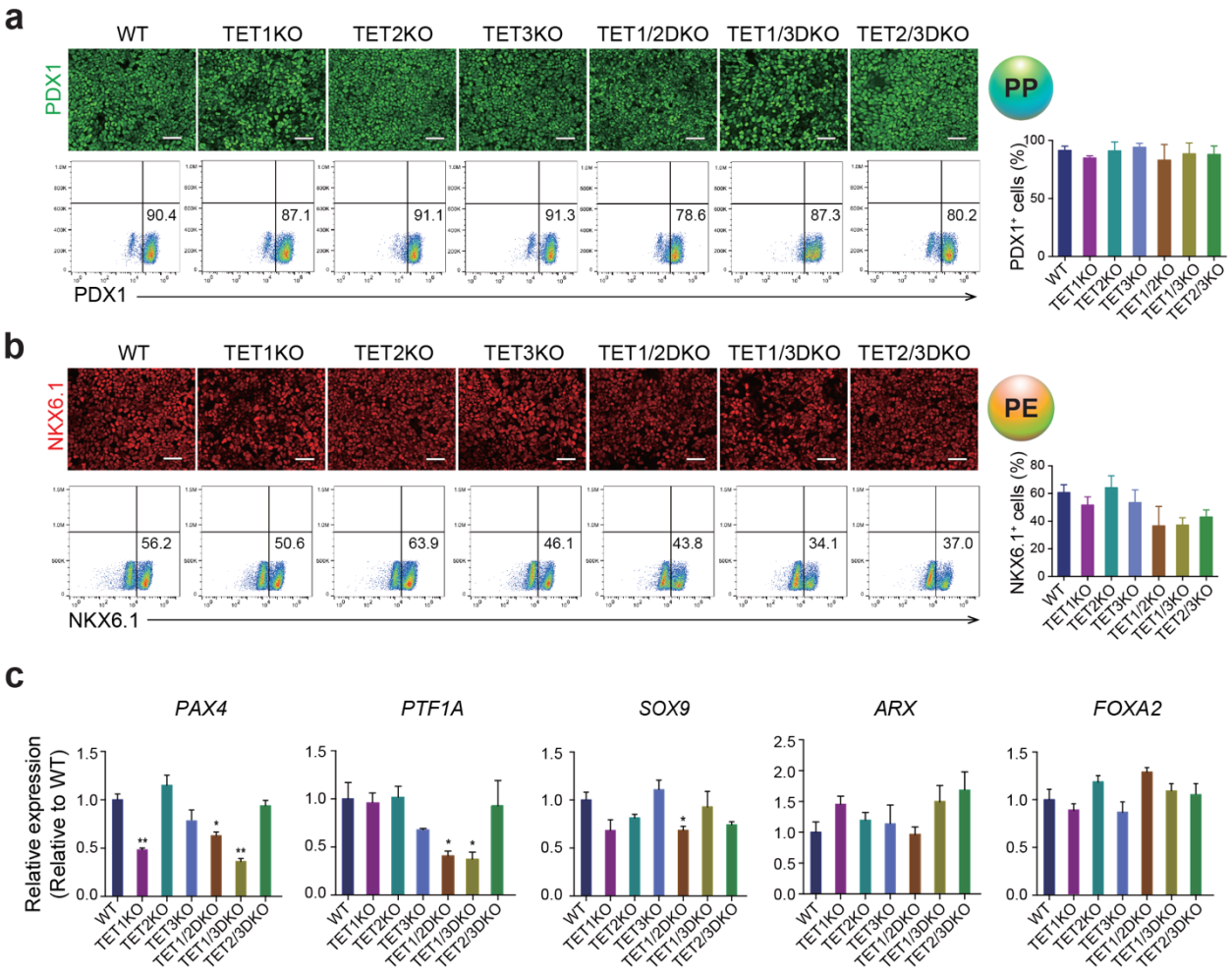
a
WT
TET1KO
TET2KO
TET3KO
TET1/2DKO
TET1/3DKO
TET2/3DKO
PDX1
PP
90.4
87.1
91.1
91.3
78.6
87.3
80.2
PDX1
PDX1+ cells (%)
b
NKX6.1
PE
56.2
50.6
63.9
46.1
43.8
34.1
37.0
NKX6.1
NKX6.1+ cells (%)
c
PAX4
PTF1A
SOX9
ARX
FOXA2
Relative expression (Relative to WT)

Fig. S5

1019

1020 **Supplementary Fig. 5 TET1 is required in pancreatic β-cell specification.**

1021 **a** Immunostaining and representative plots of flow cytometry of PDX1 at the PP stage for WT,
1022 TET1KO, TET2KO, TET3KO, TET1/2DKO, TET1/3DKO, and TET2/3DKO cells.
1023 Quantifications of the percentage of $PDX1^+$ cells are shown in the right panel. n = 3 independent
1024 differentiation. Scale bar = 50 μm. **b** Immunostaining and representative plots of flow cytometry
1025 of NKX6.1 at the PE stage for WT, TET1KO, TET2KO, TET3KO, TET1/2DKO, TET1/3DKO,
1026 and TET2/3DKO cells. Quantifications of the percentage of $NKX6.1^+$ cells are shown in the right
1027 panel. n = 3 independent differentiation. Scale bar = 50 μm. **c** Expression of *PAX4*, *PTF1A*, *SOX9*,
1028 *ARX*, and *FOXA2* in WT, TET1KO, TET2KO, TET3KO, TET1/2DKO, TET1/3DKO, and
1029 TET2/3DKO cells at the PP stage by RT-qPCR. RT-qPCR validation was performed with three
1030 independent batches of samples.

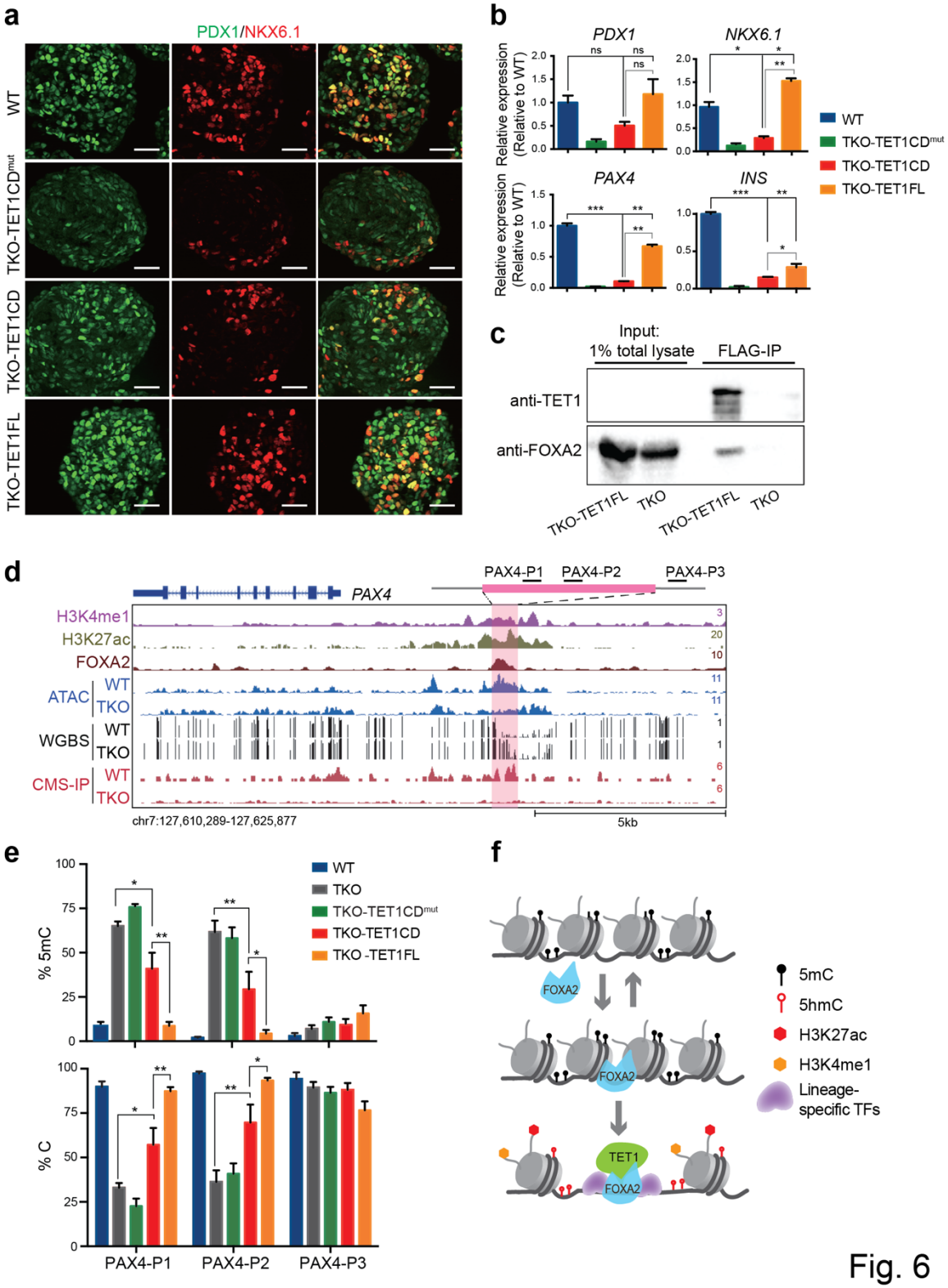
a
PDX1/NKX6.1
WT
TKO-TET1CDmut
TKO-TET1CD
TKO-TET1FL
b
PDX1
NKX6.1
PAX4
INS
Relative expression (Relative to WT)
WT
TKO-TET1CDmut
TKO-TET1CD
TKO-TET1FL
c
Input: 1% total lysate
FLAG-IP
anti-TET1
anti-FOXA2
TKO-TET1FL
TKO
d
PAX4-P1
PAX4-P2
PAX4-P3
PAX4
H3K4me1
H3K27ac
FOXA2
ATAC
WGBS
CMS-IP
chr7:127,610,289-127,625,877
5kb
e
% 5mC
% C
WT
TKO
TKO-TET1CDmut
TKO-TET1CD
TKO-TET1FL
PAX4-P1
PAX4-P2
PAX4-P3
f
FOXA2
TET1
5mC
5hmC
H3K27ac
H3K4me1
Lineage-specific TFs

Fig. 6

1031

1032 **Fig. 6 Full-length TET1 is required for the establishment of a hypomethylated PAX4**
1033 **enhancer**.

1034 **a** Immunostaining of PDX1 and NKX6.1 at the PP stage for WT, TKO-TET1CD^{mut}, TKO-
1035 TET1CD, and TKO-TET1FL cells. Scale bar=50 μm. **b** Expression of *PDX1*, *NKX6.1*, *PAX4*, and
1036 *INS* in WT, WT, TKO-TET1CD^{mut}, TKO-TET1CD, and TKO-TET1FL cells at the PE stage. **c**
1037 Western blots showing immunoprecipitation with antibodies against FLAG on cell extract of
1038 TKO-TET1FL cells followed by immunoblotting with antibodies against TET1 or FOXA2. TKO
1039 is included as control. **d** Genome-browser view of the *PAX4* locus with increased methylation and
1040 decreased chromatin associability upon TET depletion at a FOXA2-bound site featured enhancer
1041 signatures H3K4me1 and H3K27ac. **e** Locus-specific increase in 5mC at the *PAX4* enhancer in
1042 TKO or TKO-TET1CD samples compared with WT and TKO-TET1FL samples. Percentages of
1043 unmethylated cytosine and 5mC at CCGG sites are shown. n = 3 independent differentiation. **f**
1044 Schematic model depicting cooperative interaction among TET1 and TFs in lineage-specific
1045 enhancer activation. Pioneer TF FOXA2 transiently binds to enhancers. Lineage-specific TFs,
1046 such as PTF1A and NEUROD1, stabilize the binding of FOXA2 which recruits TET1 to induce
1047 DNA demethylation and subsequent chromatin opening.

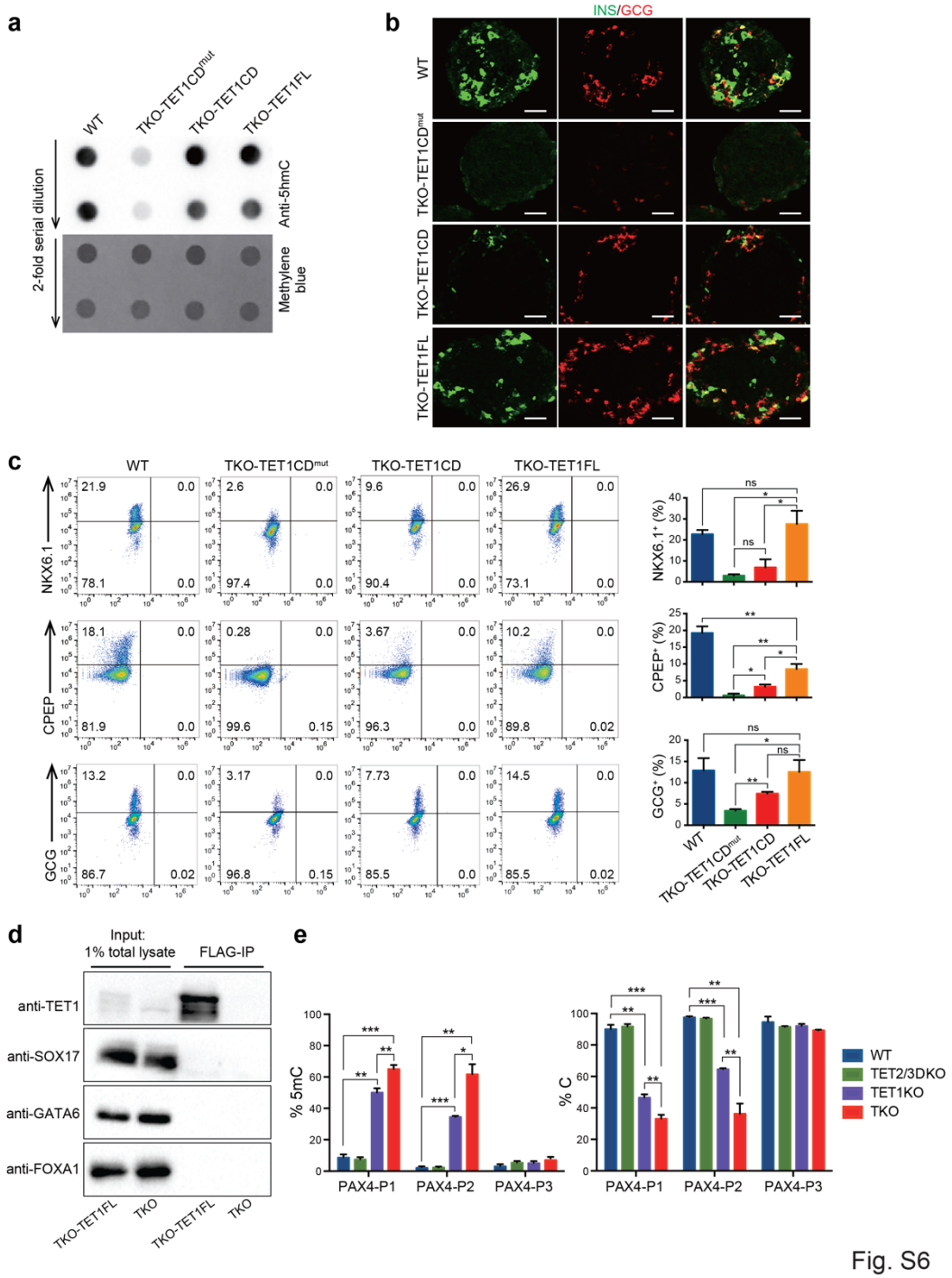
a
WT
TKO-TET1CDmut
TKO-TET1CD
TKO-TET1FL
2-fold serial dilution
Anti-5hmC
Methylene blue
b
INS/GCG
WT
TKO-TET1CDmut
TKO-TET1CD
TKO-TET1FL
c
WT
TKO-TET1CDmut
TKO-TET1CD
TKO-TET1FL
NKX6.1
CPEP
GCG
NKX6.1+ (%)
CPEP+ (%)
GCG+ (%)
d
Input:
1% total lysate
FLAG-IP
anti-TET1
anti-SOX17
anti-GATA6
anti-FOXA1
TKO-TET1FL
TKO
e
% 5mC
% C
PAX4-P1
PAX4-P2
PAX4-P3
WT
TET2/3DKO
TET1KO
TKO

Fig. S6

1049 **Supplementary Fig. 6 Reverse hypermethylation at *PAX4* enhancer through overexpression**
1050 **of full-length TET1.**

1051 **a** Analysis of global 5-hydroxymethylcytosine (5hmC) levels (top) in WT, TKO-TET1CD^{mut},
1052 TKO-TET1CD, and TKO-TET1FL cells by 5hmC dot blot analysis. The bottom panel shows
1053 methylene blue staining using the total amount of input DNA as the loading control. **b**
1054 Immunostaining of insulin (INS) and glucagon (GCG) at the PE stage for WT, TKO-TET1CD^{mut},
1055 TKO-TET1CD, and TKO-TET1FL cells. Scale bar=50 μm. **c** Representative plots of flow
1056 cytometry of NKX6.1, human C-peptide (CPEP), and glucagon (GCG) in WT, TKO-TET1CD^{mut},
1057 TKO-TET1CD, and TKO-TET1FL cells at the PE stage. Quantifications of the percentage of
1058 NKX6.1^{+}, CPEP^{+}, or GCG^{+} cells are shown in the right panel. n = 3 independent differentiation.
1059 **d** Western blots showing immunoprecipitation with antibodies against FLAG on cell extract of
1060 TKO-TET1FL cells followed by immunoblotting with antibodies against TET1, SOX17, GATA6,
1061 or FOXA1. TKO is included as control. **e** Locus-specific decreases in 5-methylcytosine (5mC) at
1062 the *PAX4* enhancer in TKO or TET1KO samples compared with TET2/3DKO samples.
1063 Percentages of unmethylated cytosine and 5mC at CCGG sites are shown. n = 3 independent
1064 differentiation.