H3K4 Methylation in β-cells Prevents Transcriptional Downregulation and Variance Associated with Type 2 Diabetes

Ben Vanderkruk[1,2,3], Nina Maeshima[2], Daniel J Pasula[1,2,3], Meilin An[2], Priya Suresh[2], Aline Daniel[2], Cassandra L. McDonald[2], Dan S. Luciani[1,2,3] and Brad G. Hoffman[1,2,3*]

[1]Department of Surgery, University of British Columbia, Vancouver, BC, V5Z 4E3, Canada

[2]Diabetes Research Group, British Columbia Children's Hospital Research Institute, 950 West 28th Avenue, Vancouver, BC, V5Z 4H4, Canada

[3]Department of Cellular and Physiological Sciences, University of British Columbia, Vancouver, BC, Canada

[*]Corresponding Author: brad.hoffman@ubc.ca

Summary

Histone 3 lysine 4 trimethylation (H3K4me3) is associated with promoters of actively expressed genes, with genes important for cell identity frequently having exceptionally broad H3K4me3-enriched domains at their TSS. While H3K4 methylation is implicated in contributing to transcription, maintaining transcriptional stability, facilitating enhancer-promoter interactions, and preventing irreversible silencing, some studies suggest it has little functional impact. Therefore, the function of H3K4 methylation is not resolved. Insufficient insulin release by β-cells is the primary etiology in type 2 diabetes (T2D) and is associated with the loss of expression of genes essential to normal β-cell function. We find that H3K4me3 is reduced in islets from mouse models of diabetes and from human donors with T2D. Using a genetic mouse model to impair H3K4 methyltransferase activity of TrxG complexes, we find that reduction of H3K4 methylation significantly reduces insulin production and glucose-responsiveness and increases transcriptional entropy, indicative of a loss of β-cell maturity. Genes that are downregulated by reduction to H3K4 methylation are concordantly downregulated in T2D. Loss of H3K4 methylation causes global dilution of epigenetic complexity but does not generally reduce gene expression – instead, genes related to β-cell function and/or in particular chromatin environments are specifically affected. While neither H3K4me3 nor H3K4me1 are strictly required for the expression of many genes, the expression of genes with critical roles in β-cell function becomes destabilized, with increased variance and decreased overall expression. Our data further suggests that, in absence of H3K4me3, promoter-associated H3K4me1 is sufficient to maintain expression. Together, these data implicate H3K4 methylation dysregulation as destabilizing β-cell gene expression and contributing to β-cell dysfunction in T2D.

Introduction

Methylation of lysine 4 of histone H3 (H3K4me) is conserved across eukarya and, along with other chromatin post-translational modifications, correlates with cell-type-specific gene expression programmes[1,2]. Distribution of H3K4me across the genome is determined not only by cell type, but is also influenced by metabolic states and by pathologies such as cancer, aging, and obesity[3–6]. H3K4 can be mono-, di-, or trimethylated, with distinct functional associations in each case. H3K4 trimethylation (H3K4me3) is abundant in close proximity with transcription start sites (TSS) of actively expressed genes[7,8]. Exceptionally broad H3K4me3-enriched domains are linked to genes with high and consistent expression and whose functions are important for cell identity[5,9]. H3K4 monomethylation (H3K4me1) is associated with active and poised enhancers and also demarcates the borders of active promoters[7,10,11]. Other prominent histone modifications are H3K27 acetylation (H3K27ac), which is enriched at active promoters and enhancers, and H3K27me3, which is correlated with Polycomb repressed DNA and anticorrelated with H3K27ac, H3K4me1, and H3K4me3[7].

Although these associations are widely recognized, findings regarding the functional impact of H3K4 methylation on transcription have been divided. H3K4 methylation may regulate gene activation via recruitment of transcription factors and cofactors[12–15], facilitating enhancer-promoter interactions[16], and/or by preventing irreversible silencing[17]. On the other hand, in genetic models global (i.e., genome-wide) decreases in H3K4me1 or H3K4me3 levels are remarkably well-tolerated in embryonic stem cells and during development in genetic models[18–20], perhaps due to redundancy with other regulatory inputs and/or environmental-context-dependency[18,21]. Therefore, the importance of H3K4 methylation as a driver of transcriptional activity versus as merely a marker of promoters remains poorly understood.

Methylation of H3K4 in mammalian cells is predominantly catalyzed by the Trithorax Group (TrxG) complexes (also known as COMPASS and COMPASS-related complexes). TrxG complexes are composed of the core subunits WDR5, RBBP5, ASH2L, and DPY30, plus one of six catalytic subunits MLL1-4 or SET1A-B, in addition to a variable retinue of accessory factors[22]. We previously defined roles for the TrxG core subunit DPY30 during pancreatic islet cell specification and maturation[23,24], finding that DPY30 fine-tunes cell fate-decisions to increase islet cell specification[23] and is needed to activate terminal marker genes during islet cell maturation[24].

Pancreatic islets are endocrine micro-organs that contain the body's β-cells that dynamically maintain normoglycemia during myriad metabolic conditions by releasing appropriate quantities of insulin. Insufficient insulin release by β-cells is the primary etiology in type 2 diabetes (T2D) and is characterized by metabolic and transcriptional remodeling in β-cells and other tissues[25–27]. Individual risk and prognosis of T2D is influenced by genetics and environmental factors including age, exercise, and nutrition[25]. Chromatin modifications act as a potential link between environmental factors and gene expression changes in β-cells[28,29]. Accordingly, H3K27ac in islet chromatin regulatory domains is correlated with acute glucose-induced gene expression changes[30,31], and demethylation of DNA and H3K27 are linked to gene activation in T2D[26,32]. Here, we show that H3K4me3 enrichment in mature β-cell chromatin corresponds with transcription of critical β-cell genes, and that it is reduced in T2D. Using a genetic model to deplete H3K4 methylation from mature β-cells, we observed downregulation of genes that are concordantly downregulated in T2D. Depletion of H3K4 methylation in β-cells is deleterious to several aspects of insulin production and glucose-responsiveness and increases transcriptional entropy, indicative of a loss of β-cell maturity. Depletion of H3K4me3 causes global dilution of epigenetic complexity but does not generally reduce gene expression – instead, genes related

to β-cell function and/or in particular chromatin environments are affected. Thus, maintenance of the unique β-cell transcriptome by H3K4 methylation is impaired in T2D.

Results

H3K4me3 stratifies β-cell genes and is dysregulated in type 2 diabetes. To begin to elucidate the relationship between H3K4me3 and gene expression in β-cells, we generated genome-wide enrichment profiles of H3K4me3, H3K4me1, H3K27ac, and H3K27me3 in healthy mature murine β-cells (sorted on the basis of Pdx1-CreER-driven *Rosa26*^{mTmG} recombination) using native chromatin immunoprecipitation sequencing (ChIP-seq). H3K4me3 was enriched in 18,936 distinct peak regions, the majority of which spanned DNA segments of less than 3 kb. Ranking H3K4me3 peaks by breadth revealed a class of exceptionally broad peaks, which included H3K4me3 peaks covering gene promoters of hallmark β-cell transcription factors (Fig. 1a), consistent with the idea that genes important for cell identity are marked with broad H3K4me3 peaks[5,9]. We detected strong enrichment for H3K4me1 and H3K27ac, and depletion of H3K27me3, in the vicinity of TSS's defined by broad H3K4me3 peaks (Extended Data Fig 1c). RNA-sequencing (RNA-seq) of mature murine β-cells (sorted as above) indicated that gene expression was mildly correlated with promoter H3K4me3 peak breadth; however, genes with broad peaks did not have exceptionally high RNA expression (Extended Data Fig. 1b). As an example, Fig. 1b shows the H3K4me3 profiles of a β-cell transcription factor gene, *Nkx6-1*, compared to an expression-matched housekeeping gene, *Rplp0*, which has a typical H3K4me3 profile. These data show that genes critical to the β-cell lineage are marked by broad H3K4me3 enrichment in β-cell chromatin, and are consistent with previous reports that broad H3K4me3 peaks coincide with active promoter markers[5], but are only modestly correlated with expression levels[9].

We next examined the relationship between H3K4me3 enrichment and transcriptional changes that occur in a mouse model of T2D. Leptin receptor-deficient *Lepr*$^{db/db}$ mice exhibit overfeeding, obesity, fasting hyperglycemia, and glucose intolerance, along with β-cell dysfunction and dedifferentiation[33]. We compared the chromatin profiles for genes upregulated, downregulated (≥2-fold, $P \leq 0.01$), or stably expressed (≤1.1-fold changed) in *Lepr*$^{db/db}$ versus *Lepr*$^{+/+}$ islets[34]. We observed an exaggerated shoulder of H3K4me3 and H3K27ac enrichment extending into the gene body of genes that are downregulated in *Lepr*$^{db/db}$ islets (Fig. 1c). In contrast, genes upregulated in *Lepr*$^{db/db}$ islets displayed weak enrichment of these marks (Fig. 1c). To conduct a quantitative analysis, we ranked all promoter-associated H3K4me3 peaks by breadth and grouped them into quantiles of H3K4me3 peaks with similar breadth. Enrichment analysis shows that a large number of housekeeping genes (GSEA:M11197, [35]), genes involved in endocrine pancreas development (GSEA:M12875), and genes related to maturity onset diabetes of the young (GSEA:M18312) have broad peaks (Fig. 1d). Further, genes downregulated in *Lepr*$^{db/db}$ mice are strongly enriched for broad H3K4me3 peaks, while upregulated genes are strongly enriched among narrow H3K4me3 peaks (Fig 1d). As H3K4me3 peak breadth is positively correlated with absolute expression (Extended Data Fig. 1b), we tested whether this observation was secondary to highly expressed genes becoming downregulated, and lowly expressed genes becoming upregulated. Ranking genes on the basis of absolute expression indicates that gene sets related to islet development and diabetes are more effectively stratified by H3K4me3 peak breadth than absolute expression (an exception is housekeeping genes, which are more strongly enriched among the top 5% of genes ranked by absolute expression than by H3K4me3 peak breadth, as expected[5]) (Extended Data Fig. 1d). This data

122 shows that genes that are differentially expressed in *Lepr*$^{db/db}$ islets are diametrically stratified according
123 to H3K4me3 breadth, with upregulated genes tending to have narrow peaks and downregulated genes
124 tending to have broad H3K4me3 peaks.

125 To extend our observations to human islets, we retrieved H3K4me3 ChIP-seq data of sorted β-
126 cells[36] and RNA-seq data of islets of nondiabetic ("Ctrl", HbA1c < 6%, n = 51) and diabetic ("T2D", HbA1c
127 > 6.5, n = 11) donors[27]. As in murine β-cells, the H3K4me3 peak breadth distribution shows a small
128 population of exceptionally broad peaks (Extended Data Fig. 1a) and a generally weak relationship
129 between peak breadth and gene expression (Extended Data Fig. 1b). Human β-cell transcription factors
130 also tend to have broad H3K4me3 peaks (Extended Data Fig. 1a). We again observed significant
131 enrichment for genes downregulated ($P \leq 0.01$) in T2D within the broadest H3K4me3 peaks, whereas
132 enrichment of upregulated genes was statistically weak in any quantile (Fig. 1d, Extended Data Fig. 1d).
133 Thus, regions of broad H3K4me3 enrichment are overrepresented among β-cell identity genes and
134 genes downregulated in human T2D.

135 In available single cell RNA-seq data (n = 18 Ctrl, 7 T2D donors)[37] we also observed elevated
136 variability of gene expression in β-cells from donors with T2D (Fig. 1e). Correlating H3K4me3 peak
137 breadth data with expression variance data shows that, up to a point, increasing H3K4me3 breadth
138 correlates with modest protection against increased variance in T2D, after which genes with the
139 broadest H3K4me3 peaks instead have increased variance (Fig. 1e). Given this, we next wondered
140 whether H3K4me3 enrichment in islets is perturbed in diabetes. Fluorescent immunohistochemical (IHC)
141 analysis showed reduction of H3K4me3 immunoreactivity (normalized to total histone H3) in islets of
142 *Lepr*$^{db/db}$ versus *Lepr*$^{+/+}$ mice, and human T2D versus non-diabetic donors ($P = 0.0528$ for the latter) (Fig.
143 1f-g, i-j) Immunoblot analysis of a random subset of the human islet samples corroborated the reduction
144 of H3K4me3 in T2D, whereas no difference was seen for H3K4me1 (Fig. 1h, k, Extended Data Fig. 1e).
145 Together, these data show that genes with broad H3K4me3 enrichment are prone to transcriptional
146 dysregulation in T2D, and that H3K4me3 is globally reduced in islets from T2D donors.

147 **Reduction to H3K4 methylation in β-cells causes glucose intolerance and hyperglycemia.** These
148 observations motivated us to test whether reducing H3K4 methylation in mature β-cells could cause the
149 cell dysfunction characteristic of T2D. The most prominent mammalian H3K4 methyltransferases are
150 associated with TrxG complexes and require DPY30 for enzymatic activity *in vivo*[23,38–40]. We therefore
151 generated *Dpy30*-deficient β-cells by crossing *Dpy30*$^{flox/flox}$ mice with *Pdx1-CreER; Rosa26*^{mTmG} mice and
152 administering tamoxifen at 8-weeks-old to induce recombination in mature mice (Fig 2a, Extended Data
153 Fig. 2a) (henceforth, KO: *Pdx1-CreER*$^{Tg/0}$, *Dpy30*$^{flox/flox}$; *Rosa26*$^{mTmG/+}$; WT: *Pdx1-CreER*$^{Tg/0}$, *Dpy30*$^{+/+}$,
154 *Rosa26*$^{mTmG/+}$ or *Pdx1-CreER*$^{0/0}$, *Dpy30*$^{flox/flox}$, *Rosa26*$^{mTmG/+}$). Loss of *Dpy30* did not affect assembly of the
155 remaining core components of the TrxG complex, or their association with chromatin, but did cause
156 progressive reduction of H3K4me3 and H3K4me1 (Fig. 2b-e). To characterize the relative role of DPY30
157 versus H3K4 methylation in regulating gene expression, we generated mRNA-seq data from sorted GFP+
158 cells (Extended Data Fig. 2a) at time points defined by loss of DPY30 with no change in H3K4 methylation
159 (15-days after tamoxifen administration) versus combined loss of DPY30 and reduction of H3K4
160 methylation (45-days) (Fig. 2d). While expression of only 3 genes was significantly altered (≥ 2-fold, $P \leq$
161 0.01) by loss of DPY30 alone, one of which was *Dpy30* itself, 828 genes were differentially expressed
162 after loss of H3K4 methylation (Fig. 2f). Thus, transcriptomic remodeling in *Dpy30*-KO β-cells is
163 secondary to loss of H3K4 methylation.

164 Physiologically, KO mice display glucose intolerance and impaired insulin secretion during
165 glucose tolerance tests (GTT) 45-days after *Dpy30* inactivation, whereas no significant differences were
166 seen after 15- or 30-days compared to WT mice (Fig. 2g). Functional annotation of differentially
167 expressed genes indicate reduction to exocytosis and glucose homeostasis processes in KO β-cells 45-
168 days after *Dpy30* inactivation (Extended Data Fig. 2b), supporting a defect in insulin secretion. Glucose
169 intolerance quickly progressed to dramatic unfasted hyperglycemia beginning 44-50 days after *Dpy30*
170 inactivation (Fig. 2h, top). Hyperglycemia in KO mice did not coincide with changes to animal body mass
171 (Fig. 2h, bottom), and occurred despite elevated β-cell proliferation (Extended Data Fig. 2c-d). We
172 further confirmed our physiological observations using an alternative Cre driver. Deletion of *Dpy30* from
173 maturing β-cells using *Ins1-Cre* caused reduction of H3K4me3 and H3K4me1 by 5 weeks of age
174 (Extended Data Fig. 3a-b). Unfasted blood glucose was dramatically elevated from 5-6 weeks of age,
175 whereas body mass was similar among KO, WT, and heterozygous *Ins1-Cre Dpy30*-KO mice (Extended
176 Data Fig. 3c-d). *Ins1-Cre Dpy30*-KO mice showed impaired glucose clearance and low serum insulin
177 during GTT (Extended Data Fig. 3f-g), corroborating β-cell-intrinsic dysfunction after reduction to H3K4
178 methylation. In summary, reduction of H3K4 methylation in mature β-cells causes highly penetrant
179 diabetes characterized by reduced glucose tolerance, reduced serum insulin, and hyperglycemia.

180 **H3K4 methylation is required for β-cell glucose-responsiveness.** We next explored how reduction of
181 H3K4 methylation alters β-cell function to cause glucose intolerance. First, we carried out calibrated
182 ChIP-seq for H3K4me3, H3K4me1, H3K27ac, and H3K27me3 in sorted WT and KO β-cells 45-days after
183 *Dpy30* inactivation. Interrogation of the *Ins1* and *Ins2* loci revealed broad H3K4me3 enrichment
184 extending several kb past the extent of each gene in WT chromatin. Strikingly, KO chromatin shows
185 complete loss of H3K4me3 and profound reduction of H3K4me1 and H3K27ac at these loci (Fig. 3a).
186 Despite the combined reduction of active promoter marks, mRNA-seq data revealed a very modest
187 reduction of *Ins1* and *Ins2* transcripts (-1.152 and -1.234-fold, respectively, Fig. 3b), indicating that
188 expression of these genes can occur in the absence of local H3K4me3. Despite the modest reductions in
189 expression of *Ins1* and *Ins2*, we found a significant reduction in overall insulin peptide content in KO
190 islets (Fig. 3c). Reduction of insulin was confirmed in individual β-cells, where reduction of insulin
191 granule size and density was apparent in transmission electron micrographs (TEM) (Fig. 3d-e), suggesting
192 a defect in insulin translation, storage, and/or maturation. Intriguingly, 2 genes downregulated in KO
193 cells, *Nnat* and *Spcs3* (Extended Data Fig. 4a), are components of the signal peptidase complex which
194 converts preproinsulin into proinsulin in β-cells[41]. Both genes are also downregulated in $Lepr^{db/db}$ islets
195 (data not shown), and *Nnat* is repressed in human obesity[42]. To determine if production of proinsulin is
196 impaired in KO islets, we treated islets with the proteasome inhibitor MG132 to prevent (pre)proinsulin
197 degradation. Immunoblot analysis of islet lysates showed an increase in the preproinsulin/proinsulin
198 ratio, consistent with impaired signal peptidase complex processing of preproinsulin (Fig. 3f-g).
199 Therefore, downregulation of insulin processing genes, caused by reduction of H3K4 methylation in β-
200 cells, impairs insulin peptide processing and reduces cellular insulin content.

201 We confirmed that glucose-stimulated insulin secretion is deficient in *ex vivo* islets (Fig. 3h), in
202 agreement with *in vivo* GTTs (Fig. 2g, Extended Data Fig. 3g). Notably, accounting for the reduced insulin
203 content of KO islets did not fully explain the impairment of insulin secretion. The persistent impairment
204 was specific to high glucose stimulation, whereas secretion under basal conditions, stimulated with the
205 Krebs cycle metabolite analogue dimethyl α-ketoglutarate, or depolarizing cells directly using KCl,
206 showed no difference between WT and KO insulin secretion (Fig. 3h), suggesting impaired coupling of

207 glucose sensing to insulin secretion. To study stimulus-secretion coupling we measured cytosolic Ca^{2+}
208 during islet stimulation. High glucose-stimulated Ca^{2+} influx showed a delayed response and lower
209 magnitude in KO islets, whereas KCl stimulation caused normal Ca^{2+} influx (Fig. 3i), consistent with a
210 defect in metabolic coupling involving glucose catabolism and/or Ca^{2+} signaling upstream of membrane
211 depolarization. Accordingly, GSEA revealed enrichment of calcium-mediated signaling genes among
212 downregulated genes (Extended Data Fig. 4b). Respirometry of dispersed islet suggests that KO cells do
213 not display differences in glucose catabolism (Fig 3j-k). Interestingly, KO cells did display elevated
214 maximal respiration (Fig. 3j-k), as well as upregulation of oxidative phosphorylation genes and expansion
215 of mitochondria area (Extended Data Fig. 4b-c), which could indicate compensation for higher metabolic
216 demand and/or lower metabolic efficiency. In a separate measure of glucose-stimulation, we tested
217 replication in WT and KO cells by measuring EdU incorporation of islet cells cultured in low versus high
218 glucose for 2 days. While high glucose culture induced a 22-fold increase of replication in WT β-cells, KO
219 β-cells were induced less than 2-fold (Extended Data Fig. 4d-e). Taken together, these results suggest
220 that reduction to H3K4 methylation impairs calcium signaling-mediated coupling of elevated glucose to
221 insulin secretion and cell replication.

222 **H3K4 methylation maintains expression of genes that are downregulated in type 2 diabetes.** We next
223 asked whether loss of H3K4 methylation recapitulates the transcriptional changes that occur in diabetic
224 islets. Genes that are downregulated in mouse *Lepr*$^{db/db}$ or human T2D islets showed marked
225 downregulation in KO β-cells (Fig. 4a), supporting a common mechanism for gene downregulation.
226 Examination of β-cell transcription factors exemplifies the prevalent, modest downregulation of genes
227 with broad H3Kme3 enrichment in diabetic islets and KO β-cells (Fig. 4b). Indeed, genes marked by
228 broad H3K4me3 peaks were similarly biased toward downregulation in mouse *Lepr*$^{db/db}$ islets, human
229 T2D islets, and KO β-cells (Fig. 4c). An interesting divergence is seen for genes marked by narrow
230 H3K4me3 peaks: these genes are biased toward upregulation in mouse *Lepr*$^{db/db}$ and human T2D islets,
231 but downregulation in KO β-cells (Fig 4c), indicating that other, opposing epigenetic factors may drive
232 gene upregulation in T2D, such as deficient Polycomb silencing[32]. No enrichment was found among
233 genes upregulated in mouse *Lepr*$^{db/db}$, or human T2D, with genes upregulated in KO β-cells (data not
234 shown), further indicating that other mechanisms drive gene upregulation in T2D. In contrast,
235 intersection of genes that are downregulated in both KO β-cells and *Lepr*$^{db/db}$ islets captures genes that
236 are highly relevant to β-cell function, with enrichment in gene ontology annotations for insulin
237 secretion, response to glucose, cell proliferation, and cation transport (Fig 4d) – further underscoring
238 that transcriptional dysregulation effects functional impairment (e.g. those explored above) in KO and
239 diabetic β-cells. Thus, H3K4 methylation maintains expression of genes that are downregulated in T2D.

240 **Reduction to H3K4 methylation increases gene expression entropy in β-cells.** To gain further insight
241 into the transcription dysregulation seen in KO cells we performed single cell RNA-seq in mouse islet
242 cells 45-days after tamoxifen administration, generating 2,510 WT and 2,597 KO single-cell
243 transcriptomes that passed quality filtering and that we identified as β-cells (see Methods and Extended
244 Data Fig. 5a). Examination of genes highly downregulated in our bulk RNA-seq data such as *Glp1r*,
245 *Slc2a2*, and *G6pc2* confirmed general agreement between the two datasets (Extended Data Fig. 5b).
246 Notably, while mean expression of *Ins1*, *Ins2* and *Iapp* were only moderately decreased between WT
247 and KO β-cells, cell-to-cell expression variance of these genes was expanded in KO islets (Extended Data
248 Fig. 5b). To explore further, we compared un-recombined *Tdtomato+* β-cells to *Egfp+* β-cells from KO
249 islets (henceforth mT and mG, respectively) as these cells allow us to directly compare *Dpy30*-expressing

250 and KO β-cells from the same organ (Fig. 5a). To model the divergence of mG from mT β-cells, we
251 performed pseudotime ordering using the mT cluster as the ground state (Fig. 5b). *Ins1* and *Ins2*
252 expression become notably more stochastic over pseudotime (Fig 5c). Calculating cell transcriptomic
253 entropy scores[43], revealed an overall increase in transcriptional entropy along pseudotime (Fig. 5d) and
254 in mG cells versus mT cells (Fig. 5e). As we had found in T2D islets, transcriptional variance was elevated
255 in mG cells (Fig. 5f) in all quantiles of H3K4me3 breadth (Fig 5g), suggesting that H3K4 methylation per
256 se, rather than strictly broad domains of H3K4me3 enrichment, reduces gene expression variance.
257 Increased expression entropy is associated with progenitor cell states[43]; however, we did not detect
258 reactivation of marker genes of immature β-cells or of endocrine progenitor cell types (Extended Data
259 Fig. 5c), indicating that *Dpy30*-deficient cells are not dedifferentiating, but are rather becoming de-
260 specialized, with decreased transcriptional stability leading to discordance in expression of genes
261 essential to β-cell function (Extended Data Fig. 5d). In summary, transcriptomic entropy and
262 transcriptional variance are increased by reduction of H3K4 methylation in mature β-cells.

263 **Dilution of epigenetic complexity in Dpy30-KO β-cells.** Reasoning that functional and transcriptomic
264 changes in *Dpy30*-KO β-cells arise from epigenetic remodeling, we next compared the chromatin
265 landscapes of WT and KO cells. H3K4me3 and H3K4me1 are strongly depleted in TSS regions in KO
266 chromatin (Fig. 6a). At active promoters in the KO residual H3K4me1 is enriched more closely to the TSS
267 than in WT chromatin (Fig. 6a), occupying areas previously marked with H3K4me3 in WT (Extended Data
268 Fig. 6a). Analysis of H3K4me1 and H3K4me3 enrichment in 10-kb windows across the genome revealed
269 that both marks are globally depleted in KO cells (Extended Data Fig. 6b). H3K27ac enrichment is also
270 markedly reduced in KO cells (Fig. 6a), even in TSS regions of stably-expressed and upregulated genes
271 (Fig. 6b-d). While the Polycomb-repressive mark H3K27me3 is modestly increased in the TSS regions of
272 genes that become downregulated in the KO (Fig. 6b, right), we found no evidence for invasion of
273 H3K27me3-enriched domains into formerly actively marked areas (Fig. 6c,e, Extended Data Fig. 6a).
274 Taken together, we find that H3K4 methylation is needed to maintain genome-wide H3K27ac, and that
275 *Dpy30*-KO chromatin is characterized by depletion of active chromatin modifications.

276 The remaining H3K4me3, H3K4me1, and H3K27ac in KO chromatin appears to become more
277 restricted to formerly H3K4me3-enriched regions, as noted above, and to occupy stretches of chromatin
278 that more closely overlap each other in KO compared to in WT chromatin (Fig 6c, Extended Data Fig.
279 6a,c). In support, Spearman's rank correlations between distinct histone modifications are higher in KO
280 chromatin than WT chromatin (Extended Data Fig. 6d-f), confirming that their genomic distributions are
281 more similar in KO cells. This higher concordance between different histone marks suggests that the
282 combinatorial complexity of the chromatin landscape is decreased in KO cells. Despite this,
283 transcriptomes of WT and KO β-cells are well correlated (Fig. 6f), with no reduction in the total RNA
284 content of KO compared to WT cells (Fig 6f, note that the regression line tracks with the diagonal).
285 Nevertheless, gene expression and histone modification distribution were still related in the KO.

286 Specifically, TSS regions that are strongly enriched for H3K4me3 in WT chromatin actually gain
287 H3K4me1 in KO chromatin (Fig. 6g), and the expression of such genes, and of genes that maintained
288 relatively high TSS-associated H3K4me1 levels, are largely unchanged. Meanwhile, genes with the
289 greatest reduction in H3K4me1, regardless of their level of H3K27ac reduction, are frequently
290 downregulated (Fig. 6h). Despite this, there was also still some correlation between expression and
291 H3K27ac enrichment. For example, while 352 genes with a ≥2-fold decrease of H3K27ac intensity in the
292 TSS were downregulated ≥2-fold in KO cells, only 35 were upregulated (Extended Data Fig. 6i).

293 Furthermore, H3K27ac peaks gaining intensity in the KO were near upregulated genes, and H3K27ac
294 peaks losing intensity were near downregulated genes (≥2-fold, $P \leq 0.01$), more frequently than
295 expected by chance (Extended Data Fig. 6g-h). This data indicates that, although H3K27ac is globally
296 reduced, H3K27ac is still positively associated with gene regulation in KO chromatin. Together, these
297 data suggest that H3K4me3 contributes to active gene expression, but that, in its absence, promoter-
298 associated H3K4me1 is sufficient to at least partially compensate for its loss.

299

300 **Discussion**

301 H3K4 methylation has been reported to aid in the recruitment of transcription factors and cofactors[12–15],
302 to facilitate enhancer-promoter interactions[16], and to prevent irreversible silencing[17]. Despite these
303 findings, mutation of H3 lysine 4 to non-methylatable residues seems to have minimal effects on gene
304 expression[20,44,45]. Other models in which H3Kme3 is globally reduced also suggest that H3K4me3 has a
305 limited role in transcription[18–20], perhaps due to redundancy with other regulatory inputs and/or
306 environmental-context-dependency[18,21]. Further, many of these studies impair the co-activator activity
307 of the TrxG complex, making it difficult to determine whether any consequences that are seen are due
308 to loss of H3K4 methylation or to the loss of recruitment of coactivators[18,19,46].

309 Here, we uncover specific features of H3K4 methylation associated with T2D in β-cells and
310 characterize metabolic, transcriptomic, and epigenomic consequences using an inducible genetic mouse
311 model. We find that DPY30-dependent TrxG activity is the dominant source of H3K4me3 and H3K4me1
312 in mature β-cells at essentially all genomic loci, although input from other methyltransferases such as
313 SETD7[47] may account for some residual methylation. Genetic inactivation of *Dpy30* causes depletion of
314 H3K4me1 and H3K4me3 during the course of several weeks. Apparent physiological and transcriptomic
315 changes are associated with loss of H3K4 methylation rather than DPY30 itself, highlighting *Dpy30*-KO as
316 a convenient and powerful tool to study the functional impact of simultaneous loss of H3K4me3 and
317 H3K4me1, without targeting non-enzymatic functions of TrxG complexes[18,19].

318 In β-cells, reduction to H3K4 methylation causes impairments to insulin production and glucose
319 stimulation, and transcriptomic remodeling including downregulation of genes that are also
320 downregulated in mouse and human diabetes. Further, *Dpy30*-KO and T2D β-cells display an increasingly
321 stochastic gene expression profile, a feature linked to loss of H3K4me3 breadth[9]. Increased gene
322 expression variability between cells within a tissue may increase adaptability to stress, but is also
323 associated with aging and several human pathologies[48–50]. It is interesting to consider stochastic
324 expression in light of transcriptional bursting kinetics, since promoters and enhancers uniquely shape
325 burst frequencies[51,52]. If H3K4me1 and/or H3K4me3 are functionally involved in enhancer-promoter
326 looping, as has been carefully demonstrated in the case of H3K4me1[16], increasingly stochastic
327 expression may be a manifestation of altered bursting frequencies. We suggest that in *Dpy30*-KO β-cells,
328 and possibly in T2D, this increased stochasticity of expression of genes essential for β-cell function
329 decreases the concordance of the co-expression of these factors, and this loss contributes to β-cell
330 dysfunction.

331 Importantly, we employed spike-in calibration for ChIP-seq and RNA-seq experiments[53,54].
332 Empirical normalization of DNA and RNA enrichment revealed a global reduction of H3K4me3,
333 H3K4me1, and H3K27ac enrichment after *Dpy30* inactivation—but, in stark contrast, no change to the

total mRNA content of the cells. These epigenetic marks may, therefore, be more important for promoting positional allocation of the transcriptional machinery, rather than their absolute activity[55]. In support, the expression of many genes lacking H3K4me3, H3K4me1, and H3K27ac at their promoters in *Dpy30*-KO cells is only modestly affected, or even completely unaffected. The genome-wide decrease in H3K27ac in *Dpy30*-KO chromatin adds to existing evidence[19,56] that H3K27ac is downstream of H3K4 methylation. H3K27ac is also, irrespective of H3K4 methylation, positively correlated with gene expression changes between *Dpy30*-WT and KO cells, consistent with recent work suggesting that H3K27ac is largely a consequence of transcription[57]. Together, these data suggest that neither H3K4me3 nor H3K4me1 are absolutely required for the expression of most genes.

Combined loss of promoter-associated H3K4me3, H3K4me1, and H3K27ac is a common feature shared by the most downregulated genes in *Dpy30*-KO cells. Our data suggest that H3K4me3 positively contributes to gene expression, but that, in its absence, promoter-associated H3K4me1 is sufficient to maintain expression. For example, the *Ins1* and *Ins2* genes remain the two most highly expressed genes in *Dpy30*-KO cells despite suffering complete loss of their H3K4me3 peaks. In fact, many genes with strong H3K4me3 density retain, or even accumulate, H3K4me1 proximal to the TSS upon loss of H3K4me3, to a degree that is related to the original enrichment level of HK4me3. An interpretation is that, in *Dpy30*-KO chromatin, depletion of H3K4 methylation progresses through a H3K4me1 intermediate stage, and that loci highly enriched for H3K4me3 thus maintain enrichment for this mark for a longer period of time. An alternative, and not mutually exclusive, interpretation, is that new H3K4 methylation in these regions is less reliant on DPY30[38].

Reduction of H3K4 methylation did not prompt canonical dedifferentiation[33] of β-cells to a precursor-like state, however, the cells displayed some features of immaturity like elevated transcriptional entropy[43], *in vivo* replication[58], reduced expression of mature β-cell transcription factors, and reduced glucose-responsiveness[59]. Dilution of epigenetic complexity characterizing *Dpy30*-KO cells is thematically similar to so-called “smoothening” of the epigenomic landscape which is proposed to cause loss of cellular identity and increase entropy during epigenetic aging[60]. In that model, both active and repressive chromatin modifications show reduced intensity at regions where they are most enriched. A similar picture emerges for β-cells during T2D: general reduction of H3K4me3 occurs simultaneously with reduction of H3K27me3 or polycomb suppression[32]. We believe that reduction to H3K4me3 reduces transcriptional consistency and impairs expression of cell lineage genes marked by broad domains of H3K4me3 enrichment, while reduction to polycomb suppression allows ectopic activation of disallowed genes and impairs expression of cell lineage genes[32]. Taken together, β-cell dysfunction during T2D may, in part, be framed as an epigenetic aging problem, in which a relative loss of H3K4me3-dependent gene expression regulation is an important component.

Acknowledgements

The authors thank the staff of the Animal Care Facility at British Columbia Children’s Hospital Research Institute for daily maintenance of the mouse colonies, and Francis Lynn for providing Pdx1-CreER animals and William Gibson for providing Ins1-Cre animals. We thank Marcia Reid at the Facility for Electron Microscopy Research of McGill University for help with transmission electron microscopy, Lisa Xu at the Flow Core of BCCHRI for help with FACS, and the staff of the Alberta Diabetes Institute IsletCore for providing human islets. This work was supported by operating grants from the British Columbia Children’s

Hospital Research Institute, the Natural Sciences and Engineering Research Council of Canada (RGPIN-2016-04292) and the Canadian Institute for Health Research (RN310864 – 375894). Fellowship support for B.V. was provided by the British Columbia Children's Hospital Research Institute, University of British Columbia, and Natural Sciences and Engineering Research Council of Canada.

Author Contributions

Experiments were conducted by B.V., N.M., P.S., A.D., C.L.M, and conceptualized by B.V., and B.G.H. D.J.P and D.S.L performed calcium imaging experiments. All bioinformatics were performed by B.V. and B.G.H. Funding and supervision was provided by B.G.H.; the manuscript was written by B.V. and B.G.H. B.G.H is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Competing Interest Statement

The authors declare no competing interests.

391 **Methods**

392 **Animal strains and care.** Mice harboring conditional alleles of *Dpy30* (ref. [23]) and *Rosa26*$^{mTmG/+}$ (Jax
393 007576) were crossed with *Pdx1-CreER*$^{Tg/0}$ (Jax 024968) or *Ins1*$^{Cre/+}$ (Jax 026801) lines to generate β-cell-
394 specific *Dpy30* KO animals. At 8 weeks old, *Pdx1CreER*$^{Tg/0}$ mice were administered 8 mg tamoxifen
395 dissolved in 100 µL corn oil by oral gavage three times, with ~48 hours between administrations. The
396 first day of injection is considered day 0. Body mass and blood glucose concentration from the tail vein
397 was monitored 1-2 times per week from 5-weeks (*InsCre*) or 8-weeks (*PdxCreER*) of age. *Lepr*$^{db/db}$ mice
398 and *Dock7*$^{m/m}$ controls (Jax 000642) were used at 12-16-weeks-old. Up to five littermates were housed
399 per cage with free access to food (Teklad 2918) and water with temperature and humidity control under
400 a 12-hour light/dark cycle. Mouse experiments were restricted to male mice and were approved by the
401 University of British Columbia Animal Care Committee under certificates A17-0045 and A18-0111.

402 **Cell care.** Mouse islet cells were maintained in RPMI 1640 medium containing 11 mM glucose, 10% (v/v)
403 FBS, 50 U/ml penicillin, 50µg/mL streptomycin (complete RPMI) in 5% CO_2 and 37°C. *Drosophila* S2 cells
404 were maintained at ambient temperature and air in Schneider's Drosophila Medium (Gibco) containing
405 10% (v/v) heat-inactivated FBS, 50 U/ml penicillin, 50µg/mL streptomycin, with passaging once per
406 week.

407 **Islet and β-cell isolation.** Unless otherwise indicated, mouse islets were isolated 45-days after tamoxifen
408 administration and processed for analysis immediately after isolation. Pancreata were perfused with 700
409 U/mL collagenase type XI (Sigma-Aldrich) through the bile duct, excised, and held for 15-minutes at 40°C
410 before manual disruption and hand-picking of islets. Human islets were procured from the Alberta
411 Diabetes Institute Islet Core or Prodo laboratories. Upon receipt, human islets were handpicked, washed
412 once with PBS, and processed immediately for immunohistochemistry or immunoblotting as described
413 below. To isolate mouse β-cells, islets from *Pdx1-CreER*$^{Tg/0}$ *Rosa26*$^{mTmG/-}$ transgenic mice were washed 3×
414 in room temperature PBS + 2 mM EDTA and then dispersed by constant gentle pipetting in 1 mL of
415 0.025% Trypsin-EDTA (Thermo Fisher Scientific) in PBS at room temperature for 8-10 minutes. Trypsin
416 was quenched by addition of 9 mL room temperature 10% (v/v) FBS in RPMI 1640, cells were
417 centrifuged for 3 minutes at 600 × *g* at 4°C, and resuspended in 0.3-0.5 mL ice-cold sorting buffer (PBS, 2
418 mM EDTA, 1% (w/v) BSA). After passing through a 70 µm cell strainer, singlet GFP^+ RFP^- cells were
419 purified by fluorescence-activated cell sorting (FACS) using a FACS-Aria II with a 100 µm nozzle.

420 **ChIP-seq.** An ultra-low-input native ChIP protocol[61] was used to generate ChIP-seq libraries in biological
421 duplicate, with modifications. Per each IP, 100,000 GFP^+ RFP^- β-cells were pooled with 50,000 *Drosophila*
422 S2 cells by FACS, centrifuged at 600 × *g* for 5 minutes at 4°C and then flash-frozen at -80°C. Inclusion of
423 *Drosophila* cells allowed internal calibration of enrichment between samples[53]. Upon thawing, cells were
424 permeabilized in nuclear isolation buffer (0.1% Triton X-100, 0.1% sodium deoxycholate) containing 1
425 mM phenylmethylsulfonyl fluoride (PMSF) and EDTA-free protease inhibitor cocktail (Roche)) on ice.
426 Chromatin was fragmented to mononucleosomes by MNase (NEB) in digestion buffer (50 mM Tris pH
427 7.9, 5 mM $CaCl_2$, 1.5 mM DTT) for 7.5 minutes at 37°C, then quenched with EDTA. Soluble chromatin was
428 pre-cleared with Protein A Dynabeads (Thermo Fisher Scientific) pre-adsorbed with normal rabbit IgG
429 (Millipore, 12-370, lot 2972424) for 2 hours at 4°C, and then incubated at 4°C overnight with 1 µg of
430 H3K4me3 (Abcam, ab1012, lot GR158530-14), 2 µg H3K4me1 (Abcam, ab8895, lot GR3198865-1), 1 µg
431 H3K27ac (Active Motif, 39034, lot 16119013), or 2 µg H3K27me3 (Millipore, 07-449, lot 2194165)
432 antibodies pre-adsorbed to 10 µL of Protein A Dynabeads (Thermo Fisher Scientific) in 150 µL of ChIP
433 buffer (20 mM Tris pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.2% Triton X-100, 1 mM PMSF, and EDTA-free

434 protease inhibitor cocktail (Roche)). After washing the beads twice with ice-cold wash buffer (20 mM
435 Tris pH 8, 150 mM NaCl, 2 mM EDTA, 1% (v/v) Triton X-100, 0.1% (w/v) SDS) and twice with ice-cold
436 high-salt wash buffer (20 mM Tris pH 8, 500 mM NaCl, 2 mM EDTA, 1% (v/v) Triton X-100, 0.1% (w/v)
437 SDS), DNA was eluted in elution buffer (100 mM $NaHCO_3$, 1% (w/v) SDS, 50 µg/mL RNase A) for one hour
438 at 55°C, with addition of 1 µL 20 mg/mL proteinase K after the first 30-minutes. DNA was purified by
439 Phenol:chloroform:isoamyl alcohol extraction and used for library preparation with the NEBNext Ultra II
440 DNA Library Prep Kit for Illumina (NEB) with 9 (H3K4me1, H3K27me3) or 13 (H3K4me3, H3K27ac)
441 amplification cycles. Indexed libraries were analyzed for size distribution using the Bioanalyzer High
442 Sensitivity DNA chip (Agilent) and for concentration using the Qubit dsDNA HS Assay (Thermo Fisher
443 Scientific), then pooled and sequenced on the NextSeq 500 platform (Illumina; NCS version 2.2.0.4) for 2
444 × 78 nucleotide paired end reads.

445 **ChIP-seq data analysis.** ChIP-seq reads were aligned to concatenated mouse (GRCm38/mm10) and
446 *Drosophila* (BDGP6.28) genome assemblies using Bowtie2 v2.3.4.1 with options '--very-sensitive --no-
447 unal --no-discordant'[62]. Multi-mapped reads, reads with mapping quality MAPQ < 20, and suspected
448 PCR-duplicates were discarded using Samtools[63]. To scale enrichment according to the *Drosophila* cell
449 spike-ins, the fraction of remaining reads that mapped to the *Drosophila* genome was determined, and a
450 scaling factor was calculated for each sample as *(minimum* Drosophila *read fraction per histone*
451 *mark)/(sample* Drosophila *read fraction)*. After determining a scaling factor, each sample was
452 downsampled to the lowest mapped read count per each histone mark using Samtools, and then reads
453 mapped to the *Drosophila* genome were removed. For visualization, libraries were converted to
454 bedgraph format using the calculated scaling factor in the '-scale' argument of BEDtools genomecov[64].
455 Genome browser views were generated for merged replicates using Spark[65] and TSS profiles using
456 Deeptools[66]. Peak calling was performed with MACS2 version 2.2.6 in each biological replicate with
457 parameters '--broad --llocal 100000'[67]. Peak regions that were detected in only one biological replicate
458 or that overlapped with high-background blacklisted regions defined by the ENCODE project[68] were
459 discarded. Chromosomes X, Y, and M were also excluded from the analyses. We defined TSS locations as
460 the 5' end of the most abundant annotated transcript of each gene detected in our RNA-seq data.
461 H3K4me3 peaks were assigned to genes they overlap the region ±1 kb of the TSS.

462 **mRNA-seq.** GFP^+ RFP^- β-cells (70,000-226,000 per mouse, biological triplicate) were purified and
463 counted by FACS with a 10% spike-in of *Drosophila* S2 cells. *Drosophila* cells were included to detect
464 global changes in cell RNA content[69]. The cells were sorted directly into Trizol LS Reagent (Thermo Fisher
465 Scientific) and total RNA was extracted according to manufacturer's instructions. Residual DNA was
466 digested by Turbo DNase (Thermo Fisher Scientific) for 30-minutes at 37°C, and purified total RNA
467 content was measured using the Qubit RNA HS assay (Thermo Fisher Scientific). mRNA was enriched
468 from 400 ng of total RNA per sample using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB)
469 and used to prepare libraries with the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina
470 (NEB) with 11 amplification cycles. Uniquely indexed libraries were quantified, pooled, and sequenced
471 for 2 × 38 nucleotide reads as described above.

472 **mRNA-seq data analysis.** Sequenced reads were aligned to the concatenated mouse (GRCm38/mm10;
473 Gencode vM24) and *Drosophila* (BDGP6.28; Ensembl release 99) genomes and transcriptomes using
474 STAR v2.7.3a[70] with default settings. Transcript abundance was calculated using Salmon v1.4.0[71] in
475 alignment mode with seqBias and gcBias options. Gene counts and differential expression were
476 calculated using DESeq2[72] with default parameters and implementing the apeglm method of effect-size

477 shrinkage[73]. Genes with fewer than two counts in any sample were filtered out and genes showing ≥ 2-
478 fold difference in expression with P ≤ 0.01 (Wald test with Benjamini and Hochberg adjustment for
479 multiple comparisons) are considered differentially expressed. Except in figure 6f, Drosophila transcripts
480 were removed after abundance calculation; for figure 6f, *Drosophila* transcripts were used as control
481 genes in the estimateSizeFactors function of DESeq2, and then removed.

482 **Enrichment analyses.** Enrichment values of gene sets with genes grouped into H3K4me3 breadth or
483 expression quantiles were determined using Fisher's exact tests. GSEA were performed with
484 clusterProfiler[74] on gene sets available from https://www.gsea-msigdb.org/ or on lists of differentially-
485 expressed genes, using biomaRt[75] when necessary to convert between mouse/human gene orthologues.
486 GO analysis was performed with the Database for Annotation, Visualization and Integrated Discovery
487 (DAVID) (v6.8)[76].

488 **Permutation tests.** Association of H3K27ac peaks with differential gene expression was compared with
489 randomized peak distributions. Genomic locations of H3K27ac peaks were iteratively shuffled 100 times
490 using BEDtools[64] while keeping peak size and intensity, and gene location, constant. Two-tailed *P*-values
491 were calculated as the number of randomized observations more extreme than the observed
492 association frequency, where P < 0.01 ~ none observed.

493 **Single-cell mRNA-seq.** Islets from one WT and one KO mouse were dissociated to single-cell suspensions
494 as described above without FACS-enrichment. The cell suspensions were processed through the
495 Chromium Single Cell 3' protocol using the Chromium Controller (firmware v4.0) with Reagent Kit v3.1
496 (10X Genomics) and Dual Index Kit TT Set A (10X Genomics) according to the manufacturer's instructions
497 for a targeted 5,000-cell recovery. Total cDNA was amplified for 11-cycles, and then 13-cycles during
498 index-ligation. cDNA concentrations were quantified by qPCR with the NEBNext Library Quant Kit for
499 Illumina (NEB). Libraries were pooled and sequenced for 28-90 paired-end nucleotide reads to an
500 average depth of ~70,000 mapped fragments per cell.

501 **Single cell mRNA-seq data analysis.** Cellranger (v5.0.0, 10X Genomics) was used to generate FASTQ files
502 and demultiplex reads. Cell barcode detection, read mapping, quality filtering, and transcript counting
503 were performed using Alevin (Salmon v1.4.0) [77] against the mouse protein-coding transcriptome
504 (Gencode vM24), which we appended with the *Tdtomato* and *Egfp* sequences, as well as the mouse
505 genome (GRCm38/mm10) as mapping decoys for selective alignment[78]. Additional cell quality filtering
506 was performed using Seurat v4.0[79] based on unique features (> 1,000 genes) and mitochondrial RNA
507 content (< 25%). Putative doublets were excluded based on a high transcript count (>90,000) or co-
508 expression of *Tdtomato* and *Egfp* transcripts, and using DoubletFilter[80]. Genes detected in < 3 cells were
509 excluded from analysis. After quality filtering, the dataset comprised 3,703 WT and 3,641 KO cells.
510 Filtered data were scaled and normalized using SCTransform[81] with default parameters. PCA and UMAP
511 dimensionality reduction using the first 60 principal components was used to cluster and visualize cell
512 populations. We then used Slingshot[82] to perform pseudotime analysis on β-cell clusters. For cell
513 entropy analysis, we used a published R function (https://github.com/skannan4/cm-entropy-score [43])
514 which calculate a cell entropy score using Shannon entropy of the 1,000 most highly expressed genes.

515 **Immunohistochemistry.** Mouse pancreata were isolated and fixed in 4% (w/v) paraformaldehyde (PFA)
516 in PBS at 4°C overnight. Human islets were handpicked and fixed in 4% (w/v) PFA in PBS for 1-hour at
517 room temperature and then embedded in 1% agarose. Following washing with PBS and dehydration in
518 ethanol and xylenes, mouse and human tissues were embedded in paraffin and sectioned with a

519 thickness of 5 µm. Immunofluorescent analyses of tissue sections was performed as previously
520 described[24] in a blinded fashion from 4-6 sections per biological sample. Primary antibodies used were:
521 H3K4me3 (Cell Signaling Technologies, C42D8, 1:500), H3K4me1 (Abcam, ab8895, 1:1,000), H3 (Abcam,
522 ab1791, 1:1,000), DPY30 (Atlas Antibodies, HPA043761, 1:500), Insulin (Agilent, IR-002, 1:4), and
523 Glucagon (Sigma Aldrich, G2654, 1:1,000). For in vivo 5-ethynyl-2'-deoxyuridine (EdU) labeling, mice
524 were injected intraperitoneally with 0.5 mg EdU (Carbosynth) twice per day for seven consecutive days,
525 and pancreata were isolated one day after the final injection. EdU detection in tissue sections was
526 performed as described previously[23].

527 **Immunocytochemistry.** Islets were dissociated to single cells 3-4 hours after isolation and plated in
528 technical duplicate in a 24-well cell culture dish with glass coverslips at 50 dispersed islets/well in
529 complete RPMI. The next day, media was changed to 0.5 mL complete RPMI containing 10 µM EdU and
530 either 5.5 mM or 16.7 mM glucose. Cumulative EdU incorporation proceeded for 48-hours in these
531 media, with refreshment after the first 24-hours. Cells were then fixed in 2% PFA in PBS for 20-minutes,
532 then washed once in PBS before labelling EdU for 20-minutes with 5 µM Alexa Fluor 647 azide (Thermo
533 Fisher Scientific) in 100 mM Tris pH 8.5, 1 mM $CuSO_4$, 100 mM sodium ascorbate. Coverslips were
534 subsequently washed three times and permeabilized with 0.5% Triton-X100 in PBS for 30 minutes.
535 Following thre more washes in PBS, sections were incubated with insulin antibody (Agilent, IR-002, 1:4)
536 in 5% FBS in PBS overnight at 4°C. The next day, coverslips were washed three times in PBS, incubated
537 with Alexa Fluor 555 anti-guinea pig (Thermo Fisher Scientific, A-21435, 1:500) and DAPI (Sigma Aldrich,
538 D9542, 1:5,000) in PBS for one-hour at room temperature, washed three more times in PBS, and
539 mounted with Prolong Gold (Thermo Fisher Scientific) mounting medium onto glass slides. Slides were
540 imaged using a Leica TCS SP8 Confocal microscope and images were quantified using Cell Profiler v2[83].
541 EdU+ GFP+ cells were normalized to total GFP+ cells, from 6,000-9,000 total cells per biological sample,
542 for EdU quantification.

543 **Transmission electron microscopy.** Islets were fixed in 2% glutaraldehyde (Sigma Aldrich) in PBS and
544 submitted to the Facility for Electron Microscopy Research of McGill University (Hamilton, ON, Canada).
545 Islets were post-fixed with 1% osmium tetroxide, dehydrated in ethanol, and embedded in Spurr's resin
546 before sectioning with a Leica UCT ultramicrotome. Sections were stained with uranyl acetate and lead
547 citrate and imaged with a JEOL JEM 1200 EX TEMSCAN transmission electron microscope (JEOL,
548 Peabody, MA, USA). Insulin granule density and size were quantified in 7 images per mouse using
549 ImageJ[84]. Mitochondria area was quantified in the same images using CellProfiler v2.2[83].

550 **Immunoblotting.** Islets were lysed in Laemmli sample buffer (2% (w/v) SDS, 10% (v/v) glycerol, 60 mM
551 Tris pH 6.8, 1mM NaF, 1 mM PMSF, protease inhibitor cocktail (Roche)) for 5-minutes at 95°C. Protein
552 concentration was measured by BCA assay (Thermo Fisher Scientific), then β-mercaptoethanol was
553 added to 5% (v/v) and lysates were reboiled for 5-minutes and stored at -80°C. 10 µg lysates were
554 resolved in polyacrylamide gels, transferred to PVDF membranes, blocked for 60-minutes in 5% (w/v)
555 skim milk powder in TBST (20 mM Tris pH 7.6, 150 mM NaCl, 0.1% (v/v) Tween-20) at room temperature
556 and probed with primary antibodies overnight in 5% (w/v) BSA in TBST at 4°C. The next day, after
557 washing membranes three times with TBST, HRP-linked secondary antibodies were added for 1-hour at
558 room temperature, membranes were washed three times with TBST, and detected using ECL system
559 with film. Band intensities were quantified using ImageJ[84]. When normalizing to total H3, membranes
560 were stripped for 20-minutes in mild stripping buffer (200 mM glycine, 0.1% (w/v) SDS, 1% (v/v) Tween-
561 20, pH 2.2) at room temperature, rinsed twice in PBS and twice in TBST for 10 minutes each, then re-

562 blocked and probed for histone H3. For detection of (pre)(pro)insulin, islets were treated for 4-hours
563 with 10 µM MG-132 (Sigma Aldrich) before lysing, and probed using a modified immunoblot protocol to
564 improve peptide detection[85]. Briefly, 10 µg reduced lysates were resolved in 15% acrylamide gel by
565 tris/tricine/urea SDS-PAGE and transferred to PVDF. Blots were blocked for 5 min in 5% (w/v) BSA in PBS
566 with 0.1% Tween-20 (PBST), rinsed in PBST, and crosslinked using 0.4% PFA in PBST for 10 minutes at
567 room temperature. Blots were rinsed three times in TBST then underwent antigen retrieval for 10
568 minutes in a vegetable steamer in antigen retrieval buffer (10 mM citric acid, 1 mM EDTA, 0.05% Tween-
569 20, pH 6.0). Blots were cooled to room temperature then quenched with 200 mM glycine for 10 minutes
570 at room temperature; subsequent steps of immunoblotting were as described above. Primary antibodies
571 used for immunoblots were: H3K4me3 (Cell Signaling Technologies, C42D8, 1:1,000), H3K4me1 (Abcam,
572 ab8895, 1:100,000), H3H27ac (Cell Signaling Technologies, D5E4, 1:1,000), H3K27me3 (Millipore, 07-
573 449, 1:100,000), H3 (Abcam, ab1791, 1:1,000,000), WDR5 (Bethyl, A302-430A, 1:2,000), RBBP5 (Bethyl,
574 A300-109A, 1:10,000), ASH2L (Bethyl, A300-489A, 1:10,000), Insulin (Cell Signaling Technologies, L6B10).

575 **Protein co-immunoprecipitation.** Immunoprecipitation of WDR5 was performed in triplicate in pools of
576 750-900 dispersed islets from 2-3 mice. First, nuclei were enriched from dispersed islet cells by
577 incubation in 0.1% Triton X-100, 0.1% sodium deoxycholate for 10-minutes on ice and pelleting at 600 ×
578 *g* for 5-minutes at 4°C. Chromatin was solubilized in IP buffer (20 mM Tris pH 7.4, 150 mM NaCl, 1 mM
579 EDTA, 5% (v/v) glycerol, 0.5% (v/v) Igepal ca-630, 5 mM $CaCl_2$, 1 mM PMSF, protease inhibitor cocktail
580 (Roche)) with MNase (NEB) for 10-minutes at room temperature. EDTA was added to 5 mM to stop
581 MNase and samples were centrifuged at 14,000 × *g* for 5-minutes at 4°C. The protein concentration in
582 the supernatant was measured by BCA assay (Thermo Fisher Scientific), and an equal quantity of WT and
583 KO nuclear lysate (~30 µg protein) was immunoprecipitated by 1 µg WDR5 (Bethyl, A302-429A) or
584 normal rabbit IgG (Millipore, 12-370) antibodies rotating overnight at 4°C. The next day, 10 µL protein A
585 Dynabeads (Thermo Fisher Scientific) were added to the antibody:lysate mixture and rotated for 2-hours
586 at 4°C. Finally, bead:antibody:antigen complexes were washed six times in ice-cold IP buffer and
587 resuspended in Laemmli sample buffer with 5% (v/v) β-mercaptoethanol for immunoblotting.

588 **Insulin secretion assays.** For in vivo glucose tolerance tests, 2 g/kg glucose in a 20% (w/v) water solution
589 was injected intraperitoneally following a 6-hour fast. Blood glucose was sampled prior to and 15, 30,
590 60, and 120-minutes after glucose injection from the lateral saphenous vein. Serum was collected at the
591 0, 15, and 60-minute time points by centrifuging blood samples at 9,000 × g for 9 minutes at 4°C and
592 stored at -80°C. Serum insulin was measured by ELISA (Alpco) according to the manufacturer's
593 instructions. Ex vivo static insulin secretion assays were performed in technical duplicate in groups of 50
594 islets per mouse one day after isolation. Islets were preincubated in KRBH buffer (2.8 mM glucose, 20
595 mM HEPES, 114 mM NaCl, 4.7 mM KCl, 1.2 mM $MgSO_4$, 1.2 mM KH_2PO_4, 2.5 mM $CaCl_2$, 24 mM $NaHCO_3$,
596 0.1% (w/v) BSA, pH 7.3) for one hour at 37°C and then sequentially for 45-minutes in KRBH, either 16.7
597 mM glucose or 15 mM dimethyl α-ketoglutarate (Sigma Aldrich), and finally 30 mM KCl. Conditioned
598 medium was collected after each incubation. At the completion of the assay islets were lysed in RIPA
599 Buffer (Pierce) containing 1 mM PMSF and EDTA-free protease inhibitor cocktail (Roche)) at 95°C for 5
600 minutes. Insulin concentration was measured in conditioned medium and islet lysate by ELISA (Alpco)
601 and normalized to DNA concentration of islet lysate determined by Qubit dsDNA HS Assay (Thermo
602 Fisher Scientific).

603 **Calcium imaging.** Intact islets were cultured for two days to allow for attachment onto glass coverslips.
604 Prior to imaging, coverslips were transferred to a 2 mL imaging chamber and stained with 5 µM Fura-

605 2AM (Molecular Probes, F1221) for 30 minutes at 37°C in complete RPMI. The imaging chamber loaded
606 with a coverslip was then mounted on a Leica SP8 Laser Scanning Confocal Microscope with a 10×
607 objective lens and continuously washed with 2.8 mM glucose HEPES-buffered Ringer's solution (5.5 mM
608 KCl, 2 mM $CaCl_2$, 1 mM $MgCl_2$, 20 mM HEPES, 144 mM NaCl, pH 7.4) for 30-minutes. Throughout the
609 experiment, islets were continuously perifused at a flow rate of 2.5 mL/min with Ringer's solution.
610 Glucose and KCl concentrations were adjusted by iso-osmotic substitution of NaCl where indicated.
611 Fura-2 was excited at 340 nm and 380 nm and emitted fluorescence was detected at approximately 510
612 nm. Cytosolic Ca^{2+} levels are expressed as the ratio of Fura-2 fluorescence emission intensity (F_{340}/F_{380})
613 of 3-9 islets per biological replicate.

614 **Respirometry.** Islets were dissociated to single cells 3-4 hours after isolation and plated at 100
615 islets/well in a Seahorse 24-well cell culture dish in technical duplicate. Cells were cultured for two days
616 in complete RPMI to allow for attachment to the culture dish and then for 1-hour in XF RPMI medium
617 (Agilent) with 2.8 mM glucose, 1% FBS, 1 mM pyruvate, and 2 mM glutamine to reach a metabolic
618 baseline. Oxygen consumption was then measured using a Seahorse XF24 Extracellular Flux Analyser
619 (Agilent). The injection ports were loaded with the following compounds in XF Assay Media (Agilent): A,
620 167 mM glucose; B, 10 µM oligomycin; C, 15 µM FCCP; D, 1 µM antimycin and 0.5 µM rotenone.
621 Measurements were normalized to the DNA content of assayed cells determined by Qubit HS DNA Assay
622 (Thermo Fisher Scientific) after lysing the assayed cells in RIPA buffer.

623 **Analysis of publicly available data.** Publicly available data were reprocessed for consistency. GEO:
624 GSE50244, GSE25126, GSE107489 were downloaded from the Sequence Read Archive (SRA) using the
625 sratoolkit (https://github.com/ncbi/sra-tools) and analyzed according to the experimental application,
626 as described above. In the case of human single-cell RNA-seq dataset (GEO: GSE124742), we used the
627 count matrix provided because our analysis pipeline was not applicable to the Smart-seq method.

628

629 **References**

630 1. Aravind, L., Abhiman, S. & Iyer, L. M. Natural history of the eukaryotic chromatin protein
631 methylation system. *Progress in Molecular Biology and Translational Science* **101**, (Elsevier Inc.,
632 2011).

633 2. Karlić, R., Chung, H. R., Lasserre, J., Vlahoviček, K. & Vingron, M. Histone modification levels are
634 predictive for gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2926–2931 (2010).

635 3. Mentch, S. J. *et al.* Histone Methylation Dynamics and Gene Regulation Occur through the
636 Sensing of One-Carbon Metabolism. *Cell Metab.* **22**, 861–873 (2015).

637 4. Greer, E. L. & Shi, Y. Histone methylation : a dynamic mark in health , disease and inheritance.
638 *Nat. Rev. Genet.* **13**, 343–357 (2012).

639 5. Chen, K. *et al.* Broad H3K4me3 is associated with increased transcription elongation and
640 enhancer activity at tumor-suppressor genes. *Nat. Genet.* **47**, 1149–1157 (2015).

641 6. Brookes, E. & Shi, Y. Diverse epigenetic mechanisms of human disease. *Annual Review of*
642 *Genetics* **48**, (2014).

643 7. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*
644 **129**, 823–837 (2007).

645 8. Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411
646 (2002).

647 9. Benayoun, B. A. *et al.* H3K4me3 breadth is linked to cell identity and transcriptional consistency.
648 *Cell* **158**, 673–688 (2014).

649 10. Cheng, J. *et al.* A role for H3K4 monomethylation in gene repression and partitioning of
650 chromatin readers. *Mol. Cell* **53**, 979–992 (2014).

651 11. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters
652 and enhancers in the human genome. *Nat. Genet.* **39**, 311–8 (2007).

653 12. Wang, W. *et al.* SET-9 and SET-26 are H3K4me3 readers and play critical roles in germline
654 development and longevity. *Elife* **7**, 1–33 (2018).

655 13. Local, A. *et al.* Identification of H3K4me1-associated proteins at mammalian enhancers. *Nat.*
656 *Genet.* **50**, (2017).

657 14. Lauberth, S. M. *et al.* H3K4me3 interactions with TAF3 regulate preinitiation complex assembly
658 and selective gene activation. *Cell* **152**, 1021–1036 (2013).

659 15. Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of Lysine 4 on Histone H3: Intricacy of
660 Writing and Reading a Single Epigenetic Mark. *Mol. Cell* **25**, 15–30 (2007).

661 16. Yan, J. *et al.* Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions
662 at enhancers. *Cell Res.* **28**, 204–220 (2018).

663 17. Kumar, D. & Jothi, R. Bivalent chromatin protects reversibly repressed genes from irreversible
664 silencing. *bioRxiv* (2020).

665 18. Rickels, R. *et al.* Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like

666 proteins at enhancers is dispensable for development and viability. *Nat. Genet.* **49**, 1647–1653
667 (2017).

668 19. Dorighi, K. M. *et al.* Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from
669 Promoters Independently of H3K4 Monomethylation. *Mol. Cell* **66**, 568-576.e4 (2017).

670 20. Hödl, M. & Basler, K. Transcription in the absence of histone H3.2 and H3K4 methylation. *Curr.*
671 *Biol.* **22**, 2253–2257 (2012).

672 21. Hu, D. *et al.* Not All H3K4 Methylations Are Created Equal: Mll2/COMPASS Dependency in
673 Primordial Germ Cell Specification. *Mol. Cell* **65**, 460-475.e6 (2017).

674 22. Piunti, A. & Shilatifard, A. Epigenetic balance of gene expression by polycomb and compass
675 families. *Science (80-.).* **352**, (2016).

676 23. Campbell, S. A., McDonald, C. L., Krentz, N. A. J., Lynn, F. C. & Hoffman, B. G. TrxG Complex
677 Catalytic and Non-catalytic Activity Play Distinct Roles in Pancreas Progenitor Specification and
678 Differentiation. *Cell Rep.* **28**, 1830-1844.e6 (2019).

679 24. Campbell, S. A. *et al.* H3K4 trimethylation is required for postnatal pancreatic endocrine cell
680 functional maturation. *bioRxiv* (2020). doi:https://doi.org/10.1101/2020.11.29.402990

681 25. Chatterjee, S., Khunti, K. & Davies, M. J. Type 2 diabetes. *Lancet* **389**, 2239–2251 (2017).

682 26. Dayeh, T. *et al.* Genome-Wide DNA Methylation Analysis of Human Pancreatic Islets from Type 2
683 Diabetic and Non-Diabetic Donors Identifies Candidate Genes That Influence Insulin Secretion.
684 *PLoS Genet.* **10**, (2014).

685 27. Fadista, J. *et al.* Global genomic and transcriptomic analysis of human pancreatic islets reveals
686 novel genes influencing glucose metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13924–13929
687 (2014).

688 28. Ling, C. & Groop, L. Epigenetics: A molecular link between environmental factors and type 2
689 diabetes. *Diabetes* **58**, 2718–2725 (2009).

690 29. Vanderkruk, B. & Hoffman, B. G. Metabolism as a central regulator of β-cell chromatin state. *FEBS*
691 *J.* 1–11 (2020). doi:10.1111/febs.15562

692 30. Miguel-Escalada, I. *et al.* Human pancreatic islet three-dimensional chromatin architecture
693 provides insights into the genetics of type 2 diabetes. *Nat. Genet.* **51**, 1137–1148 (2019).

694 31. Wortham, M. *et al.* Nutrient regulation of the islet epigenome controls adaptive insulin secretion.
695 *bioRxiv* 1–71 (2019). doi:10.1101/742403

696 32. Lu, T. T. H. *et al.* The Polycomb-Dependent Epigenome Controls β Cell Dysfunction,
697 Dedifferentiation, and Diabetes. *Cell Metab.* **27**, 1294-1308.e7 (2018).

698 33. Talchai, C., Xuan, S., Lin, H. V., Sussel, L. & Accili, D. Pancreatic β Cell Dedifferentiation as a
699 Mechanism of Diabetic β Cell Failure. *Cell* **150**, 1223–1234 (2012).

700 34. Neelankal John, A., Ram, R. & Jiang, F.-X. RNA-Seq Analysis of Islets to Characterise the
701 Dedifferentiation in Type 2 Diabetes Model Mice db/db. *Endocr. Pathol.* **29**, 207–221 (2018).

702 35. Hsiao, L. L. *et al.* A compendium of gene expression in normal human tissues. *Physiol. Genomics*
703 **2002**, 97–104 (2002).

704 36. Bramswig, N. C. *et al.* Epigenomic plasticity enables human pancreatic a to b cell reprogramming.
705 *J. Clin. Invest.* **123**, 1275–1284 (2013).

706 37. Camunas-Soler, J. *et al.* Patch-Seq Links Single-Cell Transcriptomes to Human Islet Dysfunction in
707 Diabetes. *Cell Metab.* **31**, 1017-1031.e4 (2020).

708 38. Haddad, J. F. *et al.* Structural Analysis of the Ash2L/Dpy-30 Complex Reveals a Heterogeneity in
709 H3K4 Methylation. *Structure* 1–10 (2018). doi:10.1016/j.str.2018.08.004

710 39. Tremblay, V. *et al.* Molecular basis for DPY-30 association to COMPASS-like and NURF complexes.
711 *Structure* **22**, 1821–1830 (2014).

712 40. Schuettengruber, B., Bourbon, H.-M., Di Croce, L. & Cavalli, G. Genome Regulation by Polycomb
713 and Trithorax: 70 Years and Counting. *Cell* **171**, 34–57 (2017).

714 41. Millership, S. J. *et al.* Neuronatin regulates pancreatic β cell insulin content and secretion. *J. Clin.*
715 *Invest.* **128**, 3369–3381 (2018).

716 42. Dalgaard, K. *et al.* Trim28 Haploinsufficiency Triggers Bi-stable Epigenetic Obesity. *Cell* **164**, 353–
717 364 (2016).

718 43. Kannan, S., Farid, M., Lin, B., Miyamoto, M. & Kwon, C. Transcriptomic entropy benchmarks stem
719 cell-derived cardiomyocyte maturation against endogenous tissue at single cell level. (2020).
720 doi:10.1101/2020.04.02.022632

721 44. Zhang, H., Gao, L., Anandhakumar, J. & Gross, D. S. Uncoupling Transcription from Covalent
722 Histone Modification. *PLoS Genet.* **10**, (2014).

723 45. Morgan, M. A. J. & Shilatifard, A. Reevaluating the roles of histone-modifying enzymes and their
724 associated chromatin modifications in transcriptional regulation. *Nat. Genet.* **52**, 1271–1281
725 (2020).

726 46. Sze, C. C. *et al.* Histone H3K4 methylation-dependent and -independent functions of
727 set1A/COMPASS in embryonic stem cell self-renewal and differentiation. *Genes Dev.* **31**, 1732–
728 1737 (2017).

729 47. Deering, T. G., Ogihara, T., Trace, A. P., Maier, B. & Mirmira, R. G. Methyltransferase set7/9
730 maintains transcription and euchromatin structure at islet-enriched genes. *Diabetes* **58**, 185–193
731 (2009).

732 48. Li, J., Liu, Y., Kim, T. H., Min, R. & Zhang, Z. Gene expression variability within and between
733 human populations and implications toward disease susceptibility. *PLoS Comput. Biol.* **6**, (2010).

734 49. Ecker, S., Pancaldi, V., Valencia, A., Beck, S. & Paul, D. S. Epigenetic and Transcriptional Variability
735 Shape Phenotypic Plasticity. *BioEssays* **40**, 1–11 (2018).

736 50. Rangaraju, S. *et al.* Suppression of transcriptional drift extends C. elegans lifespan by postponing
737 the onset of mortality. *Elife* **4**, 1–39 (2015).

738 51. Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166**, 358–368
739 (2016).

740 52. Larsson, A. J. M. *et al.* Genomic encoding of transcriptional burst kinetics. *Nature* **565**, 251–254
741 (2019).

742 53. Orlando, D. A. *et al.* Quantitative ChIP-Seq normalization reveals global modulation of the
743 epigenome. *Cell Rep.* **9**, 1163–1170 (2014).

744 54. Taruttis, F. *et al.* External calibration with Drosophila whole-cell spike-ins delivers absolute mRNA
745 fold changes from human RNA-Seq and qPCR data. *Biotechniques* **62**, 53–61 (2017).

746 55. Chagoyen, M. & Poyatos, J. F. Complex genetic and epigenetic regulation deviates gene
747 expression from a unifying global transcriptional program. *PLoS Comput. Biol.* **15**, e1007353
748 (2019).

749 56. Wang, C. *et al.* Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition.
750 *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11871–11876 (2016).

751 57. Martin, B. J. E. *et al.* The majority of histone acetylation is a consequence of transcription. (2019).

752 58. Puri, S. *et al.* Replication confers β cell immaturity. *Nat. Commun.* **9**, (2018).

753 59. Salinno, C. *et al.* β-Cell Maturation and Identity in Health and Disease. *Int. J. Mol. Sci.* **20**, 1–20
754 (2019).

755 60. Yang, J. H. *et al.* Erosion of the epigenetic landscape and loss of cellular identity as a cause of
756 aging in mammals. *bioRxiv* (2019). doi:10.1101/808642

757 61. Brind'Amour, J. *et al.* An ultra-low-input native ChIP-seq protocol for genome-wide profiling of
758 rare cell populations. *Nat. Commun.* **6**, 1–8 (2015).

759 62. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–
760 359 (2012).

761 63. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079
762 (2009).

763 64. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features.
764 *Bioinformatics* **26**, 841–842 (2010).

765 65. Kurtenbach, S. & William Harbour, J. SparK: A Publication-quality NGS Visualization Tool. *bioRxiv*
766 1–4 (2019). doi:10.1101/845529

767 66. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis.
768 *Nucleic Acids Res.* **44**, W160–W165 (2016).

769 67. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, (2008).

770 68. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic
771 Regions of the Genome. *Sci. Rep.* **9**, 1–5 (2019).

772 69. Taruttis, F. *et al.* External calibration with Drosophila whole cell spike ins delivers absolute mRNA
773 fold changes from human RNA seq and qPCR data. *Biotechniques* **62**, 53–61 (2017).

774 70. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

775 71. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-
776 aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

777 72. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-

778 seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

779 73. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-Tailed prior distributions for sequence count data:
780 Removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).

781 74. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological
782 themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).

783 75. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of
784 genomic datasets with the R/ Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).

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

787 77. Srivastava, A., Malik, L., Smith, T., Sudbery, I. & Patro, R. Alevin efficiently estimates accurate
788 gene abundances from dscRNA-seq data. *Genome Biol.* **20**, (2019).

789 78. Sarkar, H., Zakeri, M., Malik, L. & Patro, R. Towards selective-alignment: Bridging the accuracy
790 gap between alignment-based and alignment-free transcript quantification. *bioRxiv* 1–7 (2017).
791 doi:10.1101/138800

792 79. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

793 80. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell
794 RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**, 329-337.e4 (2019).

795 81. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data
796 using regularized negative binomial regression. *bioRxiv* 1–15 (2019). doi:10.1101/576827

797 82. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics.
798 *BMC Genomics* **19**, 1–16 (2018).

799 83. Carpenter, A. E. *et al.* CellProfiler: Image analysis software for identifying and quantifying cell
800 phenotypes. *Genome Biol.* **7**, (2006).

801 84. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis.
802 *Nat. Methods* **9**, 671–675 (2012).

803 85. Okita, N. *et al.* Modified Western blotting for insulin and other diabetes-associated peptide
804 hormones. *Sci. Rep.* **7**, 1–11 (2017).

805

Figure 1

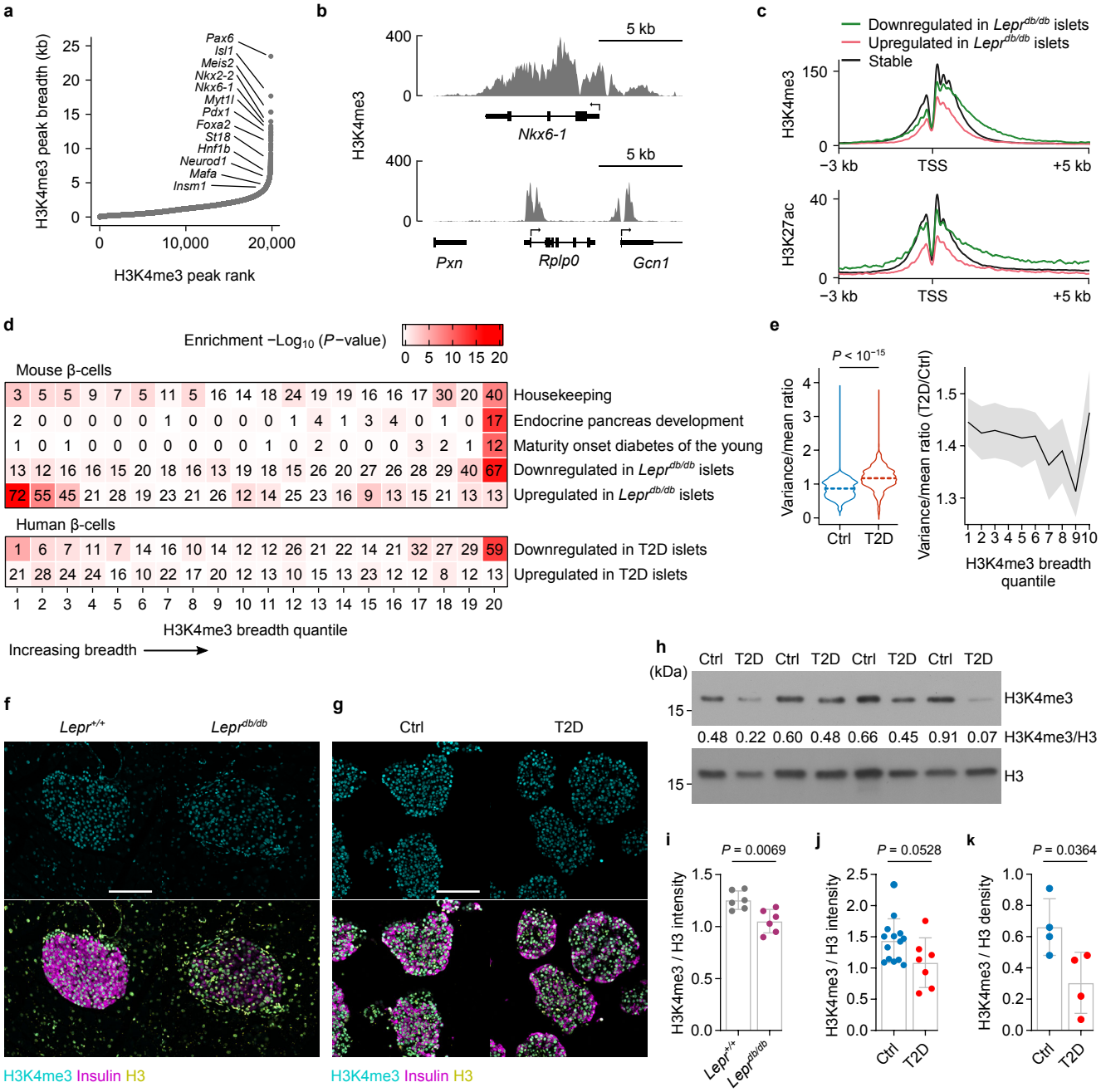

Figure 1. H3K4me3 stratifies β-cell genes and is dysregulated in type 2 diabetes.

a. H3K4me3 peaks in mouse β-cells ranked by peak breadth. Peaks associated with β-cell transcription factors are labeled.

b. Example of H3K4me3 density at the loci of *Nkx6-1*—a β-cell transcription factor—and *Rplp0*—an expression-matched housekeeping gene.

c. Average density profiles of H3K4me3 and H3K27ac in the regions spanning -3 kb to +5 kb with respect to the TSS of genes that are upregulated, downregulated, or stably expressed in *Lepr*$^{db/db}$ versus *Lepr*$^{+/+}$ islets.

d. Heatmap showing the enrichment of gene sets in mouse and human β-cell transcriptomes that have been grouped into quantiles according to H3K4me3 peak breadth at promoters, ranked from narrow to broad. The number of matched genes in each quantile (556 mouse genes, 721 human genes per quantile) are shown. *P*-values were calculated using two-tailed Fisher's exact tests.

e. Violin plot with mean showing the variance/mean ratio of gene expression across β-cells from donors with or without T2D in single-cell RNA-seq data (n = 415 Ctrl, 193 T2D β-cells). *P*-value calculated by two-tailed t-test with Welch's correction.

Figure 1. H3K4me3 stratifies β-cell genes and is dysregulated in type 2 diabetes, cont'd

f. Variance/mean ratio of T2D versus Ctrl β-cell gene expression (y-axis) plotted as a function of H3K4me3 breadth in quantiles (x-axis) (n = 84 genes/quantile). Data presented as mean with 95% confidence interval.

g. Representative fluorescence immunohistochemistry (IHC) images of islets in a *Lepr*$^{+/+}$ (left) and *Lepr*$^{db/db}$ (right) pancreas showing H3K4me3 (cyan), insulin (magenta), and histone H3 (yellow).

h. Representative IHC images of islets from a non-diabetic (left) and type-2-diabetic donor (right) showing H3K4me3 (cyan), insulin (magenta), and histone H3 (yellow).

i. Immunoblot of H3K4me3 and total histone H3 in islets from donors with and without T2D. Numbers between the bands indicate the H3-normalized H3K4me3 band intensity.

j. Bar graph showing immunofluorescence intensity of H3-normalized H3K4me3 in *Lepr*$^{+/+}$ and *Lepr*$^{db/db}$ β-cell nuclei. Data presented as mean ± SD with values from individual mice (n = 6). *P*-value calculated by two-tailed t-test with Welch's correction.

k. Bar graph showing immunofluorescence intensity of H3-normalized H3K4me3 in β-cells from human donors with and without T2D. Data presented as mean ± SD with values from individual donors (n = 14 Ctrl, 7 T2D). *P*-value calculated by two-tailed t-test with Welch's correction.

l. Bar graph showing immunoblot band densities of H3-normalized H3K4me3 in islets from human donors with and without T2D. Data presented as mean ± SD with values from individual donors (n = 4). *P*-value calculated by two-tailed t-test with Welch's correction.

Scale bars: 100 µm.

Figure 2

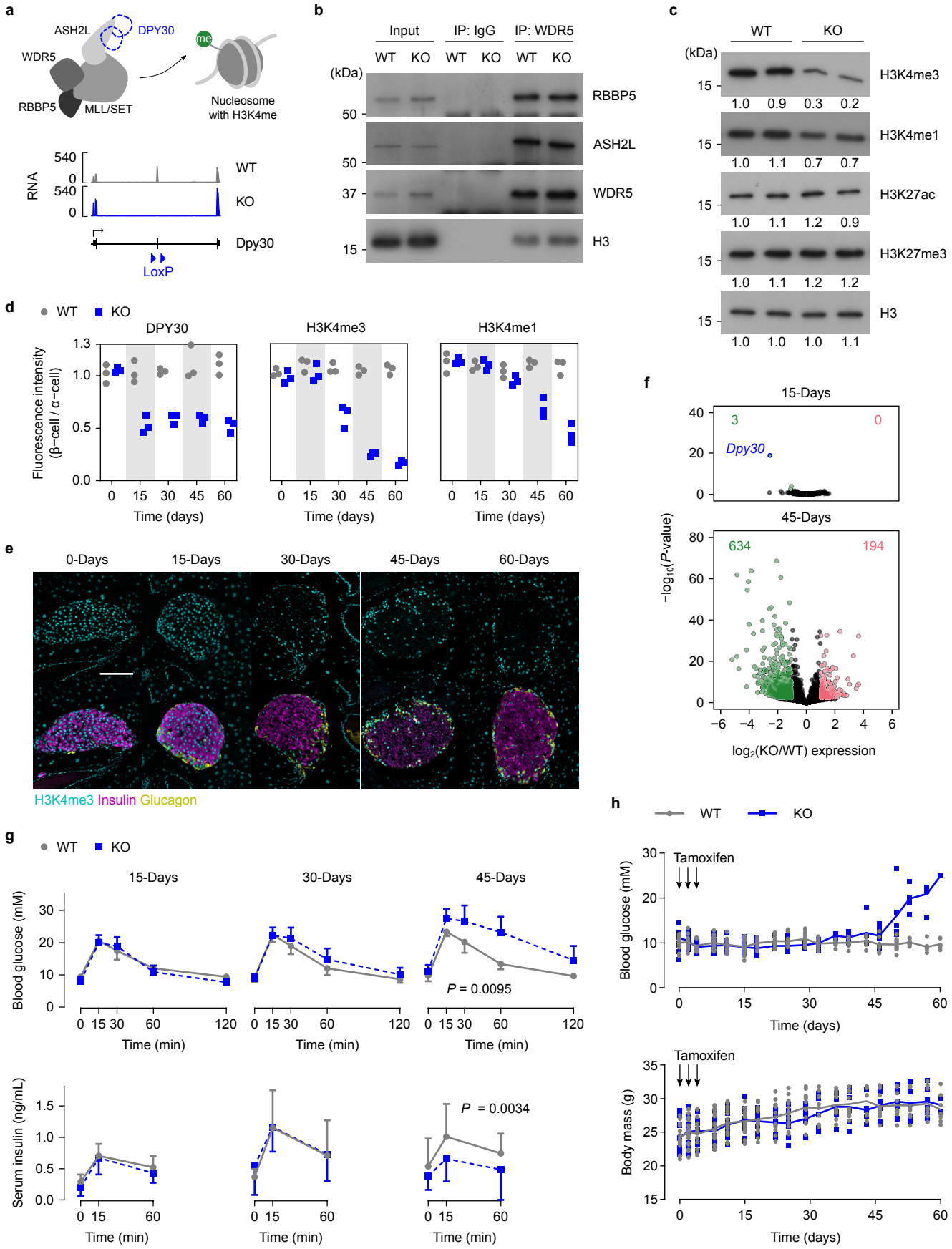
a
ASH2L
DPY30
WDR5
RBBP5
MLL/SET
me
Nucleosome with H3K4me
RNA
540
0
WT
KO
Dpy30
LoxP
b
Input
IP: IgG
IP: WDR5
WT KO WT KO WT KO
(kDa)
50
RBBP5
ASH2L
37
WDR5
15
H3
c
WT
KO
(kDa)
H3K4me3
1.0 0.9 0.3 0.2
H3K4me1
1.0 1.1 0.7 0.7
H3K27ac
1.0 1.1 1.2 0.9
H3K27me3
1.0 1.1 1.2 1.2
H3
1.0 1.0 1.0 1.1
d
WT
KO
DPY30
H3K4me3
H3K4me1
Fluorescence intensity (β−cell / α−cell)
Time (days)
e
0-Days
15-Days
30-Days
45-Days
60-Days
H3K4me3 Insulin Glucagon
f
15-Days
3
0
Dpy30
45-Days
634
194
−log10(P-value)
log2(KO/WT) expression
g
WT
KO
15-Days
30-Days
45-Days
Blood glucose (mM)
Time (min)
P = 0.0095
Serum insulin (ng/mL)
P = 0.0034
h
WT
KO
Tamoxifen
Blood glucose (mM)
Time (days)
Body mass (g)

Figure 2. Reduction to H3K4 methylation in β-cells causes glucose intolerance and hyperglycemia.
a. Top: schematic showing the core subunits of the Trithorax group (TrxG) complex and a nucleosome methylated on H3K4 by the TrxG complex. Bottom: genome-aligned RNA-seq reads over the *Dpy30* gene locus in WT and KO β-cells.
b. Immunoblots showing TrxG subunits RBBP5, ASH2L, and WDR5, and nucleosome protein histone H3, co-IP'ed with WDR5 or an IgG control from WT and KO islets. Representative immunoblots of 3 independent Co-Ips are shown.
c. Immunoblots of H3K4me3, H3K4me1, H3K27ac, H3K27me3, and total histone H3 in islets of 2 WT and 2 KO mice. Numbers beneath each band show the band intensity normalized to the left-most sample.
d. Mean immunofluorescence intensity of DPY30, H3K4me3, and H3K4me1 in WT and KO β-cell nuclei at the indicated days after tamoxifen administration. Data are normalized to fluorescence intensity of α-cell nuclei; n = 3 of each genotype and time-point.
e. Example IHC images of KO islets used for calculations in (d) showing H3K4me3 (cyan), insulin (magenta), and glucagon (yellow). Scale bar: 100 µm.
f. Scatterplots showing $\log_2$(fold-change) and $-\log_{10}$(P-value) of gene expression in KO versus WT cells 15- and 45-days after tamoxifen administration. Genes showing a ≥2-fold increase or decrease in expression with $P \leq 0.01$ (Wald test with Benjamini-Hochberg correction) are coloured red or green, respectively, and enumerated above. Genes not meeting those thresholds are black. The full-length transcript of *Dpy30* is outlined and labeled in blue.
g. Blood glucose (top) and serum insulin (bottom) concentration during IPGTT in WT and KO mice 15-, 30-, and 45-days after tamoxifen administration. Data presented as mean ± SD (n = 8 mice for glucose measurements, 8-15 mice for insulin measurements). *P*-values were calculated from AUCs using multiple two-tailed t-tests with Welch's and Benjamini-Hochberg corrections.
h. Unfasted blood glucose (top) and body mass (bottom) of WT and KO mice during the 60-days after tamoxifen administration. Data are presented as individual measurements with mean (n = 17 WT, 14 KO, however tracking was stopped after a blood glucose reading ≥ 20 mM). Days that tamoxifen was administered are indicated by downward arrows.

Figure 3

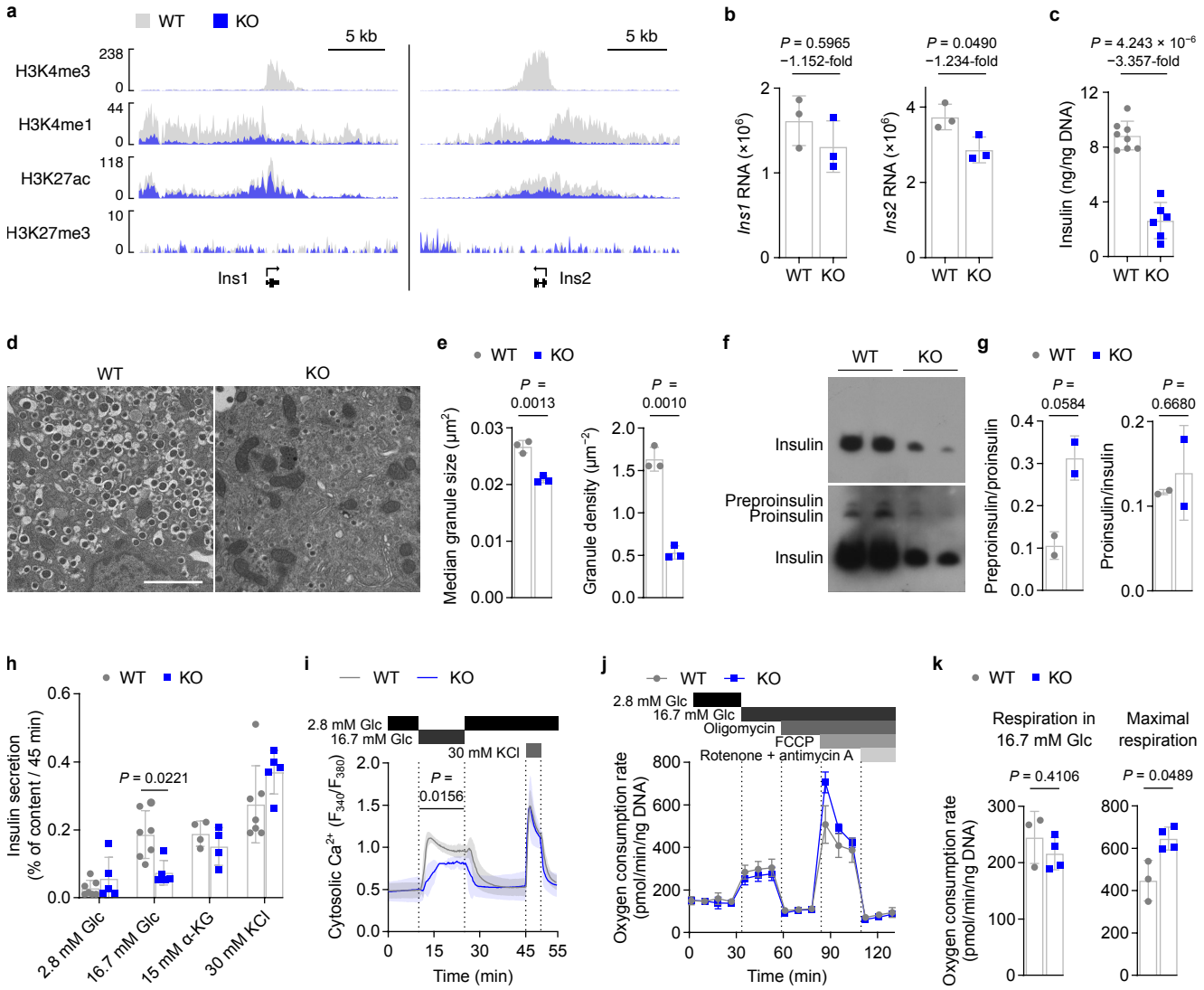

Figure 3. H3K4 methylation maintains the glucose-response in β-cells.

a. Genome browser representation of the levels of H3K4me3, H3K4me1, H3K27ac, and H3K27me3 at the *Ins1* and *Ins2* gene loci in WT and KO chromatin. Data are merged from 2 independent ChIP-seq experiments.

b. *Ins1* and *Ins2* mRNA levels in WT and KO cells. Data presented as mean ± SD with values from individual mice (n = 3). RNA expression levels and *P*-values were calculated from RNA-Seq data using DESeq2 (Wald test with Benjamini-Hochberg correction).

c. Insulin content in WT and KO islets normalized to DNA content. Data presented as mean ± SD with values from individual mice (n = 8 WT, 6 KO). *P*-value calculated by two-tailed t-test with Welch's correction.

d. Representative transmission electron micrograph (TEM) showing insulin granules in β-cells from a WT and a KO mouse. Scale bar: 2 μm.

e. Quantification of median insulin granule size and density in TEM images of β-cells from WT and KO mice. Data presented as mean ± SD with values from individual mice (n = 3). *P*-values calculated by two-tailed t-test with Welch's correction.

f. Representative immunoblot showing insulin (top) and preproinsulin, proinsulin, and overexposed insulin (bottom) protein bands in islet lysate from 2 WT and 2 KO mice.

g. Quantification of the ratio between preproinsulin versus proinsulin and between proinsulin versus insulin band intensities in immunoblots of WT and KO islet lysate. Data presented as mean ± SD with values from individual mice (n = 2). *P*-values calculated by two-tailed t-test with Welch's correction.

Figure 3. H3K4 methylation maintains the glucose-response in β-cells, cont'd

h. Insulin secretion from WT and KO islets during static stimulation with the indicated secretagogues. Insulin secretion is normalized to islet insulin content. Data presented as mean ± SD with values from individual mice (n = 4-7 WT, 4-5 KO). *P*-values calculated by multiple two-tailed t-tests with Welch's and Benjamini-Hochberg corrections; only values < 0.05 are shown.

i. Cytosolic Ca2+ concentration over time in islets from WT and KO mice during perifusion of the indicated glucose and KCl concentrations. Mean and SD (n = 4 WT, 3 KO) are shown. Dotted lines indicate times that glucose or KCl concentration of perfusate is changed. *P*-values calculated for AUC in each time block by one-way ANOVA between genotypes; only values < 0.05 are shown.

j. Oxygen consumption rate of WT and KO dispersed islets over time during sequential treatment with the indicated compounds. Oxygen consumption rate is normalized to DNA content of the assayed cells. Data presented as mean ± SD (n = 3 WT, 4 KO). Dotted lines indicate times that the indicated compounds were injected.

k. Respiration of WT and KO islet cells in 16.7 mM glucose and their maximal respiration capacity, inferred from the data in (j). *P*-values calculated by two-tailed t-test with Welch's correction of the appropriate time-point. Data presented as mean ± SD with values from individual mice (n = 3 WT, 4 KO).

Figure 4

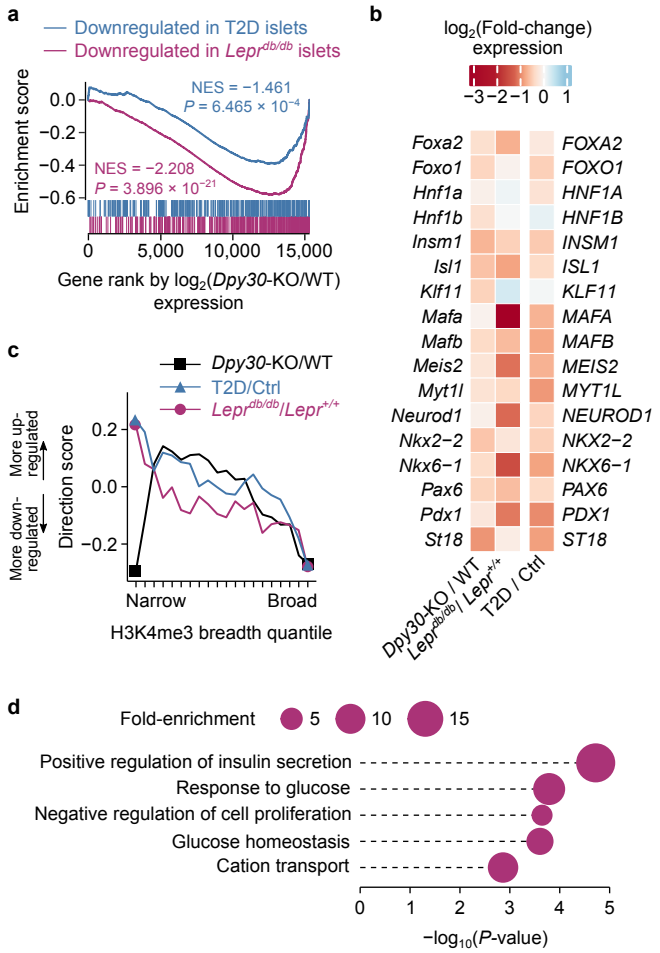

Figure 4. H3K4 methylation enforces expression of genes that are downregulated in type 2 diabetes.
a. Gene set enrichment analysis of genes that are downregulated in *Lepr*$^{db/db}$ versus *Lepr*$^{+/+}$ mouse islets ($P \leq 0.01$, Wald test with Benjamini-Hochberg correction, $\log_2$(fold-change) ≥ 1) or human donors with T2D versus nondiabetic controls ($P \leq 0.01$, Wald test with Benjamini-Hochberg correction) against mRNA-seq of Dpy30-KO islet cells. Human gene identities were converted to mouse orthologues. Enrichment *P*-values were calculated by permutation test with Benjamini-Hochberg correction. NES: normalized enrichment score.
b. Heat map showing mean expression $\log_2$(fold-change) of β-cell transcription factor genes in Dpy30-KO versus WT islet cells, *Lepr*$^{db/db}$ versus *Lepr*$^{+/+}$ mouse islets, and islets from human donors with T2D versus nondiabetic controls.
c. Promoter H3K4me3 breadth quantiles ranked from narrow to broad (x-axis) plotted against the net direction of gene expression change (y-axis) of genes in each quantile for Dpy30-KO, *Lepr*$^{db/db}$, and T2D islet cells. A direction score of 1 would indicate that mean expression of all genes in a quantile are greater in KO than WT. 556 genes per quantile.
d. Gene ontology biological processes associated with genes that are downregulated in both Dpy30-KO and *Lepr*$^{db/db}$ islets. The five terms with lowest enrichment *P*-values are shown. Size of the dot indicates the fold-enrichment of the term. *P*-values represent EASE scores calculated using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (v6.8).

Figure 5

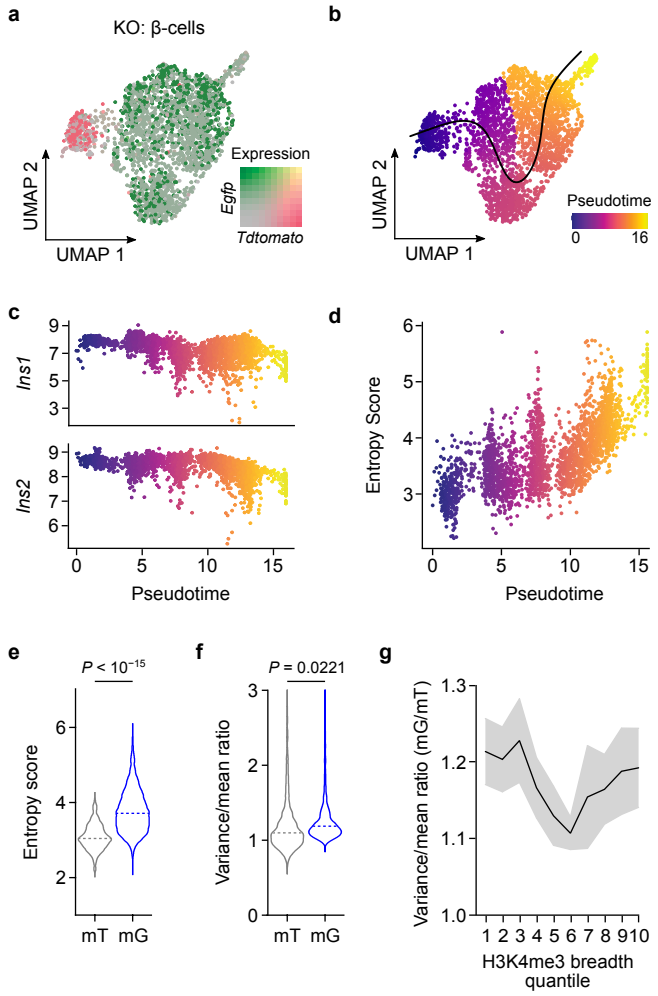

Figure 5. Reduction to H3K4 methylation increases gene expression entropy in β-cells.
a. UMAP visualization of β-cell transcriptomes from a KO mouse. Expression of *Tdtomato*, *Egfp*, and their overlap in single cells are shown as a colour gradient (n = 2,597 β-cells).
b. Pseudotime representation of (a). The *Tdtomato+* β-cell cluster was selected as a ground-state.
c. Scatterplot showing *Ins1* and *Ins2* expression in β-cells plotted against pseudotime.
d. Scatterplot showing gene expression entropy scores of β-cells plotted against pseudotime.
e. Violin plot with mean showing the transcriptional entropy of *Egfp+* cluster β-cells as compared to those in the Tdtomato+ cluster (n = 274 *Tdtomato+* cluster, 2323 *Egfp+* cluster β-cells). *P*-value calculated by two-tailed t-test with Welch's correction.
f. Violin plot with mean showing the variance/mean ratio of gene expression across β-cells in *Egfp+* cluster as compared to the *Tdtomato+* cluster (n = 274 *Tdtomato+* cluster, 2323 *Egfp+* cluster β-cells). *P*-value calculated by paired two-tailed t-test with Welch's correction.
g. Variance/mean ratio *Egfp+* cluster versus *Tdtomato+* cluster β-cell gene expression (y-axis) plotted as a function of H3K4me3 breadth in quantiles (x-axis) (n = 151 genes/quantile). Data presented as mean with 95% confidence interval.

Figure 6

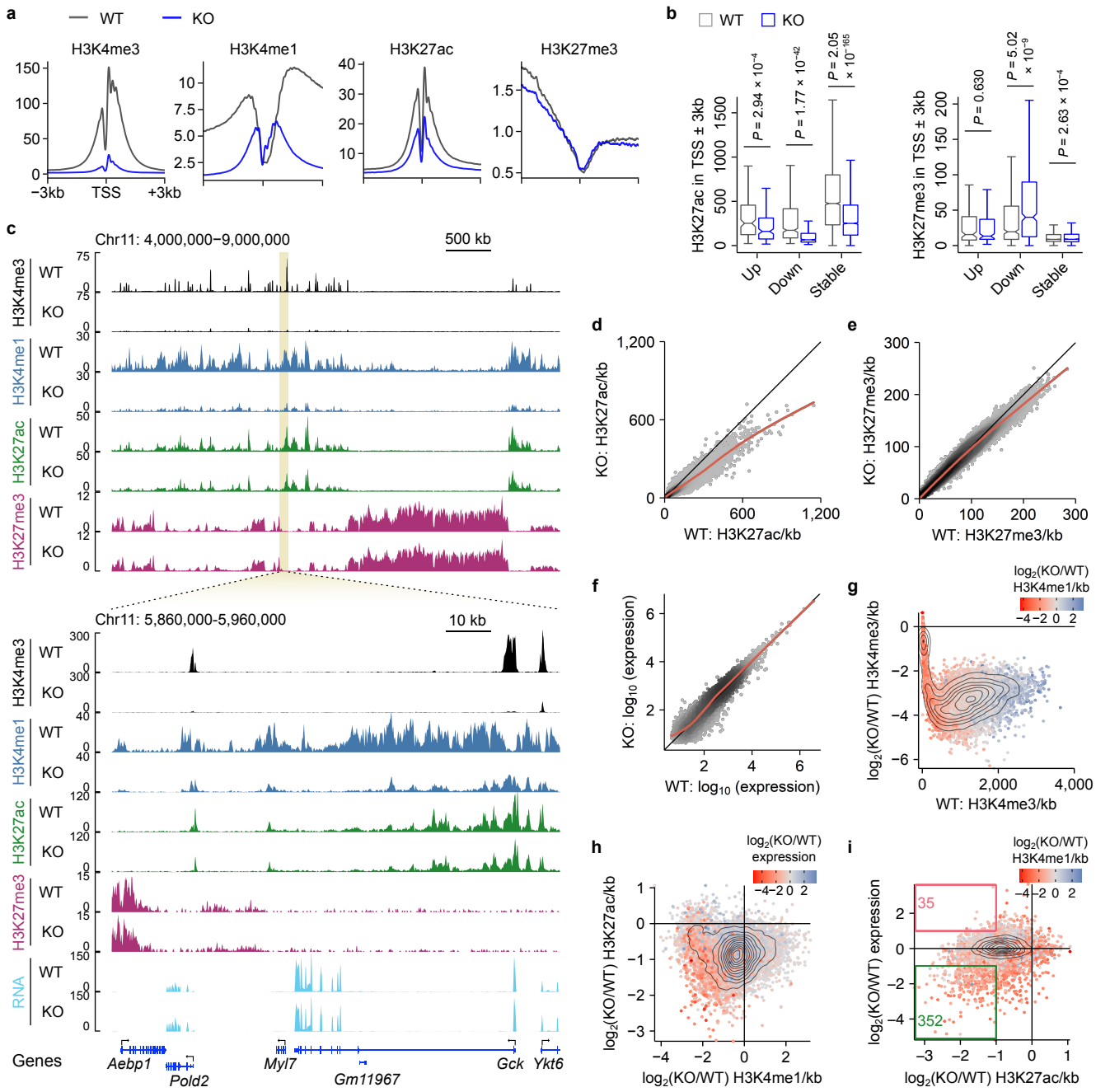

Figure 6. Dilution of epigenetic complexity in Dpy30-KO β-cells.

a. Average enrichment profiles of H3K4me3, H3K4me1, H3K27ac, and H3K27me3 in the region ±3 kb with respect to the TSS of all expressed genes in WT or KO.

b. Box and whisker plots of H3K27ac or H3K27me3 enrichment in the region ±3 kb with respect to the TSS of genes that are upregulated, downregulated, or stably expressed in KO compared to WT. Whiskers span TSS's within 1.5× the interquartile range. *P* indicates the Wilcoxon signed rank test result with Benjamini-Hochberg correction.

c. Genome browser representation of the levels of H3K4me3, H3K4me1, H3K27ac, and H3K27me3 in WT and KO chromatin. The region marked in yellow is shown zoomed-in below, including genome-mapped mRNA levels in WT and KO.

d. Scatterplot of H3K27ac enrichment per kb in WT versus KO in 10-kb bins spanning the genome. A red line shows the smoothed trend generated by a generalized additive model, while a black line of slope = 1 shows the possible values where H3K27ac enrichment is equal between WT and KO.

Figure 6. Dilution of epigenetic complexity in Dpy30-KO β-cells, cont'd

e. Scatterplot of H3K27me3 enrichment per kb in WT versus KO in 10-kb bins spanning the genome. A red line shows the smoothed trend generated by a generalized additive model, while a black line of slope = 1 shows the possible values where H3K27me3 enrichment is equal between WT and KO.

f. Scatterplot showing spike-in normalized mean gene expression in WT versus KO cells. A red line shows the smoothed trend generated by a generalized additive model, while a black line of slope = 1 shows the possible values where expression is equal between WT and KO.

g. Contour scatterplot showing the H3K4me3 enrichment per kb in WT cells (x-axis) plotted against $\log_2$(-fold-change) of H3K4me3 enrichment per kb in KO versus WT (y-axis) in the regions ±1 kb with respect to the TSS of all expressed genes. The $\log_2$(fold-change) of H3K4me1 enrichment per kb is shown as a colour gradient. The possible values where H3K4me3 is equal between WT and KO are shown by a line at y = 0.

h. Contour scatterplot showing $\log_2$(fold-change) of H3K4me1 enrichment per kb in KO versus WT (x-axis) plotted against $\log_2$(fold-change) of H3K27ac enrichment per kb in KO versus WT (y-axis) in the regions ±1 kb with respect to the TSS of all expressed genes. The $\log_2$(fold-change) of expression of associated genes is shown as a colour gradient. The possible values where H3K4me1 or H3K27ac is equal between WT and KO are shown by lines at x = 0 and y = 0, respectively.

i. Contour scatterplot showing $\log_2$(fold-change) of H3K27ac enrichment per kb in the region ±1kb with respect to the TSS in KO versus WT chromatin (x-axis) plotted against $\log_2$(fold-change) in expression of associated genes (y-axis). The $\log_2$(fold-change) of H3K4me1 enrichment per kb is shown as a colour gradient. The possible values where H3K27ac or expression is equal between WT and KO are shown by lines at x = 0 and y = 0, respectively. Genes with ≥2-fold decrease of H3K27ac enrichment with concomitant ≥2-fold up- or down-regulation are shown in red or green dotted boxes, respectively.

TSS: transcription start site.