

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

**Authors:** Xinwei Wu<sup>1#</sup>, Jianfang Li<sup>2#</sup>, Minjung Lee<sup>3</sup>, Qingping Lan<sup>1</sup>, Jia Li<sup>3</sup>, Yun Huang<sup>3</sup>, De-Qiang Sun<sup>2,3\*</sup>, Ruiyu Xie<sup>1\*</sup>

**Affiliations:**

<sup>1</sup> Centre of Reproduction, Development, and Aging, Faculty of Health Sciences, University of Macau, Taipa, Macau SAR.

<sup>2</sup> Department of Center Laboratory, The Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou 510700, China.

<sup>3</sup> Center for Epigenetics & Disease Prevention, Institute of Biosciences and Technology, College of Medicine, Texas A&M University, Houston, TX 77030, USA

# These authors contributed equally to the work.

\* Correspondence: ruiyuxie@um.edu.mo; dsun@ibt.tamhsc.edu

**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 displayed severe defects in pancreatic  $\beta$ -cell specification. Integrative whole-genome analysis identified unique cell-type-specific hypermethylated regions (hyper-DMRs) displaying reduced chromatin activity and remarkable enrichment for the binding of FOXA2, a pioneer transcription factor essential for pancreatic endoderm specification. Intriguingly, hundreds of hyper-DMRs recently identified in type-2 diabetes pathogenesis overlapped with the hyper-DMRs we found in TET-deficient cells. Furthermore, transduction of *TET1* in *TET*-deficient cells effectively rescued  $\beta$ -cell differentiation and reversed hypermethylation and suppression of the  $\beta$ -cell determinant *PAX4*. Genome-wide mapping of TET1 showed that TET1 co-localized at a subset of FOXA2 targets featuring high levels of active chromatin modifications in pancreatic progenitors. Taking these findings together with the defective generation of functional  $\beta$ -cells upon TET1-inactivation, our study not only unveils an essential role of TET1-dependent epigenetic regulation in establishing  $\beta$ -cell identity but also provides a new mechanistic clue regarding the complex crosstalk between TET dioxygenases and pioneer transcription factors in lineage specification.

## 32 **Introduction**

33 During embryonic development, pluripotent human embryonic stem cells (hESCs) differentiate into  
 34 many diverse lineages that make up the complex tissues and organs of the human body. Pancreatic  
 35 lineage specification relies on cross-talk between the genome and environmental cues in the  
 36 progenitor niche. This cross-talk is mediated by cis-regulatory elements that play a prominent role in  
 37 spatiotemporal gene regulation during embryogenesis. In particular, distal regulatory elements, such  
 38 as enhancers, serve as information integration hubs that allow binding of multiple regulators,  
 39 including lineage-specific transcription factors (TFs) as well as epigenetic readers, writers, and  
 40 erasers to ensure integration of intrinsic and extrinsic environmental cues at these loci <sup>1,2</sup>. It was  
 41 recently demonstrated that pioneer TFs, such as FOXA1 and FOXA2, are required for proper  
 42 chromatin opening and establishment of enhancer marks H3K4me1 and H3K27ac during pancreatic  
 43 fate specification <sup>3,4</sup>. These pioneer TFs bind to nucleosomal DNA to initiate chromatin remodeling  
 44 associated with DNA demethylation at newly accessible enhancers <sup>5-9</sup>.

45 DNA demethylation is mediated by ten-eleven-translocation methylcytosine dioxygenases  
 46 (TETs), which catalyze sequential oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine  
 47 (5hmC), 5-formylcytosine, and 5-carboxylcytosine <sup>10-14</sup>. Distribution of 5hmC, a novel epigenetic  
 48 modification, is dynamically changed by and positively correlated with active gene transcription  
 49 during early lineage specification <sup>15-17</sup>. Consequently, inhibition of TET family enzymes (TET1,  
 50 TET2, and TET3) impairs cell fate commitment into neural, hematopoietic, cardiac, and several other  
 51 lineages <sup>18-20</sup>. Although TETs have been shown to regulate 5mC homeostasis during early embryonic

development, the functional relevance and mechanisms by which TETs regulate pancreatic endocrine specification remain mostly unknown.

## Results

**TET deficiency impairs pancreatic endoderm differentiation.** We previously demonstrated that 5hmC positively correlates with ‘open’ chromatin at poised and active enhancers in multiple endodermal lineage intermediates using a stepwise hESC differentiation system toward pancreatic progenitors<sup>21</sup>. To determine the biological significance of TET-dependent regulation during pancreas specification, we generated *TET1*, *TET2*, and *TET3* single knock-out (KO), double knock-out (DKO), and triple knock-out (TKO) H1 hESC lines using CRISPR/Cas9 technology (Supplementary Fig. 1a). Mutations resulted in premature termination codons, which were confirmed by Sanger sequencing. Global DNA hydroxymethylation levels from positive clones were assayed by 5hmC dot blot. A significant reduction of 5hmC signals was shown in TKO, TET1KO, TET1/2DKO, and TET1/3DKO cells, whereas TET2KO, TET3KO, and TET2/3DKO cells showed minimal alterations of 5hmC levels (Supplementary Fig. 1b). To avoid functional redundancy, we focused our analysis on TKO cells devoid of any TET-mediated active demethylation. Consistent with other *TET* knock-out mESC and hESC lines<sup>18,22</sup>, TKO H1-hESCs exhibited no apparent defects in stem cell self-renewal capacity or expression of pluripotent factors (Supplementary Fig. 1e).

To determine whether TET proteins affect hESC differentiation toward pancreatic endocrine fate, we used an established stepwise differentiation platform<sup>21</sup> to induce efficient differentiation of hESCs to definitive endoderm (DE), primitive gut tube (GT), pancreatic progenitor (PP), and

pancreatic endocrine (PE) (Fig. 1a). Both TET triple-deficient lines (clones 2 and 6) displayed similar differentiation efficiency toward endoderm germ layer as the wild-type (WT) hESC line, in which over 90% of cells were SOX17<sup>+</sup> by day 3 of differentiation (Supplementary Fig. 1c). Moreover, expression levels of other endoderm markers, such as *FOXA2*, *FOXA1*, and *CXCR4*, were unchanged in TET-depleted cells compared with control cells (Supplementary Fig. 1d, e). This analysis suggests that TET dioxygenases are dispensable for endoderm specification in the context of *in vitro* hESC differentiation. To examine the effects of TET ablation on pancreatic commitment, we subsequently examined the expression of critical pancreatic markers at the PP and PE stages. Using flow cytometry to quantitate the expression of *PDX1* at the PP stage and co-expression of *PDX1* and *NKX6.1* at the PE stage, we observed a substantial decrease in TET knock-out cells (Fig. 1b). These data were confirmed by immunofluorescence staining and RT-qPCR analysis (Fig. 1c, Supplementary Fig. 1e). Consistent with these findings, expression of the pancreatic progenitor markers *SOX9* and *PTF1A* and endocrine hormones insulin, and glucagon were significantly downregulated in TKO lines (Fig. 1c, Supplementary Fig. 1e). Interestingly, *PAX4*, a key determinant for  $\beta$ -cell specification, failed to induce in TET-inactivated cells, whereas expression of the  $\alpha$ -cell determinant, *ARX*, was not affected (Supplementary Fig. 1e). These results demonstrate that TETs are required for pancreatic  $\beta$ -cell lineage specification.

To investigate TET-dependent global transcriptional changes during pancreatic endocrine cell fate commitment, we performed transcriptome analysis of WT and TKO cells at the ES, DE, and PP stages. Hierarchical clustering of genes specifically expressed at each stage revealed a remarkable downregulation of PP-specific genes in TKO\_PP cells, whereas ES-specific and DE-specific genes

showed no obvious changes in TET-deficient cells compared with WT cells (Supplementary Fig. 1f). In particular, 1,173 differentially expressed genes (DEGs) were identified in TKO\_PP cells (Fig. 1d, Supplementary Table 1). Among them, there were 921 downregulated DEGs in TET-deficient cells, including many important developmental determinants for pancreatic endocrine specification, such as *PDX1*, *NKX2.2*, *NKX6.1*, *NKX6.2*, *NEUROG3*, and *PAX4*. We also observed 252 genes, particularly HOX family members, which were significantly upregulated upon TET deletion, suggesting that failure to differentiate into pancreatic  $\beta$ -cells is tightly linked to an aberrant TF network. In addition, KEGG pathway analysis of the DEGs showed enrichment for terms associated with maturity onset diabetes of the young (MODY) and insulin secretion (Supplementary Fig. 1g), suggesting that loss of TETs impairs pancreatic endocrine formation at later stages of endoderm specification.

**FOXA motifs are enriched at distal regulatory elements displaying decreased accessibility upon TET deletion.** Given that our previous study demonstrated that TET-mediated cytosine oxidation is strongly associated with open chromatin regions during pancreatic differentiation <sup>21</sup>, we performed ATAC-seq to evaluate chromatin accessibility landscapes in WT and TET-deficient cells. We assessed differential accessible regions (DARs) between WT\_PP and TKO\_PP cells from a total of 124,322 identified ATAC peaks and found that 4.3% of ATAC peaks significantly changed upon TET deletion (Supplementary Table 2). We found that loss of TET led to an overall reduction in chromatin accessibility, with 5,060 DARs showing reduced accessibility and only 290 DARs showing increased accessibility in TKO\_PP cells (Supplementary Fig. 1h). Moreover, DARs with decreased accessibility were located primarily at distal regions (> 1 kb transcription start site (TSS)) (Fig. 1e).

Decomposition of chromatin accessibility and DEGs in TKO\_PP cells revealed that downregulated genes became less accessible (Fig. 1f). To enhance the biological insights of TET depletion-mediated changes in chromatin accessibility, we determined DNA binding motifs within DARs and found that regions with reduced accessibility were mostly enriched for motifs of pioneer TFs essential for pancreas organogenesis, such as FOXAs and GATAs<sup>23,24</sup> (Fig. 1g).

# **TET deficiency induces a differentiation-associated loss of 5hmC at pioneer TF binding sites.**

To gain insight into the TET-mediated DNA methylation network governing pancreatic differentiation, we used cytosine-5-methylenesulfonate immunoprecipitation (CMS-IP) coupled with high-throughput sequencing to profile DNA hydroxymethylation landscapes in WT\_PP and TKO\_PP cells. As a consequence of TET deletion, we found genome-wide loss of DNA hydroxymethylation in TET-deficient cells (Supplementary Fig. 2a, b). 94% of CMS-positive regions showed a significant reduction in TKO\_PP cells relative to WT\_PP cells (Supplementary Fig. 2c). These differentially hydroxymethylated regions showing reduced 5hmC (hypo-DHMRs) were primarily located at non-promoter regions (> 1 kb from TSS) (Supplementary Fig. 2d).

Reduced DNA hydroxymethylation in TET-deficient cells could be a result of initial loss of 5hmC at the ES stage or unsuccessful oxidation of 5mC during lineage progression. To provide biological insight into development-specific changes of hydroxymethylation upon TET inactivation, we systematically compared hypo-DHMRs found in TKO\_PP cells to previously identified regions showing increased 5hmC during differentiation of hESCs toward pancreatic endoderm<sup>21</sup>. Specifically, 27,113 hypo-DHMRs with higher 5hmC levels in WT\_PP cells than in WT\_ES cells (named ‘differentiation-specific hypo-DHMRs’) were first isolated from a total of 70,857 TKO hypo-

135 DHMRs (Fig. 2a). We subsequently clustered them into four categories based on the dynamic changes  
 136 of 5hmC across stages (Fig. 2b). In particular, one group gained 5hmC specifically at the PP stage  
 137 (PP-specific), whereas others gained 5hmC at the earlier DE (DE-to-PP) or GT (GT-to-PP) stages  
 138 and sustained hydroxymethylation until the PP stage. We found that GATA motifs were mostly  
 139 associated with the DE-to-PP cluster, in which gain of 5hmC began at the DE stage (Fig. 2c, purple).

140 By contrast, the PP-specific cluster displayed a prominent presence of binding sites for HNF6,  
 141 which is critical for pancreatic endocrine differentiation <sup>25</sup> (Fig. 2c, red). Most notably, FOXA motifs  
 142 were predominantly associated with the GT-to-PP(h) cluster in which 5hmC levels increased at the  
 143 GT stage and continued to be elevated until the PP stage (Fig. 2c, blue), whereas both FOXA and  
 144 GATA motifs were highly associated with the GT(h)-to-PP cluster in which 5hmC peaked at the GT  
 145 stage and subsequently decreased at the PP stage (Fig. 2c, brown). These results suggest a unique  
 146 temporal binding pattern of GATAs, FOXAs, and HNF6s associated with the dynamic distribution  
 147 of 5hmC during pancreatic differentiation, which is supported by a recent study demonstrating  
 148 sequential requirements for these TFs during transitions in pancreas development <sup>4</sup>.

149 It was previously demonstrated that the pioneer TF FOXA2 is essential for enhancer priming and  
 150 chromatin remodeling during pancreatic differentiation <sup>3,4</sup>. By analyzing FOXA2-binding datasets  
 151 previously generated <sup>4</sup> from a similar stepwise pancreatic differentiation system, we found that  
 152 binding of FOXA2 corresponded well with the dynamic changes of 5hmC across differentiation  
 153 stages in the four ‘differentiation-specific hypo-DHMRs’ clusters (Fig. 2d). These results suggest that  
 154 TET-mediated hydroxymethylation strongly correlates with FOXA2 deposition during development.  
 155 Given our previous finding that 5hmC is positively associated with chromatin accessibility and



enhancer activity<sup>21</sup>, we hypothesized that loss of TET inhibits chromatin opening and enhancer activation at FOXA2 targets during pancreatic differentiation. Thus, we conducted ChIP-seq for enhancer signatures H3K4me1 and H3K27ac in WT\_PP and TKO\_PP cells and performed integrated analysis. We found remarkable decreases in ATAC and H3K27ac signals on FOXA2 targets overlapping with ‘differentiation-specific hypo-DHMRs’ upon TET deletion (Fig. 2e), particularly at distal regions (Supplementary Fig. 2e). Moreover, ATAC, H3K4me1, and H3K27ac signals were substantially higher at FOXA2/hypo-DHMR-coexisting sites in pancreatic progenitors (Fig. 2e), suggesting that 5hmC-associated FOXA2 targets are predominantly enriched for active chromatin. Nevertheless, FOXA and GATA family members were expressed at similar levels in TKO and WT cells (Supplementary Fig. 2f), suggesting that inhibition of TET did not alter the expression levels of pioneer TFs. Collectively, these data reveal that TET deficiency results in a progressive failure of 5mC oxidation at a subset of FOXA2 targets essential for the establishment of active chromatin during differentiation.

**Pancreas-specific DMRs feature reduced chromatin activity during differentiation.** To investigate TET deficiency-mediated alterations in global DNA methylation, we performed whole-genome bisulfite sequencing (WGBS) in WT\_PP and TKO\_PP cells. Over 86% of sequencing reads were uniquely aligned to hg38 with a high sequencing depth of 24 × on CpG dinucleotides and typical bimodal distribution of methylation ratio in each sample (Supplementary Fig. 3a, b). The methylation ratio (5mC/C) was depleted at TSS but enriched across gene coding regions, especially within TKO\_PP cells (Supplementary Fig. 3c). Among 26.6 million CpG sites detected in both WT\_PP and TKO\_PP cells (depth ≥ 5), a total of 251,658 differentially methylated cytosines (DMCs; credible

difference of methylation ratio > 20%) were identified (Fig. 3a). Strikingly, 97.5% of DMCs were hypermethylated (hyper-DMCs), suggesting that TET inhibition results in a pronounced gain of methylation during pancreas differentiation. We observed enrichment of DMCs primarily at intergenic regions and introns (Supplementary Fig. 3d), whereas substantial enrichment of hypo-DMCs was also found in repeat elements, such as long interspersed elements, consistent with a suggested connection between hypomethylation and activation of transposable elements<sup>26,27</sup>. Based on an established link between DNA methylation and transcription regulation, we further calculated changes in methylation levels on cis-regulatory elements previously identified in pancreatic progenitors<sup>3,28</sup>. We found increased methylation at bivalent promoters, active enhancers, and, to a lesser extent, poised enhancers (Fig. 3b). Notably, hypermethylation was preferentially enriched at open chromatin (ATAC) as well as FOXA2, GATA4, GATA6, and PDX1 binding regions located distally (Fig. 3b, Supplementary Fig. 3e). This analysis demonstrates that active regulatory elements are hypermethylated upon TET depletion during pancreatic differentiation.

Our previous studies reveal that DNA demethylation is correlated with pancreatic endocrine patterning<sup>21</sup>. To gain better insight into the role of TET-mediated methylation changes in pancreatic differentiation, we analyzed differentially methylated regions (DMRs) by connecting at least three consecutive DMCs<sup>29</sup> between TKO\_PP and WT\_PP cells. We found a total of 16,490 hyper-DMRs and classified them into two categories based on their 5mC levels in pancreatic progenitors versus hESCs<sup>30</sup> (Fig. 3c). Notably, more than half (n = 10,254) of hyper-DMRs exhibited less 5mC in WT\_PP cells than in hESCs (named pancreas-specific hyper-DMRs), whereas the others (n = 6,236) showed similar methylation levels (non-pancreas hyper-DMRs) (Fig. 3c, Supplementary Fig. 3f).

198 Interestingly, we found that FOXA2, GATA4, and GATA6 were predominantly enriched at pancreas-  
 199 specific hyper-DMRs in pancreatic progenitors (Fig. 3d), suggesting that TET-dependent  
 200 demethylation is primarily associated with pioneer TF binding. We further compared ATAC,  
 201 H3K4me1, and H3K27ac signals at pancreas-specific hyper-DMRs with those at non-pancreatic  
 202 hyper-DMRs. We found that chromatin activity was markedly reduced in TET-deficient cells at  
 203 distally located hyper-DMRs in the pancreatic-specific group (Fig. 3e, Supplementary Fig. 3g).  
 204 Consistently, genomic loci displaying decreased accessibility showed a substantial increase in DNA  
 205 methylation upon depletion of TET (Fig. 3f). For example, we detected significant hypermethylation  
 206 at the distal regulatory elements of *PTF1A* and *NKX2.2* loci (Supplementary Fig. 3h). Altogether,  
 207 these analyses demonstrate that TET-dependent demethylation at enhancers and other distal  
 208 regulatory elements is essential for chromatin remodeling during pancreas development.

209 Additionally, we found that 774 type 2 diabetes (T2D)-associated islet hyper-DMRs, which were  
 210 recently identified <sup>31</sup>, overlapped with hyper-DMRs found in TET-deficient cells (Supplementary  
 211 Table 3). Remarkably, four of the most significant T2D-associated islet hyper-DMRs annotated to  
 212 the *PDX1* and lncRNA *PDX1-AS1* loci were hypermethylated upon TET deletion during pancreatic  
 213 differentiation (Fig. 3g, green areas). We also uncovered an adjacent hyper-DMR (Fig. 3g, pink area)  
 214 that overlapped with the pioneer TF binding sites and showed decreased ATAC, H3K4me1, and  
 215 H3K27ac signals in a TET-dependent manner. Given that both *PDX1* and *PDX1-AS1* are  
 216 downregulated in islets from donors with T2D <sup>32</sup>, these data suggest that TET inactivation leads to  
 217 remarkable DNA hypermethylation featuring loss of chromatin activity at regulatory elements, which  
 218 might be crucial for the development of diabetes.

219 **TET1 is required in pancreatic  $\beta$ -cell specification.** Our findings suggest that TET inactivation-  
 220 induced aberrant methylation/hydroxymethylation at least in part contributes to defective  $\beta$ -cell  
 221 specification. To determine the essential role of each TET family member in pancreatic differentiation,  
 222 we further analyzed mutants with single (*TET1*, *TET2*, or *TET3*) or double (*TET1/2*, *TET2/3*, or  
 223 *TET1/3*) *TET* knock-out. All mutant lines were able to induce the expression of *PDX1* and *NKX6.1*  
 224 to levels comparable to WT (Supplementary Fig. 4a). However, only those retaining intact *TET1*  
 225 expression (*TET2*KO, *TET3*KO, and *TET2/3*DKO) displayed proper induction of *PAX4* (Fig. 4a). In  
 226 comparison with *TET2/TET3* double-deletion, inhibition of TET1 alone had significant effects on the  
 227 formation of INS- and C-peptide (CPEP)-expressing  $\beta$ -cells but not GCG-expressing  $\alpha$ -cells (Fig. 4b,  
 228 Supplementary Fig. 4b). To further determine whether TET1 is responsible for  $\beta$ -cell induction, we  
 229 restored *TET1* expression using lentivirus-mediated gene transduction to TET-depleted hESCs.  
 230 *TET1*-repaired TKO cells showed a global increase in 5hmC levels, confirming the replenishment of  
 231 hydroxymethylation (Supplementary Fig. 4c). Differentiation of the TKO-TET1 line toward  
 232 pancreatic endocrine fate led to robust restoration of  $\beta$ -cell genes, including *NKX6.1*, *PAX4*, *INS*, and  
 233 *CPEP* (Fig. 4c, d, Supplementary Fig. 4d), suggesting that TET1 is critical for  $\beta$ -cell specification.

234 To evaluate global transcriptome changes in response to *TET1* deletion, we performed RNA-seq  
 235 of TET1KO cells at the PP stage. Integrated analysis of WT\_PP, TKO\_PP, and TET1KO\_PP data  
 236 revealed a total of 1,590 DEGs among the three lines (Supplementary Table 4). Specifically, 555  
 237 genes, including most pancreas developmental regulators such as *PDX1*, *SOX9*, *NKX2.2*, *NKX6.1*,  
 238 *NEUROD1*, and *PAX6*, were downregulated in TKO\_PP cells but not in TET1KO\_PP cells (Fig. 4e),  
 239 consistent with effective differentiation of TET1KO-hESCs into *PDX1*<sup>+</sup>/*NKX6.1*<sup>+</sup> cells

(Supplementary Fig. 4a). However, 536 genes, including the  $\beta$ -cell fate determinants *PAX4*, *NKX6.2*, and *FEV*<sup>33-35</sup>, were significantly inhibited in TET1-deficient cells, implicating TET1 as responsible for the activation of a subset of genes essential for  $\beta$ -cell identity. To further analyze the functional consequences of TET1 inhibition, WT-, TET1KO-, and TKO-hESCs were differentiated to the PE stage and subsequently engrafted into SCID-beige mice under the kidney capsules (Fig. 4f). Glucose-stimulated human C-peptide secretion was determined 18 weeks post-transplantation. Notably, mice engrafted with WT\_PE cells produced substantial fasting C-peptide in serum ( $1605 \pm 527$  pg/ml) and showed statistically significant glucose-stimulated C-peptide secretion ( $2,358 \pm 839$  pg/ml). By contrast, mice transplanted with TET1KO\_PE cells secreted extremely low amounts of basal C-peptide ( $163 \pm 122$  pg/ml) and showed no response to glucose stimulation ( $149 \pm 161$  pg/ml). In agreement with these functional results, excised WT\_PE cell grafts were highly composed of insulin<sup>+</sup>  $\beta$ -cells (Supplementary Fig. 4e). The TET1KO\_PE cell grafts displayed much less insulin content, and only the  $\delta$ -cell hormone somatostatin was detected in TKO\_PE cell grafts. Together, these data demonstrate that loss of TET1 impairs  $\beta$ -cell specification and maturation.

**TET1 binding sites are distinct in pancreatic progenitors.** To identify the direct targets of TET1, we performed ChIP-seq for TET1 in pancreatic progenitors and identified a total of 8,476 peaks with higher signals around TSSs (Supplementary Fig. 5a). Consistent with previous reports<sup>18,36</sup>, TET1 was enriched at CpG islands (Supplementary Fig. 5b) and was primarily associated with unmethylated genomic loci (Fig. 5a). To further examine the epigenetic features of TET1 target sites in pancreatic progenitors, we centered all TET1 peaks and analyzed their accessibility, 5hmC, and deposition of H3K4me1, H3K27ac, H3K4me3<sup>28</sup>, and H3K27me3<sup>28</sup>. We found that TET1 binding

261 sites predominantly overlapped with active chromatin modifications at proximal (Fig. 5b, left panel)  
 262 and distal (Fig. 5b, right panel) regions, suggesting that TET1 is primarily present at active regulatory  
 263 elements in pancreatic progenitors.

264 Recent studies show that TET1 safeguards bivalent promoters in ESCs to ensure proper neuronal  
 265 differentiation<sup>18</sup>. Therefore, it is interesting to examine whether TET1 occupies similar or different  
 266 loci upon hESC differentiation toward pancreatic endoderm. We compared TET1 binding sites  
 267 identified in pancreatic progenitors and pluripotent stem cells using TET1 ChIP-seq data previously  
 268 generated from hESCs<sup>18</sup>. 4,463 TET1 peaks specific to pancreatic progenitors were referred to ‘PP-  
 269 specific’, 22,177 binding sites specific to hESCs were called ‘ES-specific’, and 4,013 TET1 sites  
 270 shared between the two groups were named ‘common’ (Fig. 5c). We then analyzed the annotations  
 271 of their nearby genes with Genomic Regions Enrichment of Annotations Tool (GREAT) and found  
 272 that only PP-specific peaks were enriched for terms of endocrine pancreas development  
 273 (Supplementary Fig. 5c). In agreement with these findings, PP-specific sites prominently featured  
 274 FOXA2 and CTCF motifs, whereas insulator CTCF and its paralogue BORIS motifs were mostly  
 275 enriched at ES-specific sites (Fig. 5d). This analysis suggests that TET1 is involved in different  
 276 biological functions in ESCs and lineage-committed cells. Interestingly, in contrast to the PP-specific  
 277 sites, most of which were distally located (Fig. 5e), the ‘common’ TET1 binding sites were mainly  
 278 located at proximal promoters and enriched for gene ontology annotations connected to essential  
 279 biological processes, such as mRNA processing, biosynthesis, and translation (Supplementary Fig.  
 280 5c). We further examined chromatin accessibility and histone modifications of promoter and  
 281 enhancer regions at proximal and distal TET1 binding sites. We found that TET1 bound more strongly

282 at both proximal and distal ‘common’ sites than at ‘PP-specific’ sites (Fig. 5f, TET1). Higher active  
 283 chromatin signals were also observed at the proximal (Fig. 5f, ATAC and H3K4me3) and distal (Fig.  
 284 5f, ATAC and H3K27ac) ‘common’ sites. The more active chromatin modifications at ‘common’  
 285 TET1 targets may contribute to maintaining continually active transcripts essential for basic  
 286 biological functions in different types of cells. By contrast, cell-type-specific TET1 binding may  
 287 participate in chromatin remodeling during lineage specification.

288 Recent studies demonstrate that FOXA2 triggers chromatin opening on a subset of its potential  
 289 binding sites during endoderm differentiation <sup>37</sup>. In the present study, we found that FOXA2 target  
 290 sites acquiring 5hmC during pancreatic differentiation correlate with high levels of active chromatin  
 291 modifications (Fig. 2). We thus hypothesized that TET1 binds to a specific set of FOXA2 targets  
 292 associated with unique chromatin features. We isolated FOXA2 binding sites in pancreatic  
 293 progenitors <sup>4</sup> and analyzed DNA methylation, chromatin accessibility, and enhancer activity with or  
 294 without TET1 binding. The functional relevance of TET1 and FOXA2 co-binding is supported by  
 295 exceptionally low levels of methylation as well as more active chromatin modifications (ATAC,  
 296 H3K4me1, and H3K27ac), particularly at distal regions (Fig. 5g, Supplementary Fig. 5d). Moreover,  
 297 only the loci bound by both FOXA2 and TET1 were significantly enriched for terms associated with  
 298 pancreas development (Supplementary Fig. 5e), implicating a unique feature of TET1/FOXA2 co-  
 299 binding elements. Taken together, our data suggest that TET1 binds to a specific subset of FOXA2  
 300 targets that are confined mainly to active regulatory compartments during pancreatic differentiation.  
 301 **TET1 is required for the *PAX4* enhancer to achieve a hypomethylated state.** As TET1 depletion  
 302 inhibited the induction of genes essential for  $\beta$ -cell specification (Fig. 4), we reasoned that co-binding

303 of TET1 and FOXA2 at distal regulatory elements modulates focal DNA  
 304 methylation/hydroxymethylation and subsequent gene activation. To identify potential TET1  
 305 regulatory elements, we integrated gene expression, chromatin accessibility, and TET1/FOXA2  
 306 binding profiles into a comprehensive analysis. We found that a putative *PAX4* enhancer (~4.0 kb  
 307 upstream of the TSS), where TET1/FOXA2 co-bound, displayed hypermethylation and decreased  
 308 accessibility in TET-deficient cells (Fig. 6a, pink area). We next examined whether the deletion of  
 309 TET1 alone results in hypermethylation at this site. Locus-specific methylation was determined in  
 310 TET1KO\_PP, TET2/3DKO\_PP, TKO\_PP, and WT\_PP cells using a glucosylation and digestion-  
 311 based method followed by qPCR analysis with two independent pairs of primers (PAX4-P1 and  
 312 PAX4-P2) located within the TET1/FOXA2 co-binding site (Fig. 6a). Notably, the percentage of  
 313 5mC was increased to 35–50% for TET1KO and 60–65% for TKO compared with WT but remained  
 314 unchanged for TET2/3DKO (Fig. 6b, left panel). A corresponding decrease in unmethylated cytosine  
 315 content was observed, with values of 45–65% and ~35% of C in TET1KO and TKO samples,  
 316 respectively (Fig. 6b, right panel). As a housekeeping control, we examined methylation content at a  
 317 nearby region using the PAX4-P3 primer, and, as expected, no significant differences in 5mC and C  
 318 were observed in any samples.

319 Furthermore, we determined methylation contents at the same *PAX4* locus in TET1-repaired  
 320 TKO cells. Consistent with the results found in TET2/3DKO cells, no detectable differences in 5mC  
 321 and C were found among TKO-TET1 samples (Supplementary Fig. 6). Together with the increased  
 322 expression of *PAX4* in TKO-TET1 cells (Fig. 4d), we conclude that TET1 is responsible for the  
 323 establishment or maintenance of a hypomethylated state at the *PAX4* enhancer, which is essential for



transcription activation of *PAX4*.

## Discussion

In the present work, we dissected the roles of TET proteins in pancreatic endocrine commitment based on a stepwise hESC differentiation system. We found that the loss of all three TET family members significantly impaired the differentiation of pancreatic  $\beta$ -cells. Furthermore, we discovered that locus-specific hypermethylation was associated with genes essential for  $\beta$ -cell specification and maturation, such as *PAX4*<sup>33</sup>, *PDX1*<sup>38</sup>, and *NKX2.2*<sup>39</sup>. The reintroduction of *TET1* in TET-deficient cells effectively reversed hypermethylation and restored the expression of *PAX4*. We further demonstrated that TET1 functions as an upstream epigenetic regulator of *PAX4* through direct binding at a putative *PAX4* enhancer to preserve its unmethylated status, thereby potentiating *PAX4* expression to adopt  $\beta$ -cell fate during endocrine lineage commitment. Consistently, we observed striking increases in methylation at the *PAX4* enhancer in TET1 knock-out cells but not in TET2/TET3 DKO cells, suggesting that TET1 epigenetically regulates induction of the  $\beta$ -cell program in a locus-specific manner. Moreover, despite the successful induction of *PDX1* and *INS* upon deletion of TET1 alone, mice receiving TET1KO cell grafts showed a persistent defective insulin response to glucose, implying that TET1 is also essential for  $\beta$ -cell maturation.

Whereas the specification of  $\beta$ -cells was strongly influenced by depletion of *TET1*, we did not observe a significant impact on the expression of *ARX*, a critical  $\alpha$ -cell fate determinant<sup>40</sup>. In contrast to *PAX4*, no hyper-DMRs were identified within the *ARX* locus, where methylation levels were nearly undetectable in pancreatic progenitors. Previous studies demonstrate that several CpG-rich sites of *Arx*, including one site close to TSS and another site 2 kb upstream of TSS, are heavily methylated

345 in adult  $\beta$ -cells but not  $\alpha$ -cells<sup>41</sup>. Moreover, pharmacological inhibition of DNA methyltransferases,  
 346 *Dnmts*, in pancreatic progenitors promotes  $\alpha$ -cell specification<sup>42</sup>, whereas deletion of *Dnmt1* in  $\beta$ -  
 347 cells results in demethylation and depression of *Arx*<sup>41</sup>. We thus speculate that *ARX* is hypomethylated  
 348 in pancreatic progenitors in a TET-independent manner. The *ARX* locus maintains an unmethylated  
 349 state when progenitors differentiate into  $\alpha$ -cells, whereas it becomes hypermethylated once cells  
 350 commit to  $\beta$ -cell fate in the presence of DNMTs and other  $\beta$ -cell-specific factors. In line with this  
 351 hypothesis are findings that Nkx2.2 recruits Dnmt3a to the *Arx* promoter to repress its expression in  
 352  $\beta$ -cells<sup>43</sup>. In the future, it will be interesting to explore whether there are differences in the complete  
 353 epigenetic landscapes of endocrine progenitors, which subsequently differentiate into  $\alpha$ - or  $\beta$ -cells.

354 TET dioxygenases are critical for lineage induction in a cell-type-specific manner<sup>44</sup>. How TETs  
 355 recognize lineage-specific regulatory elements and modulate chromatin remodeling during pancreas  
 356 development remains unknown. Here, we addressed these questions by performing analyses  
 357 integrating gene expression with multiple chromatin features, including DNA methylation,  
 358 hydroxymethylation, chromatin accessibility, and histone modifications of promoters and enhancers.  
 359 We found extensive hypermethylation in TET-deficient cells that differentiated to the pancreatic  
 360 progenitor stage. A significant portion of hyper-DMRs were distributed in a cell-type-specific manner,  
 361 in which they were enriched for the binding of FOXA2, a pioneer TF essential for pancreatic  
 362 endoderm differentiation, and showed remarkable decreases in chromatin activity upon TET  
 363 inactivation. It is noteworthy that we unveiled a unique TF binding pattern associated with a stepwise  
 364 increase of 5hmC during pancreatic differentiation. The presence of GATA, FOXA, and HNF6  
 365 binding motifs in a temporal manner suggests that TET proteins are recruited to chromatin by a set

366 of lineage-specific TFs sequentially expressed during pancreatic differentiation. More intriguingly,  
367 nearly 800 hyper-DMRs recently identified in T2D pathogenesis <sup>31</sup> overlapped with hyper-DMRs  
368 found in TET-deficient cells in the present study, implying that TET dioxygenases are crucial for  $\beta$ -  
369 cell function.

370 During lineage specification, chromatin structure is dynamically changed between ‘closed’ and  
371 ‘open’ states. Open chromatin regions such as primed and active enhancers are accessible for TF and  
372 epigenetic modulator binding to initiate gene transcription. Proper chromatin remodeling is believed  
373 to be at least partly triggered by pioneer TFs, which can directly bind to nucleosomal DNA <sup>6</sup>. It was  
374 previously suggested that pioneer TF FOXA2 initiates chromatin remodeling and enhancer priming  
375 during pancreatic differentiation <sup>3,4</sup>. Interestingly, our analyses revealed that changes in 5hmC mirror  
376 the dynamic binding of FOXA2 in cells differentiated from hESCs through defined lineage  
377 intermediates toward pancreatic endocrine fate. We showed that chromatin activity was markedly  
378 decreased at the FOXA2 binding sites associated with DHMRs in TET-deficient cells, suggesting  
379 that lineage-specific hydroxymethylation is processed at a subset of FOXA2 target sites to influence  
380 chromatin opening.

381 How pioneer TFs initiate chromatin opening is still mostly unknown. One suggestion is that  
382 FOXA2 can recruit nucleosome remodeling complexes SWI/SNF to alter the surrounding chromatin  
383 structure <sup>45</sup>. Our present study demonstrates that TET1 binds to two distinct groups of regulatory  
384 elements: universal and lineage-specific TET1 sites. The fact that 1) pancreatic progenitor-specific  
385 TET1 targets are enriched for the FOXA2 binding motif and 2) co-occupancy of TET1 and FOXA2  
386 at distal regulatory elements is associated with more active chromatin suggests that TET1 residing in

specific genomic loci modulates FOXA2-mediated chromatin opening. Several studies suggest that TET can be directly recruited by pioneer TFs, such as FOXA1 and PU.1, to lineage-specific enhancers to facilitate local demethylation and gene induction<sup>46,47</sup>. However, a recent study found that chromatin opening requires cooperative binding of FOXA2 and additional TFs, such as GATA4, during endoderm differentiation<sup>37</sup>. This suggests complex cross-talk between pioneer TFs and chromatin remodelers in endodermal lineage specification. Further investigation is warranted to determine whether TET1 is recruited to distinct FOXA2-binding sites by lineage-specific TFs to synergistically regulate chromatin remodeling.

## Methods

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

**Pancreatic differentiation.** All differentiation experiments were repeated at least three times with three individual clones of the same phenotype. Pancreatic differentiation was performed as previously described with slight modifications<sup>48</sup>. In brief, hESCs were dissociated with Accutase (eBioscience) and seeded at a density of 19,000 cells/cm<sup>2</sup> on Matrigel in mTeSR1 supplemented with 10 μM Rho-associated protein kinase inhibitor Y-27632 (Miltenyi Biotec). Upon reaching 95% confluence, cells were exposed to differentiation medium with daily media feeding.

ES-to-DE (3 d): hESCs were exposed to 100 ng/ml Activin A (PeproTech) and 25 ng/ml Wnt-3a (PeproTech) in basal medium-I containing MCDB 131 medium (Gibco), 1.5 g/l NaHCO<sub>3</sub>, 1×

408 GlutaMAX (Gibco), 10 mM glucose, and 0.5% BSA for 1 day. For 2 additional days, cells were  
409 cultured in basal medium-I further supplemented with 100 ng/ml Activin A.

410 DE-to-GT (3 d): DE intermediates were incubated in basal medium-I supplemented with 50  
411 ng/ml FGF7 (PeproTech) for 3 days.

412 GT-to-PP (5 d): GT intermediates were cultured for 3 days in basal medium-II containing MCDB  
413 131 medium, 2.5 g/l NaHCO<sub>3</sub>, 1× GlutaMAX, 10 mM glucose, 2% BSA, and 1:200 ITS-X (Gibco),  
414 which was further supplemented with 50 ng/ml FGF7, 0.25 μM hedgehog inhibitor SANT-1 (Sigma),  
415 1 μM retinoic acid (Sigma), 100 nM BMP inhibitor LDN193189 (Stemgent), and 200 nM PKC  
416 activator TPB (Millipore). After 3 days of culture, cells were treated for 2 days with 2 ng/ml FGF7,  
417 0.25 μM SANT-1, 0.1 μM retinoic acid, 200 nM LDN193189, and 100 nM TPB in basal medium-II.

418 PP-to-PE (3 d): PP intermediates were differentiated in basal medium-III containing MCDB 131  
419 medium, 1.5 g/l NaHCO<sub>3</sub>, 1× GlutaMAX, 20 mM glucose, 2% BSA, and 1:200 ITS-X, which was  
420 further supplemented with 0.25 μM SANT-1, 0.05 μM retinoic acid, 100 nM LDN193189, 1 μM T3  
421 (3,3',5-Triiodo-L-thyronine sodium salt, Sigma), 10 μM ALK5 inhibitor II (Enzo Life Sciences), 10  
422 μM ZnSO<sub>4</sub>, and 10 μg/ml heparin (Sigma).

423 **Animal experiments.** All animal experiments were approved by the University of Macau Animal  
424 Ethics Committee. Immunocompromised SCID-beige mice were obtained from Charles River and  
425 maintained under a 12-h light/dark cycle with free access to standard mouse diet. For transplantation,  
426 ~5 million hESC-derived d14 cells were enzymatically dissociated and aggregated using ultra-low  
427 adherent plates (Corning) in basal medium-III further supplemented with 100 nM LDN193189, 1 μM  
428 T3, 10 μM ALK5 inhibitor II, 10 μM ZnSO<sub>4</sub>, 100 nM γ-secretase inhibitor XX (Millipore), and 10  
429 μM Y-27632 for 1 day. Cell aggregates were then transplanted into 6-week-old SCID-beige mice

430 under the kidney capsule as previously described <sup>49</sup>.

431 Glucose-stimulated human C-peptide secretion was assessed with mice 18 weeks post-  
432 transplantation. Blood samples were collected after overnight fasting and 30 min following an  
433 intraperitoneal injection of glucose (2 g/kg body weight). Human C-peptide levels in isolated plasma  
434 were quantified using the STELLUX Chemi Human C-peptide ELISA kit (ALPCO Diagnostics)  
435 according to the manufacturer's instructions.

436 **Generation of *TET* knock-out lines.** *TET* knock-out hESCs were generated using CRISPR/Cas9  
437 technology. gRNAs (Supplementary Table 5) were designed to target the sequences encoding exon 7  
438 of *TET1*, exon 3 of *TET2*, or exon 3 of *TET3* and cloned into pX330 vector (Addgene #42230) as  
439 previously described <sup>50</sup>. Constructs containing validated gRNAs were electroporated together with a  
440 vector expressing puromycin into hESCs using the P3 Primary Cell 4D-Nucleofector X kit (Lonza)  
441 following the manufacturer's instructions. The electroporated cells were plated at ~2,000 cells/cm<sup>2</sup>  
442 and cultured on Matrigel in mTeSR1 supplemented with 10 μM Y-27632 for 2 days. Successfully  
443 transfected cells were selected with 1 μg/ml puromycin in mTeSR<sup>TM</sup>1 and allowed to expand to form  
444 visible colonies from a single cell. Subsequently, clonal colonies were manually picked and reseeded  
445 individually into 24-well plates. The amplified colonies were analyzed by Sanger sequencing at  
446 targeted loci for the presence of Indel mutations.

447 ***TET1* expression cell lines.** *TET1* lentivirus expression plasmid was constructed by amplifying the  
448 human *TET1* ORF from FH-*TET1*-pEF (Addgene #49792) and inserting it into *KpnI* and *BamHI* sites  
449 of the lentiCRISPR v2 vector (Addgene #52961). Lentivirus particles were prepared as previously  
450 described <sup>51</sup>. For viral transduction, TKO-hESCs were grown on Matrigel in mTeSR1 using 24-well

451 plates and treated with 6  $\mu$ g/ml polybrene for 15 min at 37°C. Concentrated lentivirus particles (10  
452  $\mu$ l,  $1 \times 10^6$  TU/ml) were added to cell culture and incubated overnight at 37°C. On the next day, viral  
453 infection was repeated (30  $\mu$ l,  $1 \times 10^6$  TU/ml) to increase transduction efficiency. Infected cells were  
454 cultured in mTeSR1 for 2 days and then exposed to 1  $\mu$ g/ml puromycin for 10 days. The TKO line  
455 TKO-TET1, in which an intact *TET1* cDNA was expressed, was amplified and frozen down. Control  
456 cells (TKO-CRTL) carrying the lentiCRISPR v2 plasmid were established with the same  
457 experimental procedure.

458 **Locus-specific detection of 5mC.** Detection of 5mC content at particular CCGG sites was performed  
459 using the Epimark 5hmC and 5mC analysis kit (New England Biolabs) following the manufacturer's  
460 instructions. Briefly, genomic DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen)  
461 followed by RNase treatment. RNase-treated DNA samples were incubated with T4  $\beta$ -  
462 glucosyltransferase at 37°C for 16 h. The glycosylated DNA was subsequently digested with *MspI* or  
463 *HpaII* for 8 h at 37°C. Samples were treated with Proteinase K at 40°C for 30 min and then at 95°C  
464 for 10 min to inactivate the enzymes. Site-specific methylation contents were examined by RT-qPCR  
465 using the primers listed in Supplementary Table 6. The percentage of 5mC and unmodified cytosine  
466 were calculated using the comparative  $C_t$  method.

467 **5hmC dot blot assay.** Genomic DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen)  
468 according to the manufacturer's instructions. DNA was denatured in 1 M NaOH supplemented with  
469 25 mM EDTA at 95°C for 10 min and then neutralized with 2 M ice-cold ammonium acetate for 10  
470 min. Two-fold serial dilutions of the DNA samples were spotted onto nitrocellulose membrane. The  
471 air-dried membrane was fixed with UV irradiation (CL-1000 UV crosslinker, Ultra-Violet Products),

472 blocked with 5% non-fat milk, and incubated with a rabbit anti-5hmC antibody (1:10,000, Active  
473 Motif) followed by an HRP-conjugated anti-rabbit antibody (1:5,000, Jackson ImmunoResearch).  
474 Signal was visualized with SuperSignal West Pico PLUS chemiluminescent substrate (Thermo  
475 Fisher Scientific). The same membrane was subsequently stained with 0.02% methylene blue in 0.3  
476 M sodium acetate to ensure equal loading of input DNA.

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

483 **Immunocytochemical analysis.** Differentiated cells were fixed in 4% paraformaldehyde for 30 min  
484 at room temperature, washed, and then permeabilized with 0.15% Triton X-100 at room temperature  
485 for 1 h. Following blocking with 5% normal donkey serum (Jackson Immuno Research Laboratories),  
486 samples were incubated with primary antibodies at 4°C overnight and then appropriate secondary  
487 antibodies for 1 h at room temperature. Images were acquired using the Zeiss Axio Observer  
488 microscope. For transplant grafts, tissues were fixed with 4% paraformaldehyde overnight at 4°C,  
489 washed with PBS, and subsequently exposed to 30% sucrose overnight at 4°C. Samples were  
490 mounted with Optimal Cutting Temperature Compound (Tissue-Tek) and sectioned at 10 µm.  
491 Immunofluorescent staining was performed on cryosections as described above. Antibodies used for  
492 immunofluorescent staining are listed in Supplementary Table 7.



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

499 **RNA-seq library construction and data analysis.** Total RNA was isolated using Trizol reagent  
 500 following the manufacturer's instructions. RNA integrity was determined using an Agilent 2100  
 501 Bioanalyzer (Agilent Technologies). Subsequently, polyA-tailed RNA was selected using Dynabeads  
 502 oligo(dT) (Thermo Fisher Scientific), and libraries were prepared using the NEBNext Ultra RNA  
 503 Library Prep kit for Illumina (New England Biolabs). Libraries were subjected to high-throughput  
 504 sequencing on an Illumina HiSeq 2500 system (150 bp, paired-end) at Novogene (Tianjin, China).  
 505 To process the sequencing data, low-quality bases and adaptor were trimmed using TrimGalore  
 506 v.0.5.0 (options: --quality 20 and --length 50, <https://github.com/FelixKrueger/TrimGalore>). Clean  
 507 reads were aligned to the hg38 reference genome using STAR v.2.5.3<sup>52</sup> with default parameters, and  
 508 only uniquely mapped reads were used for downstream analysis. A count matrix for each gene was  
 509 generated using htseq-count (HTSeq package<sup>53</sup>). DESeq2<sup>54</sup> was used to identify significant DEGs  
 510 in knock-out samples compared with WT samples at different differentiation stages (fold change (FC)  
 511  $\geq 2$ ; False Discovery Rate (FDR)  $< 0.05$ ). Hierarchical cluster analysis<sup>55</sup> of the union DEGs was used  
 512 to determine stage-specific signature genes. The 'ClusterProfiler' package in R<sup>56</sup> was used for the  
 513 functional enrichment analysis of DEGs in KEGG pathways.

514 **ATAC-seq library preparation and data analysis.** ATAC-seq libraries were prepared as previously  
 515 described<sup>57</sup>. In brief, cells were enzymatically dissociated and lysed in lysis buffer (10 mM Tris-HCl,  
 516 pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% IGEPAL CA-630). Immediately after centrifugation,  
 517 transposition reactions were carried out by adding Tn5 transposases from the Illumina Nextera DNA  
 518 library preparation kit to the isolated nuclei and incubation at 37°C for 30 min. DNA fragments were  
 519 purified using the MinElute PCR Purification kit (Qiagen) and amplified using the KAPA real-time  
 520 library amplification kit (Roche). Libraries were purified using PCRClean DX beads (Aline  
 521 Biosciences) and subsequently subjected to high-throughput sequencing on an Illumina NextSeq 500  
 522 instrument (75 cycle, paired-end).

523 Adaptor trimming of raw reads was performed using TrimGalore v0.5.0 (options: --quality 20  
 524 and --length 20), and high-quality ( $Q \geq 20$ ) reads were uniquely aligned to the hg38 reference genome  
 525 using Bowtie2 with the '--very-sensitive' option. Reads mapped to mitochondrial DNA and PCR  
 526 duplicate reads were removed, and uniquely mapped reads were extracted for downstream analysis.  
 527 Genrich v.0.5 (<https://github.com/jsh58/Genrich>) with ATAC-seq mode (option: -j,-q 0.01) was  
 528 applied for each sample (with two biological replicates) to call ATAC peaks. In total, 52,817 and  
 529 38,697 peaks were detected in WT\_PP and TKO\_PP cells, respectively. Bedtools intersect was used  
 530 to count the reads falling into 124,322 non-overlapping peak regions, and significant DARs were  
 531 detected using DESeq2 by the criteria  $FC \geq 2$  and  $FDR < 0.05$ . Volcano plots were generated using  
 532 the R package ggplot2. Motif annotation of DARs was performed using HOMER software<sup>58</sup>.

533 **CMS-IP-seq library preparation and data analysis.** CMS-IP-seq libraries were performed as  
 534 previously described with minor modifications (Huang et al., 2012). Purified genomic DNA (with 5%

535 mouse DNA and 0.5% lambda DNA spike-in) was sheared to 200–500 bp fragments using a M220  
536 Focused-ultrasonicator (Covaris). Bisulfite-converted DNA libraries with methylated adapters were  
537 enriched using an in-house anti-CMS antibody bound to protein A/G Dynabeads. Amplified libraries  
538 were purified by AmpuXP beads (Beckman Coulter) and then sequenced using the Illumina NextSeq  
539 (75 and 40 cycles, single-end) system.

540 Analysis of CMS-IP data was performed by in-house software ‘CMSIP’ v0.1.1  
541 (<https://github.com/lijinbio/cmsip>). Briefly, raw reads were mapped to hg38 and mm10 genome  
542 references using bsmmap v.2.89<sup>59</sup> (options: -n 1 -q 3 -r 0). After removing PCR duplicates and reads  
543 mapped to both human and spike-in mouse genome, species-specific reads were used to perform  
544 normalization for each sample according to the spike-in size factors. The mean wigsum for every  
545 200-bp window of the whole human genome was used to call the hydroxymethylation-enriched peak  
546 regions for each group against input. In total, 75,324 and 503 peaks were detected in WT\_PP and  
547 TKO\_PP samples, respectively. Differentially DHMRs between WT\_PP and TKO\_PP were  
548 identified (g-test; FDR < 0.05). GREAT analysis with single-nearest genes option was used to  
549 perform functional annotation of hypo-DHMRs.

550 CMS-IP-seq datasets of hESCs and multiple pancreatic lineage intermediates (DE, GT, and PP)  
551 were downloaded from GSE97992<sup>21</sup> and mapped to the hg38 reference genome using bsmmap with  
552 the same parameters as in the previous analysis. Bam2wig.py in RSeQC<sup>60</sup> was used to transform the  
553 bam file to normalized bigWig files (option: -t 20000000000). The average 5hmC signals within each  
554 hypo-DHMR for all stages were calculated using deepTools<sup>61</sup>. Compared with hESCs, increased  
555 5hmC peaks were defined as peaks in PP cells with a  $\geq 1.5$ -fold increase.

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

564 For data analysis, paired-end 100 bp reads were mapped against hg38 using bsmep v.2.89<sup>59</sup> with  
 565 paired mode (options: -n 1 -q 3 -r 0), and only uniquely mapped reads were retained. More than 26.6  
 566 million CpG sites with coverage of  $\geq 5$  reads were detected in both WT\_PP and TKO\_PP samples,  
 567 which were used for downstream analyses. BSeQC<sup>62</sup> and the mcall module in MOABS<sup>29</sup> was applied  
 568 to perform quality control and calculate the methylation ratio for each CpG site (options: --  
 569 trimWGBSEndRepairPE2Seq 40). Bisulfite conversion efficiencies were estimated using spike-in  
 570 unmethylated lambda phage DNA. The Mcomp module was used to call significant DMCs and DMRs  
 571 (absolute credible difference of DNA methylation ratio  $> 20\%$  and adjusted  $p < 0.05$ ). The output  
 572 bedGraph files from mcall included single-base resolution DNA methylation ratios, which were  
 573 transformed into bigwig file format. The bigwig files were uploaded to the UCSC genome browser  
 574 for visualization. Motif enrichment analysis of DMRs was performed using HOMER software,  
 575 functional annotation was performed using GREAT<sup>20</sup> with default settings, and many plots related  
 576 to WGBS was performed using MMINT.

577 **ChIP-seq library preparation and data analysis.** ChIP-seq was performed for TET1 (WT\_PP cells),  
 578 H3K4me1 (WT\_PP and TKO\_PP cells), and H3K27ac (WT\_PP and TKO\_PP cells). For TET1 ChIP-  
 579 seq,  $\sim 5 \times 10^7$  cells were crosslinked, washed, and snap-frozen following the Cell Fixation protocol  
 580 from Active Motif. TET1 ChIP-seq library preparation and sequencing were conducted by Active  
 581 Motif. Chromatin immunoprecipitation of H3K4me1 and H3K27ac were performed as previously  
 582 described<sup>28</sup>. WT\_PP and TKO\_PP cells were fixed, washed, and lysed in nuclear lysis buffer (50  
 583 mM Tris HCl, pH 8.0, 5 mM EDTA, 1% SDS, and 1× protease inhibitor cocktail). Chromatin was  
 584 sheared to 200–500-bp fragments using Bioruptor (Diagenode), and DNA fragments were  
 585 precipitated with appropriate antibodies. ChIP-seq libraries were prepared using the NEBNext Ultra  
 586 II DNA Library Prep kit (New England Biolabs) following the manufacturer's instruction and  
 587 subjected to high-throughput sequencing on an Illumina HiSeq 2500 system (150 bp, paired-end) at  
 588 Novogene (Tianjin, China).

589 Quality control and alignment of raw reads of ChIP-seq data were performed similarly to the  
 590 ATAC-seq data analysis described above. TrimGalore (options: --quality 20 and --length 50) was  
 591 used to remove the adaptor, and only uniquely mapped reads were retained. Bam2wig.py was used  
 592 to transform the bam file to normalized bigWig files (option: -t 2000000000). For WT\_PP TET1  
 593 ChIP-seq, Macs2<sup>63</sup> was used to call ChIP-seq enriched peak regions with default parameters.

594 **Integration of analyses.** RNA-seq, WGBS, CMS-IP-seq, ATAC-seq and ChIP-seq library  
 595 preparations were performed as previously described. Detail experimental procedures can be found  
 596 in Supplementary Materials. The number of mapped reads and mapped ratios are listed in  
 597 Supplementary Table 8. To compare 5hmC and FOXA2 enriched signals, FOXA2 ChIP-seq datasets  
 598 of multiple pancreatic lineage intermediates were downloaded from GSE117136. Data were analyzed

599 as described above. Deeptools was used to calculate average signals at pancreas-specific hypo-  
600 DHMRs (up- and down- 3 kb) in each group. The FOXA2 bound sites were detected using Macs2 <sup>63</sup>.

601 To compare DNA methylation between TKO\_PP and WT\_PP cells at annotated genomic  
602 features, ChIP-seq datasets of FOXA2, GATA4, GATA6, and PDX1 were downloaded from  
603 GSE117136. Bivalent promoters and poised and active enhancers in hESC-derived pancreatic  
604 progenitors were downloaded from EMTAB1086 and GSE54471, respectively. The analysis was  
605 similar to that described above.

606 **Quantification and statistical analysis.** All statistical analyses were performed using GraphPad  
607 Prism software. Student's unpaired two-tailed *t*-tests were used for qRT-PCR and flow cytometry  
608 experiments. Student's paired one-tailed *t*-tests were used for the glucose-stimulated human C-  
609 peptide secretion experiment. Quantification data are presented as mean ± SD. For all statistical  
610 analyses, \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.001, and \*\*\*\**p* < 0.0001. Statistical analyses for RNA-seq,  
611 WGBS, CMS-IP-seq, and ChIP-seq data are described in the corresponding sections.

612

## 613 **Data availability**

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

619

## 620 **Acknowledgments**

621 We thank the University of Macau, Faculty of Health Sciences, Animal Research Facility for  
622 animal housing. This work was supported by the National Natural Science Foundation of China  
623 (NSFC 31701276) and the University of Macau Multi-Year Research Grant (MYRG2016-00065-  
624 FHS) to R.X.

625

## 626 **Author contributions**

627 R.X. conceived the project. R.X. and D.S. directed and oversaw the project. X.W. performed most  
628 experiments and collected data. J.F.L. performed computational analysis for WGBS, RNA-seq, CMS-  
629 IP-seq, ATAC-seq, and ChIP-seq. M.L. prepared the CMS-IP library and performed sequencing. Q.L.  
630 assisted with the animal studies. J.L. provided expert advice on bioinformatic analyses. Y.H. critically  
631 reviewed the manuscript. R.X., X.W., D.S., and J.F.L. wrote the manuscript; all other authors  
632 provided editorial advice.

633

634    **Competing interests**

635    The authors declare that no conflicts of interest exist.



# References

1. Buecker, C. & Wysocka, J. Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet.* **28**, 276-284 (2012).
2. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167**, 1170-1187 (2016).
3. Wang, A. et al. Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* **16**, 386-399 (2015).
4. Lee, K. et al. FOXA2 is required for enhancer priming during pancreatic differentiation. *Cell Rep* **28**, 382-393 (2019).
5. Cirillo, L. A. & Zaret, K. S. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol. Cell* **4**, 961-969 (1999).
6. Cirillo, L. A. et al. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279-289 (2002).
7. Iwafuchi-Doi, M. et al. The pioneer transcription factor foxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell* **62**, 79-91 (2016).
8. Serandour, A. A. et al. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res.* **21**, 555-565 (2011).
9. Donaghey, J. et al. Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat. Genet.* **50**, 250-258 (2018).
10. He, Y. F. et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307 (2011).
11. Ito, S. et al. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129-1133 (2010).
12. Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300-1303 (2011).
13. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935 (2009).
14. Zhang, H. et al. TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res.* **20**, 1390-1393 (2010).
15. Tsagaratou, A. et al. Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. *Proc. Natl. Acad. Sci. USA* **111**, E3306-3315 (2014).
16. Tan, L. et al. Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method. *Nucleic Acids Res.* **41**, e84 (2013).
17. Kim, M. et al. Dynamic changes in DNA methylation and hydroxymethylation when hES cells undergo differentiation toward a neuronal lineage. *Hum. Mol. Genet.* **23**, 657-667 (2014).
18. Verma, N. et al. TET proteins safeguard bivalent promoters from de novo methylation in human embryonic stem cells. *Nat. Genet.* **50**, 83-95 (2018).
19. Ko, M. et al. Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc. Natl. Acad. Sci. USA* **108**, 14566-

- 14571 (2011).
20. Fang, S. et al. Tet inactivation disrupts YY1 binding and long-range chromatin interactions during embryonic heart development. *Nat. Commun.* **10**, 4297 (2019).
21. Li, J. et al. Decoding the dynamic DNA methylation and hydroxymethylation landscapes in endodermal lineage intermediates during pancreatic differentiation of hESC. *Nucleic Acids Res.* **46**, 2883-2900 (2018).
22. Lu, F. L., Liu, Y. T., Jiang, L., Yamaguchi, S. & Zhang, Y. Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev.* **28**, 2103-2119 (2014).
23. Rodriguez-Segui, S., Akerman, I. & Ferrer, J. GATA believe it: new essential regulators of pancreas development. *J. Clin. Invest.* **122**, 3469-3471 (2012).
24. Kaestner, K. H. The FoxA factors in organogenesis and differentiation. *Curr. Opin. Genet. Dev.* **20**, 527-532 (2010).
25. Zhang, H. et al. Multiple, temporal-specific roles for HNF6 in pancreatic endocrine and ductal differentiation. *Mech. Dev.* **126**, 958-973 (2009).
26. Lopez-Moyado, I. F. et al. Paradoxical association of TET loss of function with genome-wide DNA hypomethylation. *Proc. Natl. Acad. Sci. USA* **116**, 16933-16942 (2019).
27. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* **20**, 116-117 (1998).
28. Xie, R. et al. Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells. *Cell Stem Cell* **12**, 224-237 (2013).
29. Sun, D. et al. MOABS: model based analysis of bisulfite sequencing data. *Genome Biol.* **15**, R38 (2014).
30. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
31. Volkov, P. et al. Whole-genome bisulfite sequencing of human pancreatic islets reveals novel differentially methylated regions in type 2 diabetes pathogenesis. *Diabetes* **66**, 1074-1085 (2017).
32. Akerman, I. et al. Human pancreatic beta cell lncRNAs control cell-specific regulatory networks. *Cell Metab.* **25**, 400-411 (2017).
33. Sosa-Pineda, B., Chowdhury, K., Torres, M., Oliver, G. & Gruss, P. The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. *Nature* **386**, 399-402 (1997).
34. Nelson, S. B., Schaffer, A. E. & Sander, M. The transcription factors Nkx6.1 and Nkx6.2 possess equivalent activities in promoting beta-cell fate specification in Pdx1<sup>+</sup> pancreatic progenitor cells. *Development* **134**, 2491-2500 (2007).
35. Ohta, Y. et al. Convergence of the insulin and serotonin programs in the pancreatic beta-cell. *Diabetes* **60**, 3208-3216 (2011).
36. Wu, H. et al. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389-393 (2011).
37. Cernilogar, F. M. et al. Pre-marked chromatin and transcription factor co-binding shape the pioneering activity of Foxa2. *Nucleic Acids Res.* **47**, 9069-9086 (2019).
38. Ahlgren, U., Jonsson, J., Jonsson, L., Simu, K. & Edlund, H. Beta-cell-specific inactivation

of the mouse *Ipfl/Pdx1* gene results in loss of the beta-cell phenotype and maturity onset diabetes. *Genes Dev.* **12**, 1763-1768 (1998).

39. Sussel, L. et al. Mice lacking the homeodomain transcription factor *Nkx2.2* have diabetes due to arrested differentiation of pancreatic beta cells. *Development* **125**, 2213-2221 (1998).

40. Collombat, P. et al. Opposing actions of *Arx* and *Pax4* in endocrine pancreas development. *Genes Dev.* **17**, 2591-2603 (2003).

41. Dhawan, S., Georgia, S., Tschen, S. I., Fan, G. & Bhushan, A. Pancreatic beta cell identity is maintained by DNA methylation-mediated repression of *Arx*. *Dev. Cell* **20**, 419-429 (2011).

42. Liu, J. et al. *Neurog3*-independent methylation is the earliest detectable mark distinguishing pancreatic progenitor identity. *Dev. Cell* **48**, 49-63 (2019).

43. Papizan, J. B. et al. *Nkx2.2* repressor complex regulates islet beta-cell specification and prevents beta-to-alpha-cell reprogramming. *Genes Dev.* **25**, 2291-2305 (2011).

44. Wu, X., Li, G. & Xie, R. Decoding the role of TET family dioxygenases in lineage specification. *Epigenetics Chromatin* **11**, 58 (2018).

45. Serandour, A. A. et al. Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res* **40**, 8255-8265 (2012).

46. Yang, Y. A. et al. *FOXA1* potentiates lineage-specific enhancer activation through modulating TET1 expression and function. *Nucleic Acids Res.* **44**, 8153-8164 (2016).

47. Lio, C. W. et al. Tet2 and Tet3 cooperate with B-lineage transcription factors to regulate DNA modification and chromatin accessibility. *Elife* **5**, e18290 (2016).

48. Rezania, A. et al. Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 1121-1133 (2014).

49. Zmuda, E. J., Powell, C. A. & Hai, T. A method for murine islet isolation and subcapsular kidney transplantation. *J. Vis. Exp.* **50**, 2096 (2011).

50. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).

51. Tu, S. et al. Takusan: a large gene family that regulates synaptic activity. *Neuron* **55**, 69-85 (2007).

52. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).

53. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).

54. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

55. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).

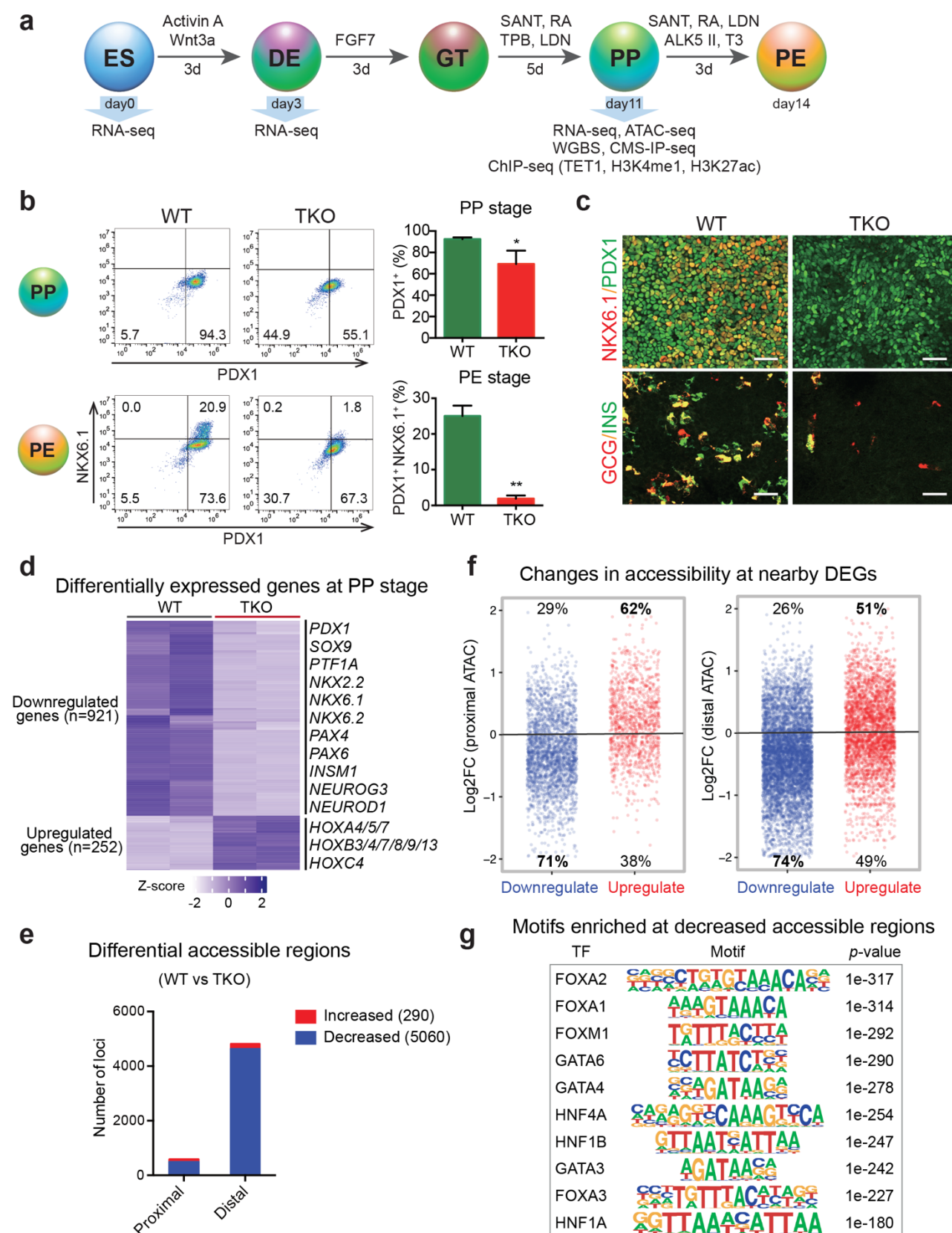
56. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287 (2012).

57. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.21-21.29.29 (2015).

58. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576-589 (2010).

- 761 59. Cowley, M. et al. Effects of cadmium exposure on DNA methylation at imprinting control  
762 regions and genome-wide in mothers and newborn children. *Environ. Health Perspect.* **126**,  
763 037003 (2018).
- 764 60. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics*  
765 **28**, 2184-2185 (2012).
- 766 61. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. DeepTools: a flexible platform  
767 for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187-191 (2014).
- 768 62. Lin, X. et al. BSeQC: quality control of bisulfite sequencing experiments. *Bioinformatics* **29**,  
769 3227-3229 (2013).
- 770 63. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using  
771 MACS. *Nat. Protoc.* **7**, 1728-1740 (2012).

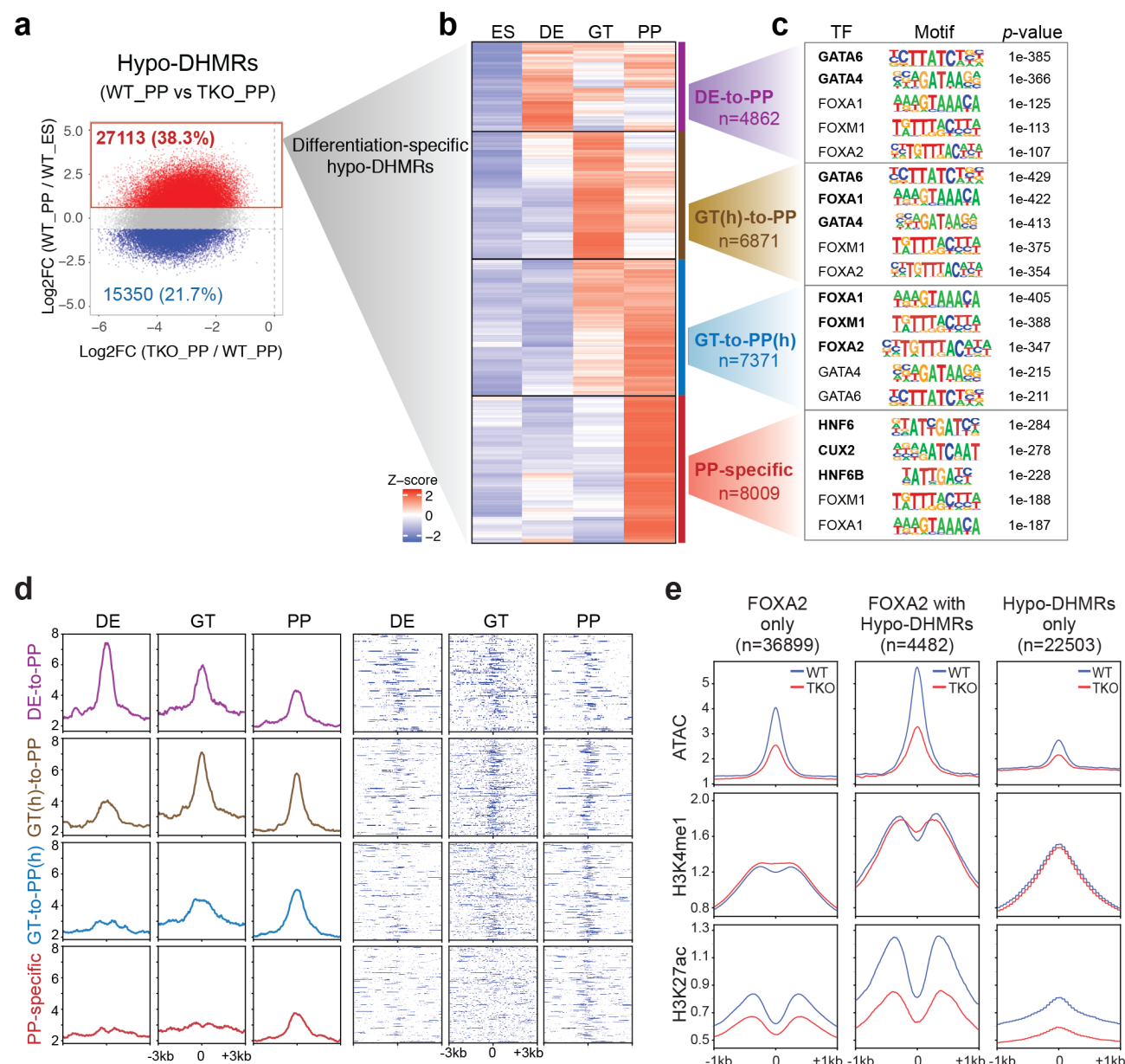
772





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

**a** Stepwise differentiation of hESCs to pancreatic endocrine cells. Wild-type (WT) and *TET1/TET2/TET3* triple knock-out (TKO) hESC lines were differentiated and analyzed by RNA-seq, WGBS, CMS-IP-seq, ATAC-seq, and ChIP-seq at the indicated stages (ES: embryonic stem cell; DE: definitive endoderm; GT: primitive gut tube; PP: pancreatic progenitor; PE: pancreatic endocrine). **b** Representative flow cytometry plots for PDX1 and NKX6.1 in WT or TKO cells at the PP and PE stages. Quantification of the percentage of PDX1<sup>+</sup> NKX6.1<sup>+</sup> cells is shown in the right panel. *n* = 3 independent differentiations. **c** Immunostaining of PDX1/NKX6.1 at the PP stage and insulin (INS)/glucagon (GCG) at the PE stage are shown in WT and TKO cells. Scale bar = 50  $\mu$ m. **d** Genes with significant changes in expression in TKO cells relative to WT cells at the PP stage. Differentially expressed genes (DEGs) are classified into down- and upregulated groups. Each row corresponds to one individual gene and each column to a different biological repeat. The color scale from white to blue represents Z-score normalized gene expression levels from low to high (FDR < 0.05). **e** Number of regions with significant changes in chromatin accessibility upon TET depletion at proximal ( $\leq$  1 kb from TSS) and distal (> 1 kb from TSS) regions (*p* < 0.05). **f** Dot plots depicting the ratios (WT\_PP over TKO\_PP) of ATAC signals (y-axis) at proximal (left) or distal (right) regions of DEGs that were downregulated (blue) or upregulated (red) in TKO\_PP cells compared with WT\_PP cells. Values indicate the fraction of increased or decreased chromatin accessibility regions in down- and upregulated genes, respectively. **g** Transcription factor (TF) motif enrichment analysis of genomic regions showed significantly decreased ATAC-seq signals in TKO\_PP cells. The top 10 significant motifs are shown after removing redundant motifs.

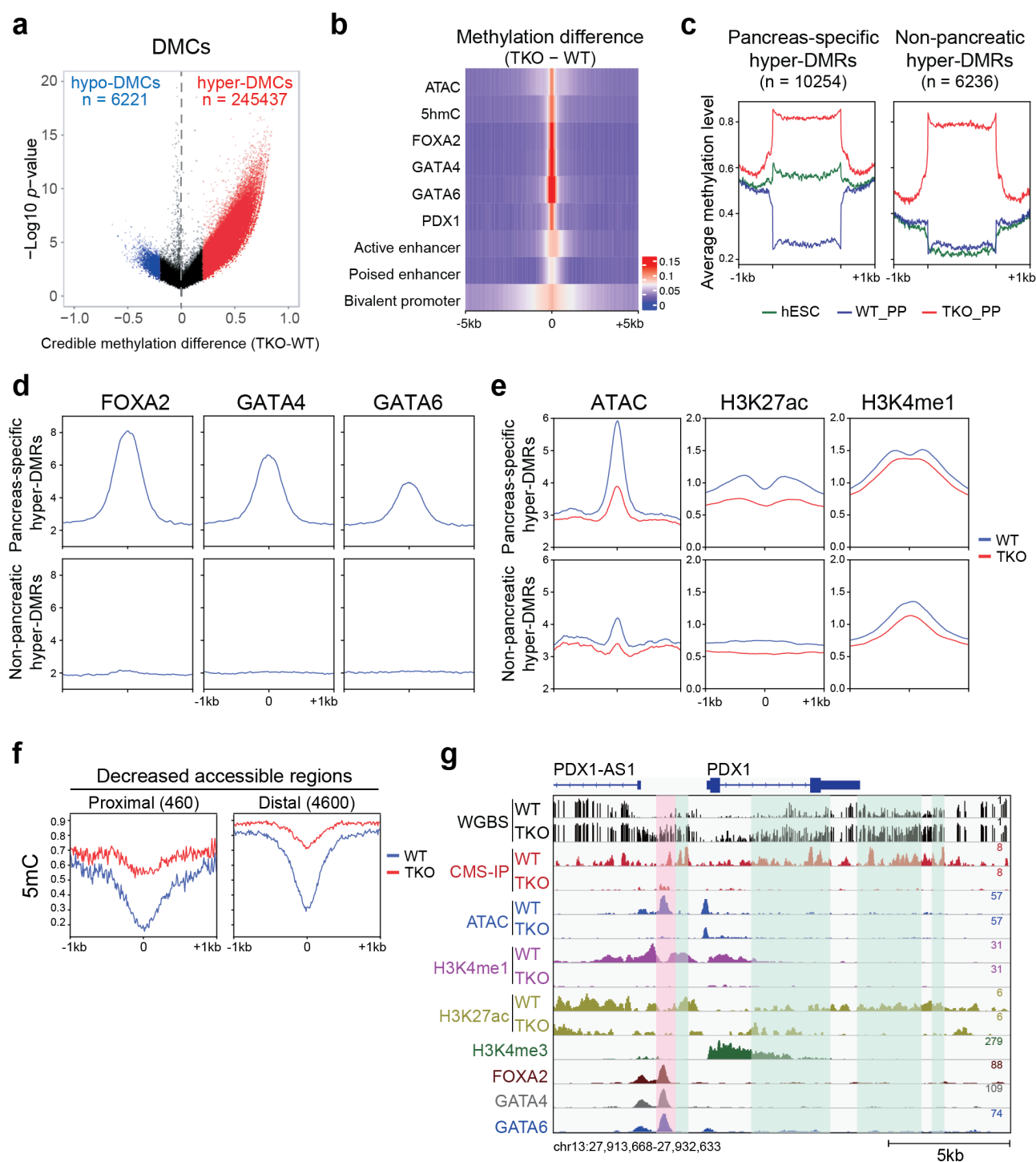


**Fig. 2 Loss of TET disrupts epigenetic dynamics at pancreas-specific hypo-DHMRs associated with distinct pioneer TFs motifs.**

**a** Scatterplot presenting 5hmC differences between pancreatic progenitors (WT\_PP) and hESCs (WT\_ES) at hypo-DHMRs identified in TKO\_PP cells. Red indicates increased 5hmC (fold change > 1.5), and blue shows decreased 5hmC in pancreatic progenitors relative to hESCs. **b** Clustering of 5hmC signals in WT cells at four differentiation stages (ES, DE, GT, and PP). Each row represents one hypo-DHMR showing increased 5hmC in pancreatic progenitors compared with hESCs. Red

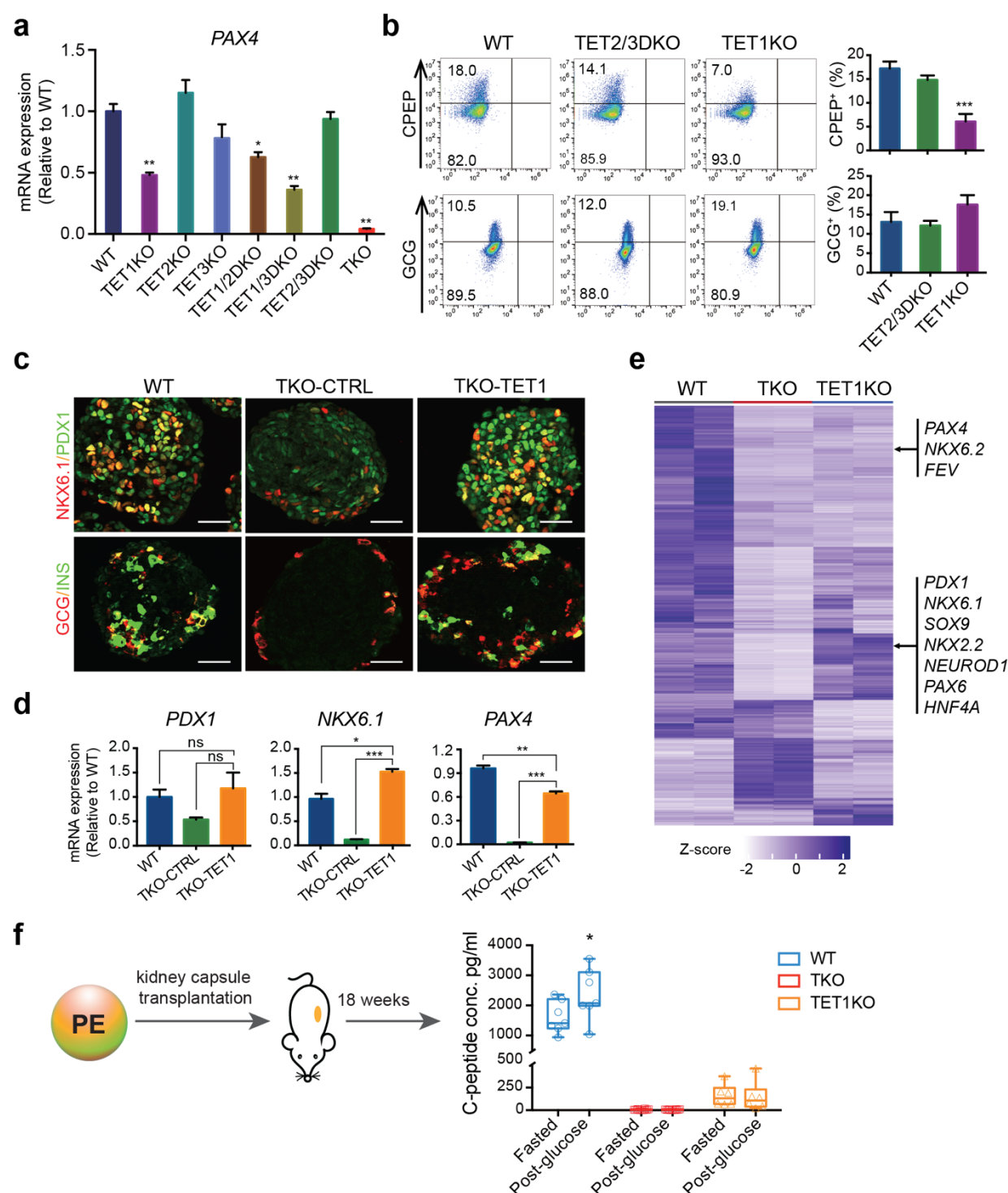
803 represents high 5hmC; blue indicates low 5hmC. **c** Motif enrichment analysis among the four clusters  
 804 defined in **b**. The top five known motifs are shown. **d** Heatmap and density plots showing dynamic  
 805 changes of FOXA2 in the DE, GT, and PP lineage intermediates across the four clusters defined in **b**.  
 806 The color scale from white to blue represents normalized FOXA2 signal from low to high. **e**  
 807 Enrichment profile of chromatin accessibility (ATAC), H3K4me1, and H3K27ac at regions of  
 808 FOXA2 binding alone (left column), FOXA2 binding together with ‘differentiation-specific hypo-  
 809 DHMRs’ (middle column), and ‘differentiation-specific hypo-DHMRs’ alone (right column). The  
 810 WT signal is marked in blue and the TKO signal in red.





**Fig. 3 Pancreas-specific hyper-DMRs show reduced chromatin activity during pancreatic differentiation.**

**a** Volcano plot of WGBS data illustrating differentially methylated CpGs (DMCs) identified in TKO\_PP cells compared with WT\_PP cells. Red and blue represent increased and decreased 5mC in TKO\_PP cells, respectively (credible methylation difference > 0.2). **b** Heatmap illustrating methylation difference between TKO\_PP and WT\_PP cells at centers of annotated genomic features ( $\pm 5$  kb) for chromatin accessibility (ATAC), hydroxylation (5hmC), TF binding (FOXA2, GATA4, GATA6, and PDX1), bivalent promoters, poised enhancers, and active enhancers. Average 5mC signals of every 100-bp bin were calculated. **c** Classification of TKO hyper-DMRs based on 5mC levels in hESCs (green), WT\_PP cells (blue), and TKO\_PP cells (red). **d** Average plots of FOXA2 (left column), GATA4 (middle column), and GATA6 (right column) signal at pancreas-specific hyper-DMRs or non-pancreatic hyper-DMRs in pancreatic progenitors. **e** Average plots of ATAC (left column), H3K27ac (middle column), and H3K4me1 (right column) at pancreas-specific hyper-DMRs or non-pancreatic hyper-DMRs in WT\_PP (blue) and TKO\_PP (red) cells. **f** Average plots of 5mC at proximal ( $\leq 1$  kb from TSS) and distal ( $> 1$  kb from TSS) decreased accessible regions in WT\_PP (blue) and TKO\_PP (red) cells. **g** Genome-browser view of the *PDX1/PDX1-AS1* locus. Four type 2 diabetes-associated islet hyper-DMRs<sup>31</sup> overlapping with TKO hyper-DMRs are highlighted in green. A specific TKO hyper-DMR showing decreased ATAC, H3K4me1, and H3K27ac signals is highlighted in pink.



**Fig. 4 TET1 is required in pancreatic  $\beta$ -cell specification.**

**a** Expression of *PAX4* in WT, TET1KO, TET2KO, TET3KO, TET1/2DKO, TET1/3DKO, TET2/3DKO, and TKO cells at the PP stage by RT-qPCR. RT-qPCR validation was performed with three independent batches of samples. **b** Representative plots of flow cytometry of human C-peptide

836 (CPEP) and glucagon (GCG) in WT, TET1KO, and TET2/3DKO cells at the PE stage.

837 Quantifications of the percentage of CPEP<sup>+</sup> or GCG<sup>+</sup> cells are shown in the right panel. n = 3

838 independent differentiation. **c** Immunostaining of NKX6.1, PDX1, INS, and GCG in WT, TKO-

839 CTRL, and TKO-TET1 cells at the PE stage of pancreatic differentiation. Scale bar = 50  $\mu$ m. **d**

840 Expression of *PDX1*, *NKX6.1*, and *PAX4* in WT, TKO-CTRL, and TKO-TET1 cells at the PP stage

841 by RT-qPCR. RT-qPCR validation was performed with three independent batches of samples. **e**

842 Heatmap showing the hierarchical clustering of DEGs among WT\_PP, TKO\_PP, and TET1KO\_PP

843 cells. Each row represents one DEG, and each column represents one biological replicate. The color

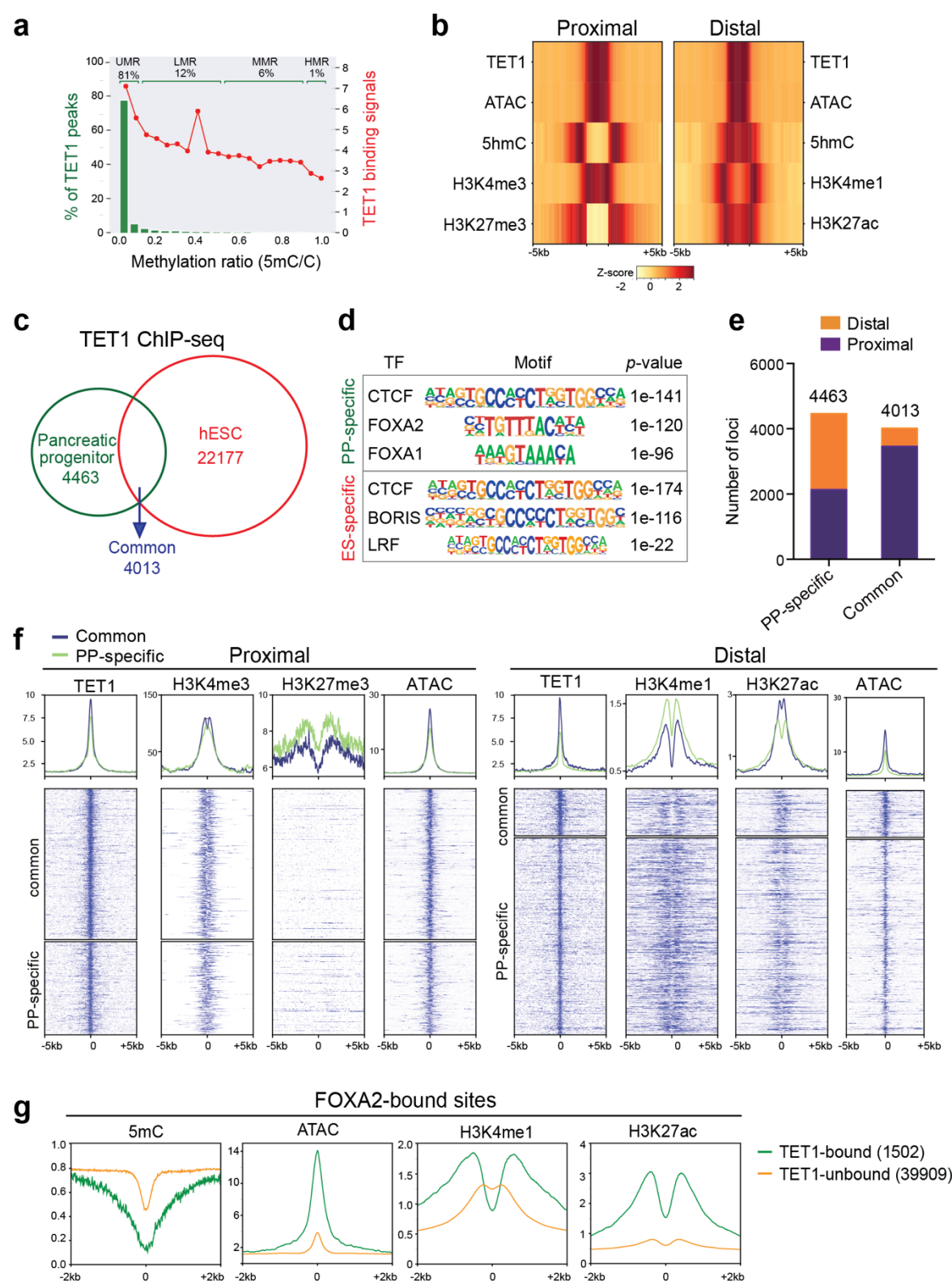
844 scale from white to blue represents normalized gene expression levels from low to high (FDR < 0.05).

845 **f** WT, TKO, and TET1KO cells were differentiated to the PE stage and transplanted under the kidney

846 capsule of nondiabetic female SCID-beige mice. Eighteen weeks post-implantation, human C-peptide

847 levels were measured after an overnight fast and 30 min following an i.p. glucose injection. C-peptide

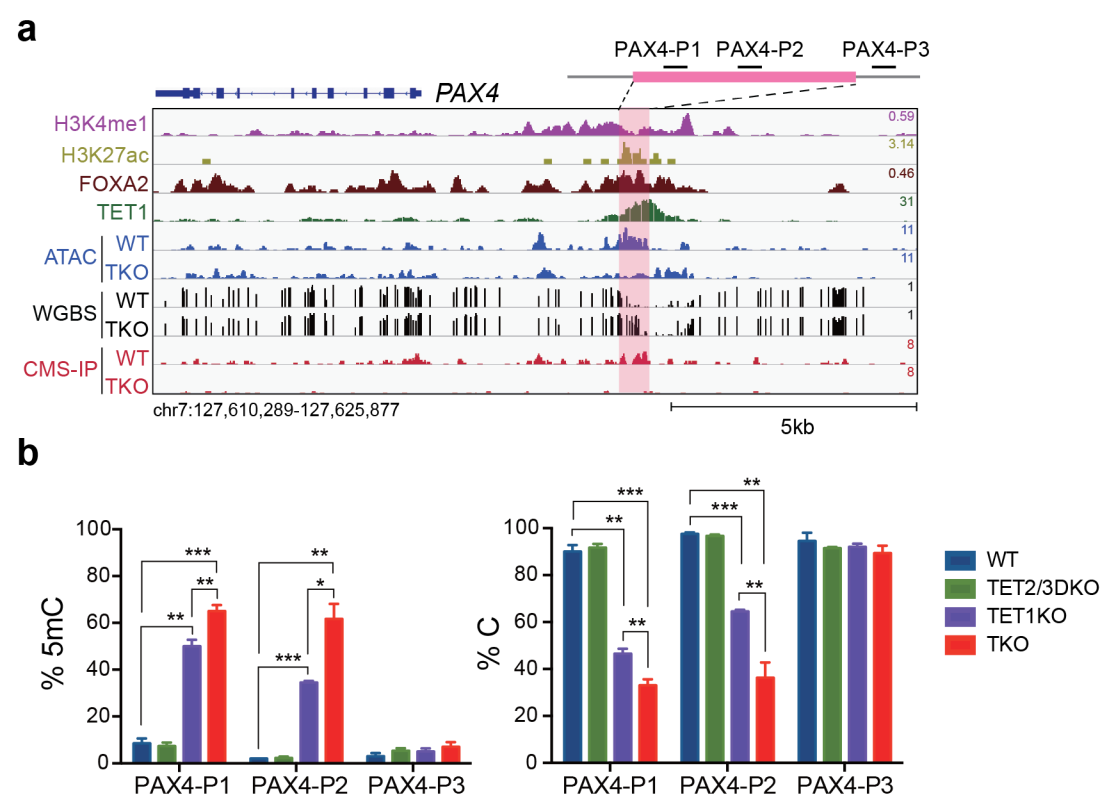
848 levels from individual mice are shown in box and whisker plots.



**Fig. 5 TET1 binding sites are distinct in pancreatic progenitors.**

**a** Methylation levels at TET1 binding sites in WT\_PP cells. The x-axis represents the average methylation ratio of CpG sites located within each peak (at 5% methylation ratio intervals). The green

853 bar represents the percentage of TET1 peaks within the corresponding methylation ratio interval (left  
854 y-axis). The red curve indicates TET1-enriched signals within each corresponding methylation ratio  
855 interval (right y-axis). The numerical labels at the top represent the fraction of the four class peaks  
856 with different average methylation ratios: UMR, unmethylated regions (0-0.1); LMR, low methylated  
857 regions (0.1-0.5); MMR, medium methylated regions (0.5-0.9); and HMR, high methylated regions  
858 (0.9-1). **b** Heatmaps demonstrating enrichment patterns of ATAC, 5hmC, and histone modification  
859 (H3K4me1, H3K27me3, H3K4me1, and H3K27ac) relative to the center of TET1 peaks ( $\pm 5$  kb) at  
860 proximal ( $\leq 1$  kb from TSS) and distal ( $> 1$  kb from TSS) regions. **c** Venn diagram illustrating distinct  
861 and overlapping TET1 binding sites between hESCs and pancreatic progenitors ( $p < 0.05$ ). **d** TF motif  
862 enrichment analysis of TET1 targets identified specifically in pancreatic progenitors (top, PP-specific)  
863 or hESCs (bottom, ES-specific). The top three motifs are shown. **e** Number of TET1 binding sites  
864 identified specifically in pancreatic progenitors (PP-specific) or in both hESCs and pancreatic  
865 progenitors (common) at proximal and distal regions ( $p < 0.05$ ). **f** Average plots (top) and heatmaps  
866 (bottom) of TET1, H3K4me3, H3K27me3, H3K4me1, H3K27ac, and ATAC at ‘common’ or ‘PP-  
867 specific’ TET1 peaks ( $\pm 5$  kb) at proximal or distal regions. Heatmaps are ranked by decreased TET1-  
868 binding signal of ‘common’ and ‘PP-specific’ regions, respectively. The color scale from white to  
869 blue represents the normalized signal from low to high. **g** Distribution profiles of average methylation  
870 (5mC), ATAC, H3K4me1, and H3K27ac located at FOXA2-bound sites with (green) or without  
871 (orange) co-binding of TET1.



**Fig. 6 TET1 is required for the PAX4 enhancer to achieve a hypomethylated state.**

**a** Genome-browser view of the *PAX4* locus with increased methylation and decreased chromatin associability upon TET depletion at a TET1/FOXA2 co-bound region featuring enhancer signatures H3K4me1 and H3K27ac. **b** Locus-specific increase in 5mC at the *PAX4* enhancer in TKO or TET1KO samples compared with TET2/3DKO samples. Percentages of unmethylated cytosine and 5mC at CCGG sites are shown. n = 3 independent differentiation.