1 **Retrospective analysis of enhancer activity and transcriptome history**

2 Ruben Boers[1*], Joachim Boers[1*], Beatrice Tan[1*], Evelyne Wassenaar[1], Erlantz Gonzalez Sanchez[1], Esther

3 Sleddens[1], Yasha Tenhagen[1], Marieke E. van Leeuwen[1], Eskeatnaf Mulugeta[2], Joop Laven[3], Menno

4 Creyghton[1], Willy Baarends[1], Wilfred F. J. van IJcken[4] and Joost Gribnau[1]

5   1) department of Developmental Biology, Erasmus University Medical Center Rotterdam,

6   Oncode Institute,

7   2) department of Cell Biology, Erasmus University Medical Center Rotterdam

8   3) department of Gynaecology, Erasmus University Medical Center Rotterdam

9   4) Erasmus Center for Biomics, Erasmus University Medical Center Rotterdam

10   *) these authors contributed equally to this manuscript

11

12 **Abstract**

13 Cell state changes in development and disease are controlled by gene regulatory networks, the

14 dynamics of which are difficult to track in real time. Here, we utilize an inducible DCM-RNA-

15 polymerase-subunit-b fusion protein, to label active genes and enhancers with a bacterial methylation

16 mark that does not affect gene transcription and is propagated in S-phase. We applied this DCM-time-

17 machine (DCM-TM) technology to study intestinal homeostasis, following enterocyte differentiation

18 back in time, revealing rapid and simultaneous activation of enhancers and nearby genes during

19 intestinal stem cell (ISC) differentiation. We provide new insights in the absorptive-secretory lineage

20 decision in ISC differentiation, and show that ISCs retain a unique chromatin landscape required to

21 maintain ISC identity and delineate future expression of differentiation associated genes. DCM-TM has

22 wide applicability in tracking cell states, providing new insights in the regulatory networks underlying

23 cell state changes in development and differentiation.

24

Introduction

26   Embryonic development and cell differentiation are intricate processes directed by cross talk between

27   cells that affect cell fate decisions and the establishment of cell type specific gene expression programs

28   (Bradner et al., 2017; Lee and Young, 2013; Stadhouders et al., 2019). Lineage tracing studies have

29   been crucial to understand these processes. Initial studies applied light microscopy to follow cleavage

30   divisions, but since then bar-coding, cre-lox and other genetic systems have been utilized to mark

31   precursors or progenitors for readout at later stages of development or differentiation (Alemany et al.,

32   2018). The present advance of the single cell RNA sequencing (scRNA-seq) technology provides a

33   wealth of expression data that can be used to predict developmental trajectories *in silico* and can be

34   linked to genetic lineage tracing techniques to rebuild lineage trees (Bowling et al., 2020; Herman et

35   al., 2018; Schiebinger et al., 2019).

36   Application of these tracing technologies to study the epithelium of the small intestine provided critical

37   insights in homeostasis and regeneration. Turnover of this epithelium happens within 7 days and starts

38   with division of the intestinal stem cell (ISC) located at the bottom of the intestinal crypt (Beumer and

39   Clevers, 2020). ISCs give rise to progenitors that divide moving up the intestinal crypt meanwhile

40   committing to absorptive or secretory lineage. Absorptive progenitors mature into enterocytes,

41   whereas secretory progenitors give rise to Paneth, tuft, entero-endocrine and goblet cells. ISCs are

42   flanked by Paneth cells that provide Wnt, Notch and EGF signals required for self-renewal. Loss of ISC-

43   Paneth cell contact facilitates cell differentiation, aided by BMP signalling that further supports

44   maturation of differentiated cell types. Notch signalling also plays a crucial role in lineage commitment

45   remaining high in absorptive progenitors and is downregulated in secretory progenitors. Lineage

46   tracing and scRNA-seq experiments have been instrumental in identification and characterisation of

47   the crypt based columnar cell as the ISC (Barker et al., 2007), but also showed that several other cell

48   types including entero-endocrine, Paneth and immature enterocytes provide a reservoir of cells that

49 can replenish the ISC niche in injury induced regeneration (Tetteh et al., 2016; Yan et al., 2017; Yu et

50 al., 2018).

51 While these examples highlight the successful application of lineage tracing and scRNA-seq

52 technologies to build relationships between cellular trajectories they cannot keep track of cell state

53 changes following this trajectory and provide limited depth and temporal information with respect to

54 gene expression changes (Baron and van Oudenaarden, 2019). To facilitate whole genome cell state

55 tracing, we developed a system to epigenetically tag transcribed genes to be examined at later stages

56 of development or differentiation. We made use of a fusion between DCM and RNA-polymerase-2-

57 subunit-b to epigenetically label gene bodies of transcribed genes. DCM methylation of $C_{me}C(A/T)GG$

58 penta-nucleotides is a bacterial form of cytosine methylation only detected at very low levels in most

59 mammalian cell types, but is maintained when introduced on transgenes in somatic cells without

60 affecting transgene expression (Clark et al., 1995). Our study demonstrates that DCM-TM marks both

61 active genes as well as enhancers, both of which are occupied by RNA-polymerase-2 and confirms that

62 DCM methylation is propagated to daughter cells with limited effect on gene expression. Thus DCM-

63 TM provides a powerful technology to trace genome wide gene transcription and enhancer activity

64 back in time without relying on in silico assumptions. We applied DCM-TM to study homeostasis in the

65 small intestine, generating gene and enhancer activity maps that trace the ISC state to the enterocyte

66 state. We discovered that gene and enhancer activity changes during enterocyte differentiation are

67 not mediated by heterochromatin changes and show that the H2A variant H2A.Z is preloaded at ISC

68 enhancers that will become activated in the enterocyte. Application of DCM-TM also indicated that

69 commitment of progenitors to the absorptive lineage is a one-way event which does not involve a

70 temporarily dynamic absorptive-secretory intermediate state.

71

72

Results

DCM-Polr2b marks active genes and enhancers.

To develop a gene activity tagging system we fused DCM to the N-terminal end of mouse RNA-polymerase-2 subunit-b (*Polr2b*, Fig 1a), and introduced this DCM-Polr2b fusion gene into the *Col1a1* locus, in an ESC line harbouring the m2rtTA trans-activator expressed from the *Rosa26* locus (Sup Fig 1a-b)(Beard et al., 2006). Addition of doxycycline leads to expression of the fusion protein at levels comparable to endogenous POLR2B, and expression is depleted 24 hours after removal of doxycycline (Sup Fig 1c-d). To detect DCM methylation we developed methylated DNA sequencing (MeD-seq), a technology based on LpnPI mediated digestion of CpG and DCM methylated target sites resulting in 32 base pair fragments that are sequenced (Fig 1a)(Boers et al., 2018). LpnPI recognizes 50% of all methylated CpG di-nucleotides ($C_{me}CG$, $_{me}CGG$, and $G_{me}CGC$), as well as DCM methylated $C_{me}C(A/T)GG$ penta-nucleotides.

Addition of doxycyclin (dox) to DCM-Polr2b:m2rtTA ESCs for five days resulted in a 5-fold induction of DCM methylation genome wide (Fig 1b, Sup Fig 1e, Sup Table 1). DCM methylation clearly increased in gene bodies of genes expressed in ESCs (*Nanog*, *Zfp42*, *ActB*), while no accumulation was observed in genes not expressed in ESCs (*Lgr5*, *Alpi*, Fig 1c-d, Sup Fig 1f). The distribution of DCM sites is clearly different from CpG sites, showing no clear enrichment at the TSS (Sup Fig 1f). In genes with at least 10 DCM sites, uninduced gene body DCM methylation displayed little correlation with gene expression, while after dox induction this correlation became robust (Fig 1e, Sup Fig 1g). Gene-meta-analysis indicated that the DCM methylation profile before dox induction resembled the distribution of CpG methylation that was present in gene bodies of active genes and possibly introduced as an accidental by-product of CpG methylation (Fig 1f, Sup Fig 1k-m) (Arand et al., 2012). After induction, the DCM methylation profile displayed increased DCM methylation at the transcription start sites (TSS), gene body and transcription end sites (TES), with a direct relationship between gene expression and DCM methylation levels (Fig 1f-g).

4

98   Comparison with published ChIP-seq data confirmed accumulation of DCM methylation at the TSS and

99   gene body (H3K36me3) after dox induction. In addition, we found DCM methylation to accumulate at

100   regions marked by enhancer specific modifications or protein recruitment (P300, H3K27Ac and

101   H3K4me1, Fig 1c-d, h, and Sup Fig 2a-b) which is consistent with Pol2 recruitment occurring at

102   enhancers (Kim et al., 2010). Whole genome differentially methylated region (DMR) calling (+dox vs –

103   dox, Mann-Whitney significance test) identified 5,973 regions, displaying significantly increased DCM

104   methylation levels, enriched for enhancer specific histone modifications, DNase sensitivity, and

105   pluripotency factor binding (Sup Fig 2c). DCM methylation levels were significantly elevated in genes

106   located in closest proximity to these intergenic DCM DMRs, and enhancer density was proportional to

107   activity of the closest gene (Sup Fig 2d-e)(Kim et al., 2010). Only 6 genes responded, with a change in

108   gene expression, following induction of the DCM-Polr2b fusion gene (Fig 1i), confirming an earlier

109   report that DCM directed methylation of gene bodies has little effect on gene expression (Clark et al.,

110   1995). These results indicate that DCM-Polr2b is functional and labels active genes and enhancers with

111   minimal effect on gene expression.

112

113   ***DCM methylation propagation in vivo.***

114   To monitor accumulation, maintenance and propagation of DCM methylation *in vivo* we generated

115   DCM-Polr2b transgenic mice. Total epithelium of jejunum was isolated through mechanical shearing

116   from transgenic mice treated with doxycyclin from day 0 through day 18. MeD-seq analysis indicated

117   induced DCM methylation to plateau around day 6 with a >25-fold induction over endogenous DCM

118   methylation levels (Fig 2a-b, Sup Table 1, Sup Fig 3a). Similar as observed in ESCs DCM methylation of

119   TSS, gene body and TES, correlated with gene expression level, Polr2a binding, H3K36me3 deposition

120   (transcribed genes), and H3K27Ac enriched regions (active promoters and enhancers, Fig 2b-e). In

121   addition, DCM gene density distribution was clearly distinct from CpG methylation (Sup Fig 3b). To

122   determine the *in vivo* DCM methylation propagation rate in jejunum, mice were administered dox for

123 7 days, followed by a chase period of 0-7 days after which DCM/CpG methylation ratios were

124 established in triplicate by MeD-seq. With an estimated average cell cycle of 18.3 hours, and taking

125 into account that MeD-seq analysis does not detect hemi-methylated templates, DCM propagation

126 was estimated to be 0.78 per cell division (Fig 2f)(Parker et al., 2017). We found no difference in the

127 propagation rate for specific genomic regions such as gene bodies, exons, introns, CpG islands, or

128 intergenic regions (Sup Fig 3c). Longer chase periods for up to two months after induction, still revealed

129 DCM methylation gene density profiles that were clearly distinct from controls (Sup Fig 3d) suggesting

130 our method is compatible with cell state tracing across longer temporal windows.

131

132 ***Gene activity dynamics in ISC to enterocyte differentiation.***

133 As turnover of the epithelium is very high in the intestine, we tested our DCM-TM technology following

134 the differentiation of ICSs into enterocytes that are eventually shed from the top of the villi. We

135 isolated enterocytes, through sequential purification of small intestinal epithelium, followed by FACS

136 sorting of Epcam+/SLC2A2+(GLUT2) enterocytes (Fig 3a, Sup Fig 4a). Comparison of RNA-seq data of

137 isolated cells with published scRNA-seq data obtained from intestinal epithelium confirmed proper

138 isolation of proximal enterocytes (Sup Fig 4a-c)(Haber et al., 2017). Subsequently, the DCM-Polr2b

139 fusion gene was induced with doxycycline and enterocytes were isolated at different time points across

140 an eight-day window, followed by MeD-seq and normalization for DCM induction efficiency (Sup Fig

141 4d, Sup Table 2). In this setting turn-over of the fusion protein is not required and non-dividing cells do

142 not affect the assay. As time passes, DCM methylation levels of stem cell specific genes are expected

143 to increase relative to the total pool of DCM sequencing reads as their profile will be propagated in the

144 transit amplifying (TA) cells and enterocytes (Fig 3a). In contrast, DCM reads from enterocyte specific

145 genes as well as ubiquitous expressed genes are expected to decline over time as they are shed from

146 the top of the villi. Indeed, stem cell markers *Lgr5*, *Sorbs2*, *Smoc2*, and *Slc12a2*, Wnt target gene *Cd44*,

147 and ephrin receptors *EphB2* and *EphB3* displayed an increase in DCM gene body labelling reaching a

6

148 maximum signal at day 6 (Fig 3b-c, Sup Fig 4e). Ephrin receptor ligands expressed in the villus, *Efnb1*

149 and *Efnb2*, and enterocyte markers *Cbr1*, *Slc2a2*, *Apoa1*, *Alpi* and *Lct* displayed a decrease in DCM

150 methylation (Fig 3b,d, Sup Fig 4e). Ubiquitously expressed genes resembled enterocyte specific genes

151 but with slower kinetics and less dynamic behaviour (Sup Fig 4f). Finally, genes associated with other

152 differentiated cell types or cell types implicated in injury-induced plasticity, including the Paneth,

153 goblet, tuft, entero-endocrine and +4 cell, either did not increase over time, or remained below

154 background levels, indicating no role for these cell types in intestinal homeostasis in the measured

155 time span (Fig 3e-f). One clear exception was *Bmi1*, contrasting with other +4 cell markers in behaviour

156 by resembling other ISC genes. This emphasizes a role for *Bmi1* in ISC homeostasis in line with studies

157 indicating *Bmi1* to be essential for ISC maintenance and intestinal homeostasis (Lopez-Arribillaga et

158 al., 2015).

159 To generate temporal gene activity maps throughout ISC to enterocyte differentiation, genes with a

160 DCM signal significantly higher than background were clustered according to their temporal signal

161 strength using maximum signal day, and average expression of the different gene clusters was

162 displayed on a UMAP that was based on scRNA-seq data from intestinal epithelium (Fig 3g-h)(Haber et

163 al., 2017). This analysis showed that genes with a temporal methylation profile peaking at day 1 (cluster

164 1) are enriched in enterocytes, whereas genes that peak at day 6 (cluster 6) and more prominently at

165 day 8 (cluster 8) are enriched in ISCs (Fig 3h), suggesting that our analysis traces all the way back

166 through intestinal development. Moreover, DCM-TM detects significantly more genes than detected

167 by scRNA-seq, which misses lowly expressed genes due to limited sensitivity (Islam et al., 2014). Using

168 immuno-cytochemistry, we confirmed exclusive expression of cluster 1 specific proteins (SGLT1,

169 SLC43A2) in the villus. Furthermore, cluster 6 and 8 proteins (GNL3 and NUP54) were expressed in the

170 crypt and could be verified as novel gene markers for ISCs (Fig 3i-j, Sup Fig 4g). Genes with a maximum

171 temporal signal at day 2 displayed the highest expression level and were expressed in most single cells

172 across different cell types, indicating that this cluster mostly represents ubiquitously expressed genes

7

173   (Fig 3h, Sup Fig 4h). These results further confirm that the DCM signal is propagated in S-phase and

174   that DCM-TM can be used to retrieve gene activity maps retrospectively.

175

176   ***Enhancer activity dynamics in ISC differentiation.***

177   We next explored whether enhancer activity could be tracked across ISC to enterocyte differentiation

178   using ChIP- and ATAC-Seq data generated in epithelium isolated from the villus (Saxena et al., 2017).

179   We found a clear correlation between H3K27ac, which marks active promoters and enhancers (Calo

180   and Wysocka, 2013), DNA accessibility and DCM methylation (Sup Fig 5a,d)(Saxena et al., 2017). 42%

181   of these H3K27ac peaks were enriched for DCM (with 80% of the high-DCM cluster labelled), whereas

182   the majority of the remaining enhancer peaks lacked sufficient DCM sites for high confidence analysis

183   of their state (Fig 4a, Sup Fig 5b-d). Interestingly, DCM-TM detected a limited number of bivalent

184   enhancers marked by both H3K27ac and H3K9me3 suggesting Rpol2B is in rare cases recruited to

185   poised enhancer states. Nevertheless, most intergenic DCM labelled DMRs are marked by active

186   enhancer chromatin likely reflecting active intestinal enhancers (Sup Fig 5a). DCM methylation was

187   evident at known enhancers near enterocyte (*Fabp1, Cbr*), ubiquitous (*Actb*), and ISC (*Olfm4, Znhit3*)

188   genes peaking at different time points after the start of doxycyclin treatment (Fig 4b, Sup Fig 6b, Sup

189   Table 3) (Kaaij et al., 2013). We annotated 51,779 intergenic DCM DMRs (> 1kb from TSS) between -

190   dox and +dox (all stages). Clustering of these intergenic DMRs based on their temporal peak values

191   highlighted the dynamic behaviour of enhancer activity during cell state specification (Fig 4c). As

192   expected, enrichment of H3K27ac at DMRs, was more pronounced at early than later time points as

193   these DMRs reflect enterocyte specific and ubiquitous enhancers that are active in villi (Fig 4c).

194   Interestingly, comparison of the different enhancer clusters based on peak day with ATAC-seq data

195   indicates that enterocyte enhancers are accessible, whereas enhancers active at earlier stages of

196   differentiation lose accessibility in enterocytes. We did not observe this dynamic behaviour in

197   accessibility for TSSs, (Sup Fig 6b). Density analysis of enhancers of the different clusters around the

8

198   different gene clusters showed a coordination in peak days of enhancers and nearby genes in the

199   enterocyte differentiation process (Fig 4e, Sup Fig 6c). Genes displaying a maximum at day 2 display

200   different enhancer kinetics as this cluster consists of both ubiquitously expressed and stage specific

201   genes (Sup Fig 6d). ChromVAR motif analysis on enhancer regions confirmed enrichment of motifs for

202   ISC specific transcription factors (TFs) peaking at day 8 (TCF4 and TEAD1). Enterocyte differentiation

203   associated TFs, including ELF3, KLF5 as well as HNFA/G, known to have a crucial role in enterocyte

204   differentiation, peak at early time points (Fig 4f, Sup Fig 6e, Sup Table 4)(Chen et al., 2019; Ng et al.,

205   2002). In addition, by selecting TFs displaying coordinated timing of gene body and enhancer DCM

206   labelling (Pearson r>0.3) we were able to identify *Mef2b* and *Tgif1* as potentially novel candidate TFs

207   in ISC homeostasis and enterocyte differentiation similar to their proposed roles in other systems (Fig

208   4g)(Ito et al., 2017; Lee et al., 2015). In addition, reverse-coordinated peak timing (Pearson <-0.3) with

209   maximum DCM gene body labelling in ISCs and maximum motif labelling in enterocytes identified

210   known and putative new repressors (*Atf7*, *Glis2*, and *Mixl1)* of the ISC state (Liu et al., 2019). These

211   results show that DCM-TM can be utilized to detect enhancer activity and relate these to underlying

212   transcription factor dynamics.

213

214   ***ISC gene expression dynamics is independent of polycomb mediated repression.***

215   DCM-TM revealed *Bmi1,* a member of non-canonical polycomb complex PRC1.4*,* to be expressed in

216   the ISC (Figure 3c and f). The family of PRC1 complexes is formed around a core of RING1A/B that

217   interacts with PCGF partner proteins to form canonical PRC1 (PCGF2) and non-canonical PRC1

218   complexes (PCGF1, PCGF3, BMI, PCGF5, PCGF6, Fig 5a) involved in the formation of facultative

219   heterochromatin through deposition of H2A119ub (Aranda et al., 2015). In the intestine studies

220   involving loss of PRC1 members *Bmi1* or *Ring1b* suggested that PRC1 is required for ISC maintenance

221   by preventing the ectopic expression of non-lineage genes through active repression (Chiacchiera et

222   al., 2016a; Lopez-Arribillaga et al., 2015). Indeed DCM-TM indicated several other members of PRC1

223   complexes including *Ring1b* and *Cbx3* to show maximum gene body DCM labelling on days 6 and 8,

224   and enrichment of RING1B and CBX3 was confirmed in CD44 positive ISCs (Fig 5b, Sup Fig 7a). In

225   contrast, DCM gene body labelling of canonical PRC1 members *Cbx2,4,6,7* and 8 was below background

226   levels, indicating that involvement of PRC1 in ISC maintenance is mediated through ncPRC1 complexes

227   (Fig 5a)(Blackledge et al., 2014). To investigate the role of PRC1 and PRC2 in promoter and enhancer

228   regulation during ISC to enterocyte differentiation, we examined enrichment of H2A119ub or

229   H3K27me3, catalysed by PRC1 and PRC2 respectively, using ISC, crypt and enterocyte ChIP-seq data

230   sets (Chiacchiera et al., 2016a; Ferrari et al., 2021). This analysis revealed a lack of enrichment of both

231   H2Aub119 and H3K27me3 at enhancers at any stage of ISC differentiation (Fig 5c). Similarly, these

232   modifications were excluded from TSSs at all stages of ISC differentiation (Sup Fig 7b).

233   Several *Hp1* associated factors involved in maintenance of constitutive heterochromatin also belonged

234   to clusters 6 and 8 (Sup Fig 7c). Nevertheless, only moderate enrichment of its target H3K9me3 (and

235   no enrichment of H3K9me2) was detected in the crypt (Sup Fig 7d). Similar to PRC1/2 mediated histone

236   modifications, analysis of published H3K9me3 ChIP-seq data did not reveal enrichment of H3K9me3

237   on enhancers and genes active at different stages of the ISC to enterocyte differentiation (Fig 5c, Sup

238   Fig 7b). These results suggest that chromatin modifying complexes PRC1, PRC2 and HP1 complex

239   members are enriched in the ISC but challenge the notion that they regulate differentiation associated

240   genes through active repression of promoter and enhancer sequences.

241

242   ***H2A.Z is loaded on enterocyte specific enhancers in ICS.***

243   The absence of heterochromatin mediated regulation of intestinal specific enhancers and promoters

244   suggests that activation signals and TF networks may play a dominant role during ISC differentiation.

245   Histone variant H2A.Z has been implicated in lineage specific gene activation (Giaimo et al., 2019).

246   H2A.Z incorporation is mediated by *Srcap* and *P400*, which both peak on days 6 and 8, and is preceded

247   by H2A acetylation, which showed marked enrichment in ISCs whereas HA2.Z is more uniformly

248   distributed in the crypts and villi (Fig 6a-c). Using ISC and enterocyte specific ChIP-seq data we found

249   that H2A.Z accumulation at ISC and enterocyte specific TSSs (Sup Fig 8a)(Kazakevych et al., 2017).

250   Interestingly, enhancers that become active in enterocytes peaking on day 1 or day 2 showed

251   enrichment of H2A.Z both in ISCs and enterocytes (Fig 6d). No H2A.Z enrichment was observed at ISC

252   specific enhancers peaking on day 6 or day 8, indicating that H2A.Z pre-marks enhancers in ISCs that

253   are poised for activation in enterocytes, and suggest a distinct role for H2A.Z in enhancer and gene

254   activity regulation. HOMER motif enrichment analysis of H2A.Z enhancer peaks present in enterocytes

255   revealed enrichment of TF binding sites for factors involved in Notch signalling (RBPJ and SPDEF) and

256   several targets of the EGF signal transduction pathway including ELK1, ELK4, cMYC, and cJun,

257   suggesting a role for these pathways in H2A.Z recruitment (Sup Fig 8b). These findings emphasize the

258   presence of an ISC-specific chromatin landscape to maintain ISC stemness and lineage identity, and to

259   prepare and delineate enterocyte specific enhancers and genes for future activation upon cell

260   differentiation.

261

262   ***The absorptive-secretory switch in ISC differentiation***

263   To better understand cell state changes in enterocyte differentiation we performed KEGG pathway

264   analysis on DCM-labelled genes and found enrichment of pathways including absorption and TGF-beta

265   signalling at early time points. Pathways including cell cycle, Wnt, EGF and Notch signalling showed

266   enrichment at later time points, consistent with their lineage history (Fig 7a, Sup Table 5). Notch

267   signalling controls the absorptive versus secretory cell fate decision dictating repression of *Atoh1* in

268   the ISC and enterocyte progenitors through action of *Hes1*,*3* and *5* (Fig 7b). Loss of contact of

269   proliferating ICSs (Notch+) with Paneth cells expressing the Notch ligand *Dll1* leads to downregulation

270   of *Notch1* and subsequent upregulation of *Atoh1* in future secretory cells (Beumer and Clevers, 2020).

271   Notch mediated repression of *Atoh1* and its key target genes *Spdef* and *Gfi1* could involve a gradual

272   transition from a bi-potential progenitor to one cell state or may involve a binary switch towards the

11

273   absorptive or secretory lineage. Single cell RNA-seq data demonstrated predominant expression of

274   *Atoh1*, *Spdef* and *Gfi1* in secretory cell types, but also revealed several cells that appear committed to

275   the absorptive lineage to express at least one of these genes (Fig 7c). In addition, a few cells express

276   both *Notch* and *Atoh1* or don't express *Notch* and *Atoh1* at all, making it difficult to discern how the

277   absorptive-secretory switch is mediated (Sup Fig 8c). Examination of DCM-TM data indicated that

278   *Notch1* as well as Notch target genes including *Hes1, Hes3* and *Hes5* are expressed throughout

279   enterocyte differentiation (Fig 7d,e). In contrast, DCM methylation in gene bodies of *Spdef* and *Gfi1*

280   never got above background levels (Mann-Whitney significance test $p > 0.05$), suggesting their

281   activation results in irreversible commitment towards the secretory lineage. These results

282   demonstrate that, unlike the transition from the ISC into the absorptive state, the switch towards a

283   secretory state represents a binary one directional switch rather than a gradual transition of,

284   transcriptional programs (Fig 7f,g).

285

286

Discussion

287

288 To facilitate whole genome cell state tracing, we developed a system to epigenetically tag transcribed

289 genes to be examined at later stages of development or differentiation in vivo. We applied this DCM-

290 TM technology to perform whole transcriptome and enhancer activity lineage tracing of specific cell

291 types, and demonstrated the possibility to establish TFs and signal transduction roadmaps through

292 isolation of a differentiated cell type without the need to isolate progenitor or stem cells and without

293 the need to infer connectivity in silico. We identified novel marker genes for different cell states and

294 provide new insights into the transcriptional dynamics during cellular differentiation in the mouse

295 intestine.

296 Key parameters for an epigenetic lineage tracing system to work are the application of an epigenetic

297 tag that is normally absent in mammalian cells, is maintained and propagated upon DNA synthesis,

298 and does not interfere with gene expression. Our study showed that DCM methylation approaches

299 these criteria simultaneously. DCM methylation is only present at low levels in wildtype ES cells and

300 intestinal epithelium (2-3%), while only a 5-fold and 25-fold induction in ES cells and intestinal

301 epithelium, respectively, is sufficient to reliably identify active genes and enhancers and trace their

302 activity back in time. In contrast to other forms of bacterial methylation which are not propagated such

303 as DAM, propagation of DCM methylation is 78% in the intestinal epithelium (van Steensel and

304 Henikoff, 2000). This is lower than previously described, but sufficient to detect gene body and

305 enhancer labelling over at least 10 cell divisions (Clark et al., 1995). Lastly, we found that only a limited

306 number of genes is affected by induction of the fusion protein. This lack of interference with

307 transcription might be related to the fact that gene bodies of active genes already accumulate CpG

308 methylation, thought to repel intragenic initiation of RNA pol2 (Neri et al., 2017). In addition, as the

309 DCM motif is found much less frequently in genes it may thus have only limited effect on transcription.

310 In this study we applied DCM-TM to understand the mechanism directing the absorptive-secretory

311 switch in the intestinal epithelium involving the Notch signalling pathway. Activation of *Notch* and its

312 downstream *Hes* family targets is mediated by cell-cell contact through direct contact of Notch and its

313 ligand Dll. Notch signalling is required for maintenance of the ISC where Notch ligands are expressed

314 by Paneth cells, but also during ISC differentiation to consolidate the absorptive lineage (VanDussen

315 et al., 2012). In the secretory lineage Notch is downregulated resulting in de-repression of *Atoh1* and

316 its downstream targets *Gfi1* and *Spdef*. Our DCM-TM data demonstrates that Notch signalling remains

317 active throughout enterocyte differentiation and that *Spdef* and *Gfi1* are never activated. This indicates

318 that the absorptive-secretory lineage fate decision involves a committed rather than a temporarily

319 dynamic absorptive-secretory intermediate state. Nevertheless, *Atoh1* is active in ISCs and cells that

320 commit to the absorptive lineage, probably as a consequence of fluctuating Notch activity and to be

321 able to quickly respond when *Notch* levels decrease below a specific threshold for locking in the

322 secretory state. This finding explains why *Atoh1* lineage tracing studies revealed *Atoh1* positive cells to

323 contribute to the ISC pool (Ishibashi et al., 2018), and also indicates that this *Atoh1* expression level is

324 too low to activate *Spdef* and *Gfi1* and the downstream secretory program.

325 Our study and studies of others indicate that the role for complexes in heterochromatin formation and

326 maintenance in ISC differentiation associated gene expression dynamics is very limited. Previous

327 studies involving loss of PRC1 members *Bmi1* or *Ring1b* suggested that PRC1 is required for ISC

328 maintenance by preventing ectopic expression of non-lineage genes (Chiacchiera et al., 2016a; Lopez-

329 Arribillaga et al., 2015). Application of DCM-TM revealed that PRC1 mediated gene repression is

330 mediated by non-canonical PRC1 complexes that contain RYBP which catalyses H2A ubiquitination

331 independent of PRC2. This is in line with the observation that loss of PRC2 and H3K27me3 does not

332 affect H2A119ub in the intestine (Chiacchiera et al., 2016b) . The present study also revealed a static

333 landscape of repressive chromatin modifications H2A119ub, H3K27me3 and H3K9me3 in intestinal

334 homeostasis. Similarly, DNA methylation changes were found to be very limited between different

335 epithelial cell types (Kaaij et al., 2013), suggesting that the main role of facultative and constitutive

336 heterochromatin in ISCs is dictating repression of non-lineage genes. Therefore, activation and

337 repression of enhancers and genes in intestinal homeostasis appears to be regulated by other

338   epigenetic mechanisms and transcription factor networks. We found histone variant H2A.Z to be

339   loaded on enhancers in ISCs that are destined to become activated in enterocytes, indicating that the

340   ISC dictates and limits enhancer activity in its decedents through H2A.Z recruitment to enhancers. The

341   clear enrichment of motifs of EGF and Notch regulated transcription factors in these enhancers make

342   both signalling pathways the likely candidate signal for H2A.Z recruitment.

343   Several cell state tracing technologies detecting the history of gene expression have been described

344   before. These include CRISPR spacer mediated recording of DNA or RNA to monitor complex cellular

345   behaviour retrospectively, as well as smFISH based detection of Crispr/Cas mediated targeted

346   disruption of expressed recording elements (Frieda et al., 2017; Schmidt et al., 2018). Unfortunately,

347   all these technologies are restricted by a limited number of genes that can be recorded. The recent

348   developments in scRNA-seq provides alternative means to detect cell states and gene expression

349   changes in relation to developmental trajectories. Temporal changes in abundance of spliced and un-

350   spliced gene products (La Manno et al., 2018), and minimum spanning tree analysis (Monocle) have

351   been applied to predict cell trajectories in silico (Trapnell et al., 2014), but these analyses are limited

352   by the temporal resolution and the number of genes detected, and therefore often fail to detect the

353   changes in gene expression from one cell state to the next, and are more difficult to apply along

354   developmental trajectories. The present DCM-TM technology circumvents these issues providing a

355   genome wide picture of gene and enhancer activity at any timepoint during development or

356   differentiation. The DCM-TM transgene can be combined with conventional lineage tracing

357   technologies to fine map cell fate decisions, discriminate between lineage paths and keep track of

358   network changes. In addition, DCM-TM can be applied to follow embryonic development and tissue

359   regeneration, providing a powerful system to identify temporal maps of transcription factor networks

360   and signal transduction pathways that can be used to improve stem cell expansion and cell

361   differentiation models.

367 **Author contributions**

368 R.G.B., J.B.B, B.T, M.C. and J.G. conceived and performed the experiments and data analysis. E.W., E.S,

369 W.IJ., E.M. and Y.H. assisted in immunocytochemistry, FACS analysis, sequence analysis and modelling.

370 J.L. E.G.S. and W.B. aided in development of DCM-TM and interpreting the results, and M.L. helped

371 with motif analysis. All authors discussed the results and contributed to the final manuscript.
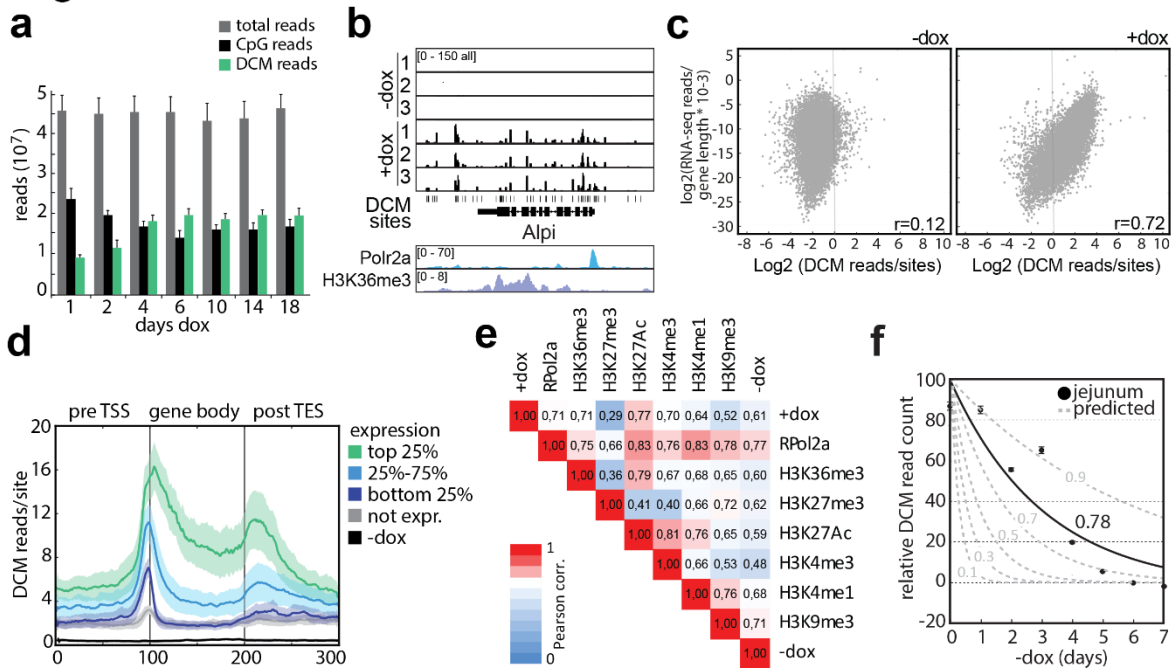
Figure 1
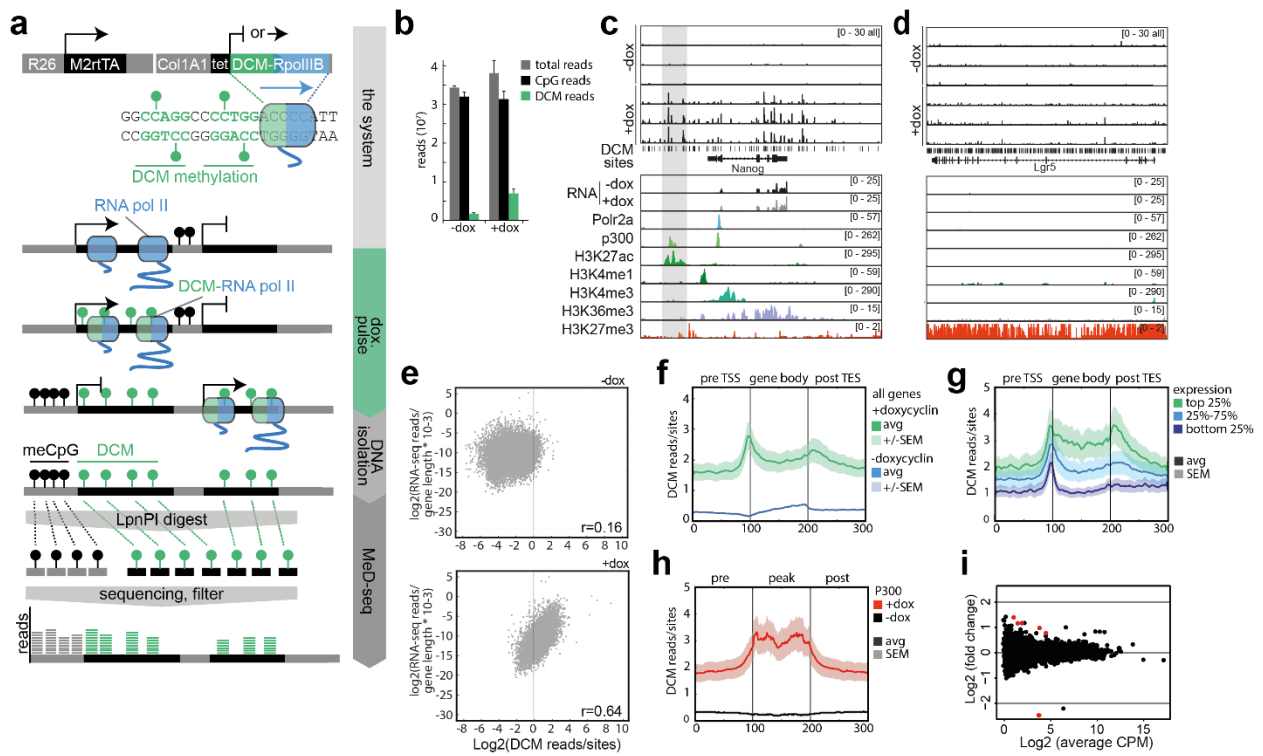
Figure 1. *The DCM-time-machine in embryonic stem cells.*

(a) Overview of the DCM-TM and MeD-seq pipelines. (b) Induction of DCM labelling measured 5 days after start of dox treatment. (c,d) Genome browser view of DCM specific MeD-seq reads (n=3), RNA-seq (+/- dox) and ChIP-seq tracks (ENCODE) in the *Nanog* (c, enhancer indicated in grey) and *Lgr5* (d) loci. (e) Scatter plot displaying RNA-seq gene expression level in relation to DCM read count per gene before (top panel) and after dox induction (bottom panel). (f) Gene meta-analysis showing binned distribution of DCM reads overlapping the gene body before (blue) and after (green) 5 days of dox treatment. The average (avg) and standard error of the mean (SEM) are depicted with a darker line and lighter region, respectively. (g) Gene meta-analysis showing distribution of DCM reads of expressed genes split in three clusters based on expression (top 25%, 25%-75% and bottom 25%) after 5 days of dox treatment. (h) DCM read meta-analysis showing binned distribution of DCM read counts over p300 positive genomic regions and 1kb proximal and distal flanking regions. (i) RNA-seq analysis comparing average gene expression values before and 5 days after dox induction (genes indicated in red show significant expression change).

17

Figure 2

387

Figure 2. *DCM labelling and propagation in the small intestine.*

(a) Genome wide DCM and CpG methylation at different time points after start of dox induction. (b) Genome browser view of *Alpi* locus showing DCM specific MeD-seq reads before and after 1 day dox treatment. Polr2a and H3K36me3 ChIP-seq tracks from ENCODE are shown below. (c) Scatter plot displaying RNA-seq gene expression level in relation to DCM read count per gene before (left panel) and after 10 days of dox induction (right panel) in epithelium of jejunum. (d) Gene meta-analysis showing distribution of DCM reads in the top 25%, 25%-75% and bottom 25% expressed genes after 3 days of dox treatment. (e) Pearson correlation analysis comparing DCM and ChIP-seq read count distribution. (f) Relative DCM read count measurement at different time points after withdrawal of dox and the estimated propagation rate per cell division in jejunum. The dotted grey lines show the estimated curves based on several different propagation rates, whereas the propagation rate fitted to the DCM data is plotted in black.

Figure 3
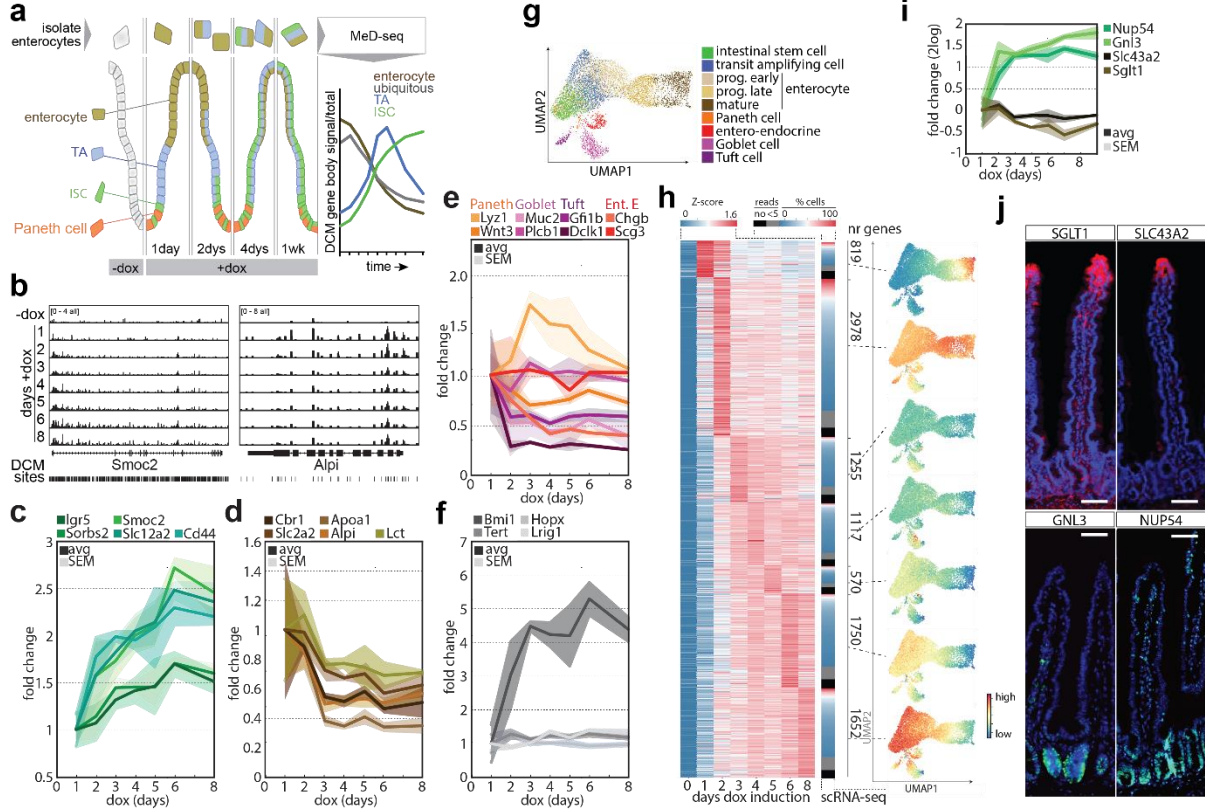
Figure 3. *DCM-Rpol2b labelling reveals gene activity maps from ISC to enterocyte.*

(a) Overview of experimental procedure: mice are labelled with dox and sacrificed at different time points to isolate Epcam+/SLC2A2+ positive enterocytes that are subjected to MeD-seq. ISC, TA, enterocyte and ubiquitously expressed genes are expected to display different dynamic behaviour in time. (b) Genome browser view of the average normalized MeD-seq DCM reads (n=3) in the *Lgr5* and *Alpi* loci at different time points after start of dox treatment. (c-f) DCM labelling (fold change in DCM reads relative to total and normalized to T=1d) of ISC (c), enterocyte (d), Paneth, Goblet, Tuft, Entero-endocrine (e) and +4 cell (f) specific genes. (g) UMAP of jejunum scRNA-seq data showing clusters annotated as specific cell types. (h) DCM labelling of all significantly labelled genes (negative (T=0) samples compared to all days after start of dox treatment) clustered according to the maximum DCM signal, capture of clustered genes by scRNA-seq (for each gene with >5 reads percentage of cells with signal is indicated), and average expression of clustered genes in UMAP shown in (g). (i-j) DCM labelling

19

413   (i) and validation by immuno-cytochemistry (j) of SGLT1, SLC43A2, GNL3 and NUP54 expression (FITC,

414   DNA is DAPI stained, scale bar: 50μm).

415
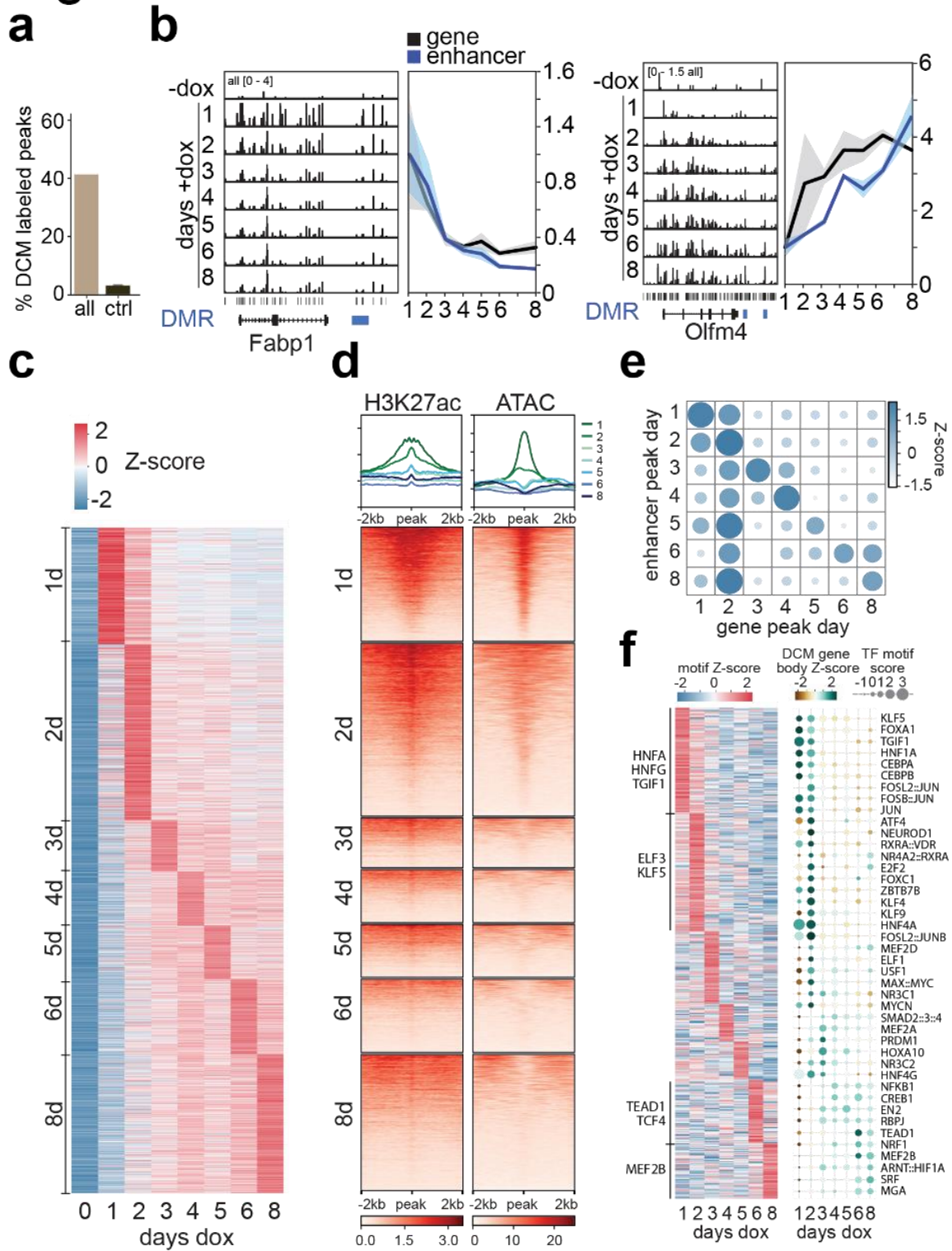
Figure 4

419   **Figure 4.** *Temporal changes in TF and enhancer activity from ISC to enterocyte.*

420   (a) Percentage of H3K27ac peaks that are labelled by DCM (i.e. peaks with ≥ 1 significant DCM site <750

421   bp from peak). Random control based on 100 sets of reshuffled H3K27ac peaks is added to show

422   expected random overlap (mean ± SD). (b) Left panels display genome browser view showing DCM

423   labelling of enterocyte specific (*Fabp1*) and ISC specific (*Olfm4*) genes with nearby enhancers (marked

424   in blue) showing coordinated behaviour in time. Right panels show the average profiles over time for

425   each gene and the average of the closest significant DCM sites. (c) Heatmap of DCM labelling of

426   enhancers (Z-scores of mean normalized DCM reads). (d) Heatmap showing H3K27ac ChIP-seq and

427   ATAC-seq overlap with the regions around enhancer DMRs peaking at different days of dox induction.

428   Each profile plot has the same y-axis range as its corresponding heatmap. (e) Correlation between gene

429   peak day and peak day of close-by enhancers (z-score of proportion of enhancers per day). (f) Heatmap

430   showing TF motif dynamics observed in intergenic DCM DMRs in time (left) and combined analysis of

431   motif enrichment and DCM gene body labelling dynamics of TFs displaying a positive correlation in

432   time (right).

433

22

434
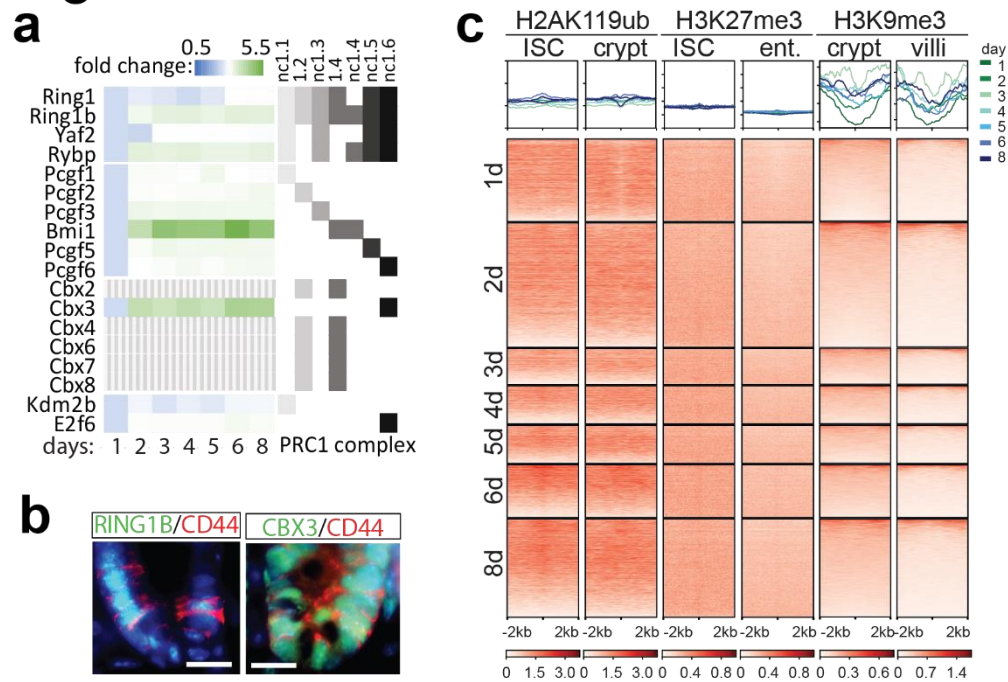
Figure 5. *Lack of heterochromatin dynamics at intestinal enhancers.*

(a) Temporal behaviour of DCM methylation (normalized to t=1 day) of members of different PRC1 complexes (genes indicated in dashed grey do not accumulate DCM signal above background). (b) Immunofluorescence detection of RING1B (FITC), CBX3 (FITC) and CD44 (Texas Red) in the intestinal crypt (DNA=DAPI, scale bar: 16μm). (c) Heatmap showing H2A119ub, H3K27me3 and H3K9me3 ChIP-seq overlap in indicated cell types with the regions around enhancer DMRs peaking at different days of dox induction. Each profile plot has the same y-axis range as its corresponding heatmap.

442

Figure 6

443
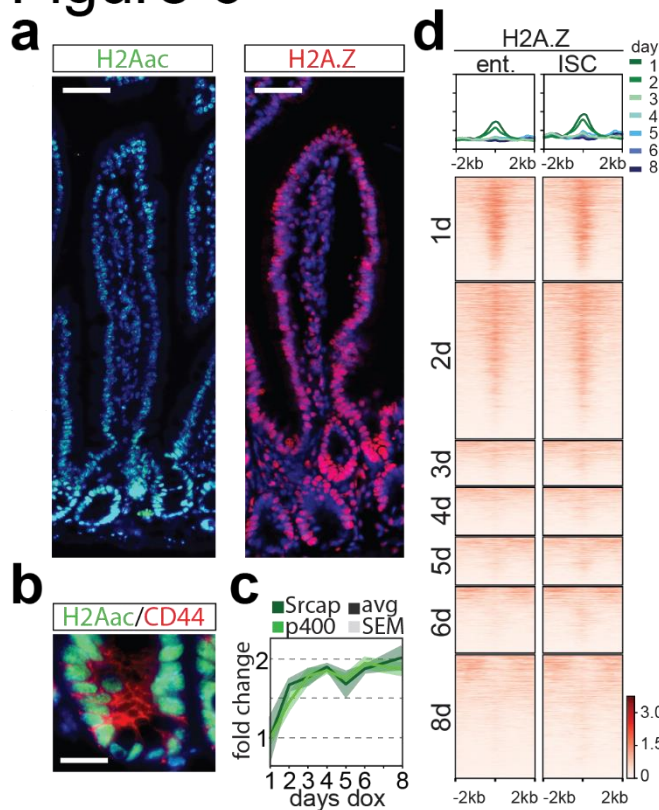
Figure 6 *H2A.Z is recruited to enterocyte specific enhancers in ISCs.*

(a) Immunocytochemistry detecting H2Aac (FITC) and H2A.Z (Texas Red) (DNA=DAPI, scale bar: 50μm), and (b) combined detection of H2Aac (FITC) and CD44 (Texas Red) in the intestinal crypt (DNA=DAPI, scale bar: 16μm). (c) DCM labelling (fold change in DCM reads relative to total and normalized to T=1d) of *Srcap* and *P400*. (d) Heatmap showing H2A.Z ChIP-seq overlap in enterocytes and ISC with the regions around enhancer DMRs peaking at different days of dox induction. Enhancers were ordered according to H3K27ac and ATAC-seq enrichment (fig 4d). Each profile plot has the same y-axis range as its corresponding heatmap.

452

Figure 7
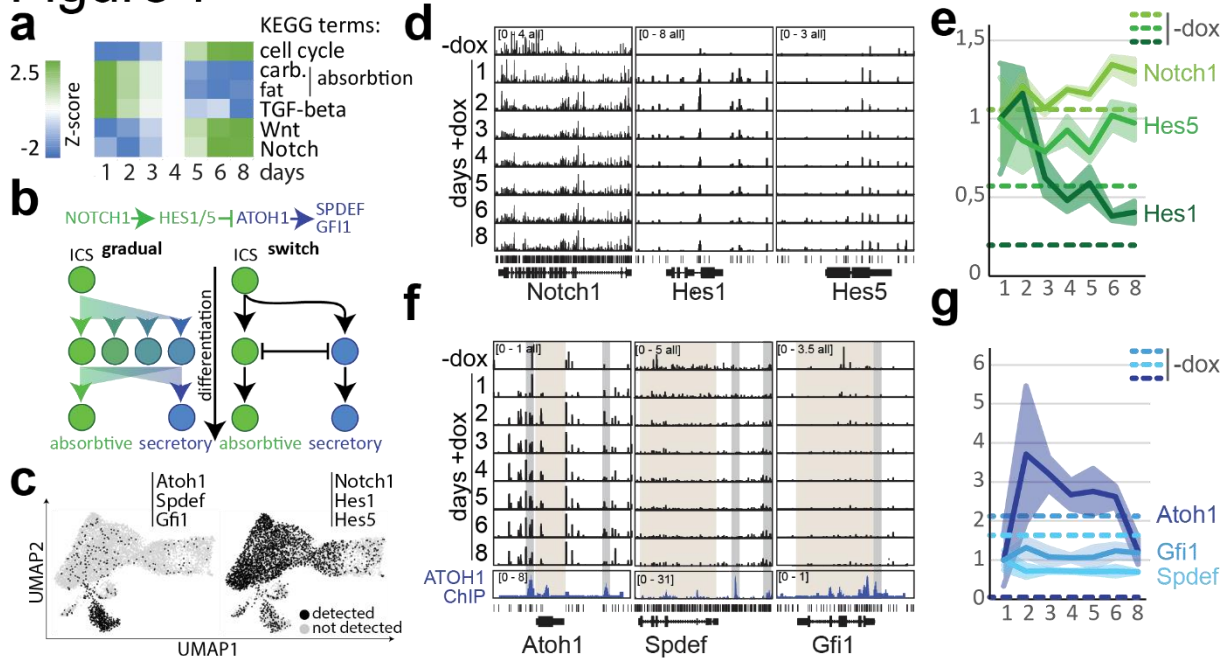
Figure 7. *Notch signalling in absorptive versus secretory cell fate decision.*

(a) KEGG pathway enrichment analysis at different time points after start of dox treatment. (b) Notch signalling pathway in intestinal stem cell differentiation towards absorptive and secretory lineage; two possible mechanisms involving a gradual change or an on/off switch in Notch signalling are shown. (c) Cells expressing at least one of following three genes; *Atoh1*, *Spdef*, and *Gfi1* (left) or *Notch1*, *Hes1* and *Hes5* (right) in scRNA-seq shown in UMAP of Figure 4g. (d,e) Genome browser view of MeD-seq DCM reads in *Notch1* and its target genes *Hes1* and *Hes5* (d), and quantification of DCM signal normalized to day 1 (dashed lines represent -dox signal per gene). (f,g) As in (e,f) but now for *Atoh1*, *Spdef* and *Gfi1.* Bottom tracks in (f) show ATOH1 ChIP-seq signal from *Atoh1* GFP+ cells from the small intestine (ATOH1 targeted regulatory regions are indicated in grey and gene bodies in brown).
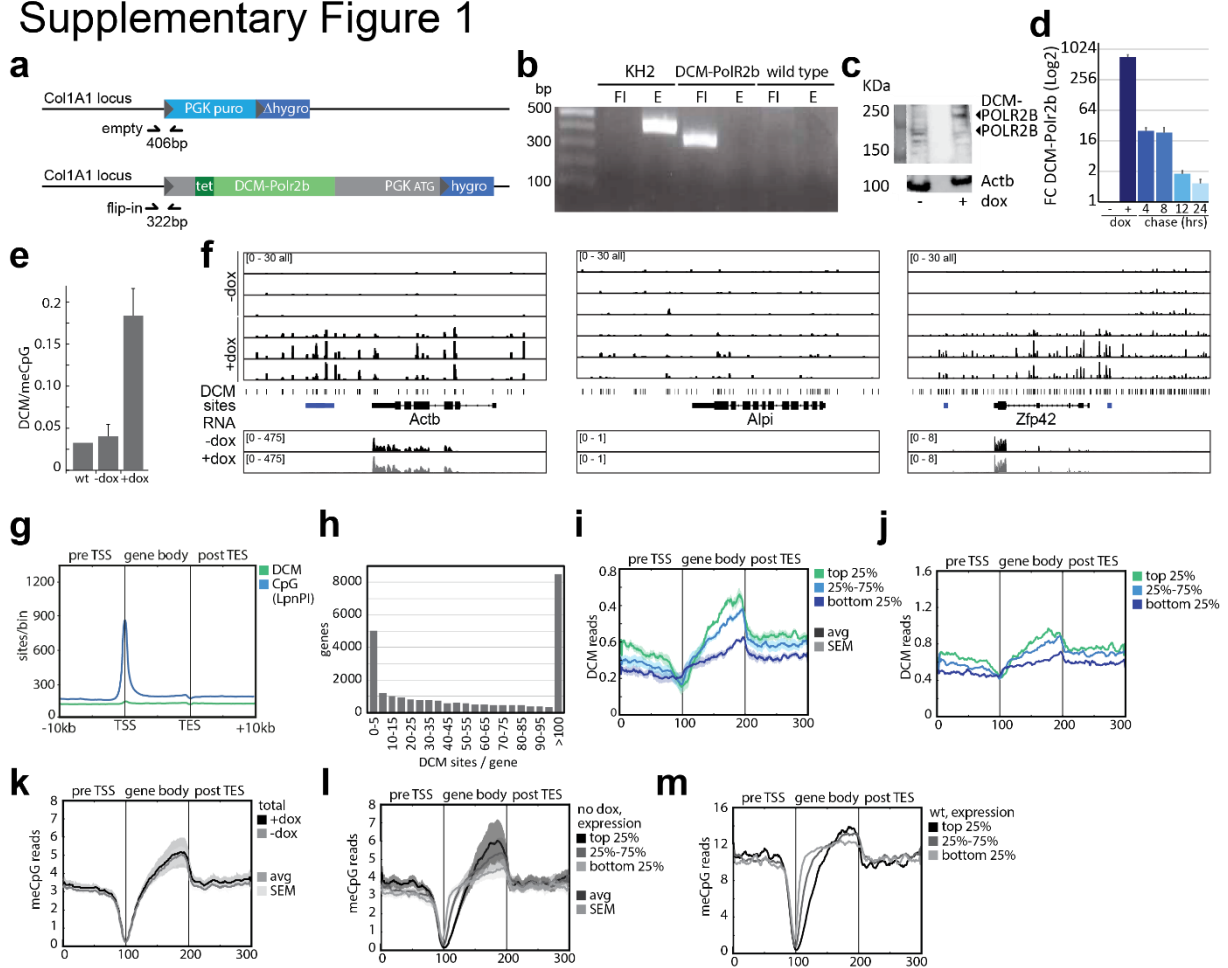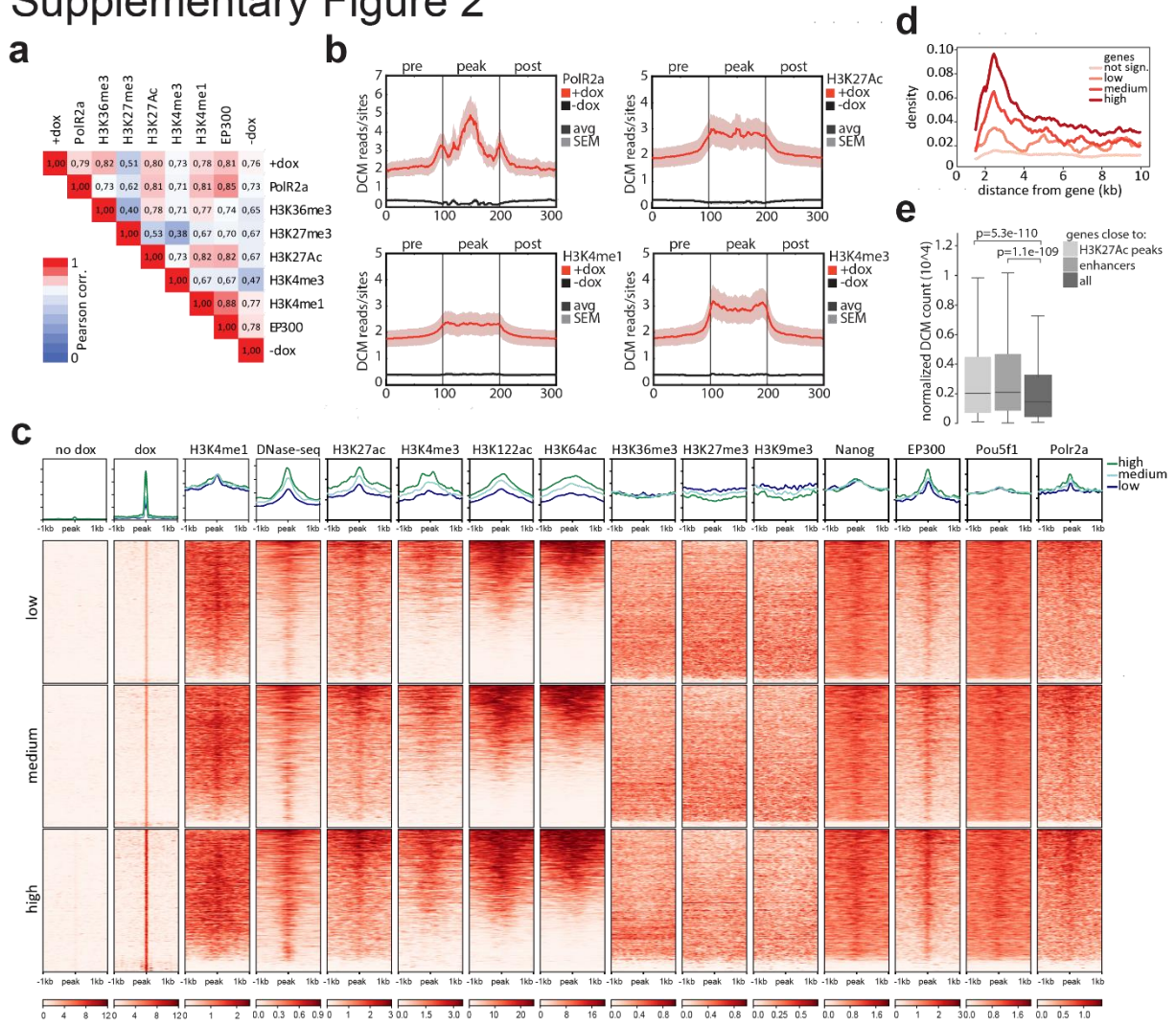
Supplementary Figure 1

464

Supplementary Figure 1

(a) The DCM-Polr2b fusion gene was introduced in the Col1A1 locus by Flipase mediated insertion. (b) PCR with primers for flip-in (FI) and empty (E) indicated in (a) verifying proper integration of the transgene. (c) Western blotting analysis detecting POLR2B, DCM-POLR2B and ACTB in uninduced and induced ES cells. (d) qRT-PCR analysis detecting DCM-POLR2B transcript in ES cells at different timepoints (hours) after removal of dox. (e) The ratio of DCM/CpG methylation genome wide in wild type (WT), DCM- Polr2b -dox and +dox ESCs. (f) Genome browser view of DCM specific MeD-seq reads in the *Actb*, *Alpi* and *Zfp42* loci (intergenic DMRs are indicated in blue). (g) Gene meta-analysis showing binned distribution of DCM and CpG sites. (h) Histogram showing distribution of DCM sites per gene. (i) Gene meta-analysis showing binned distribution of DCM reads in the top 25%, 25%-75% and bottom 25% expressed genes in DCM- Polr2b ESCs in the absence of dox. (j) As in (i) but now for wild type ESCs

476   (n=1). (k) Gene meta-analysis showing binned distribution of all CpG reads before and after dox

477   treatment (5 days). (l) Gene meta-analysis showing binned distribution of CpG reads in the top 25%,

478   25%-75% and bottom 25% expressed genes in DCM- Polr2b ESCs in the absence of dox. (m) As in (l)

479   but now for wild type ESCs (n=1).

Supplementary Figure 2

Supplementary Figure 2

(a) Pearson correlation analysis comparing DCM and ChIP-seq read count distribution. (b) Enhancer meta-analysis showing binned distribution of DCM read counts over Polr2a, H3K27Ac, H3K4me1 and H3K4me3 positive genomic regions and 1kb proximal and distal flanking regions. (c) Heatmap showing ChIP-seq overlap with the regions around enhancer DMRs. DMRs are split in three equal clusters based on the normalized number of reads for +dox. Each profile plot has the same y-axis range as its corresponding heatmap. (d) Density plot showing the number of enhancer DMRs in the 10kb region around genes that were either not significantly labelled by DCM or genes split in three equal clusters based on fold change between +dox and −dox. (e) Normalized DCM count of +dox samples for genes

490   close to the enhancer DMRs, close to H3K27ac peaks and all genes. P-values were calculated using a

491   one-sided Wilcoxon rank-sum test.

492

29
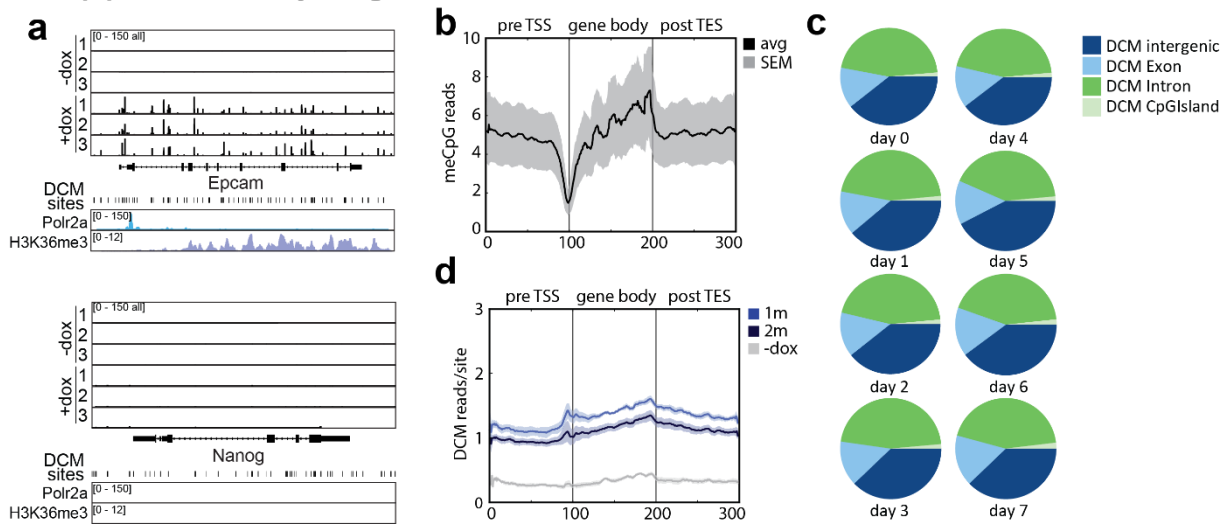
Supplementary Figure 3

Supplementary Figure 3

(a) Genome browser view of *Epcam* and *Nanog* loci showing DCM specific MeD-seq reads before and after 1 day dox treatment. POLR2a and H3K36me3 ChIP-seq tracks are shown below. (b) Gene meta-analysis showing binned distribution of all CpG reads in the absence of dox. (c) Relative distribution of DCM reads in intergenic, exonic, intronic and CpG island sequences at indicated days after removal of doxycycline. (d) Gene meta-analysis showing binned distribution of DCM reads in -dox control mice and mice treated for 7 days with dox followed by a 1 and 2 month chase prior to tissue harvesting.
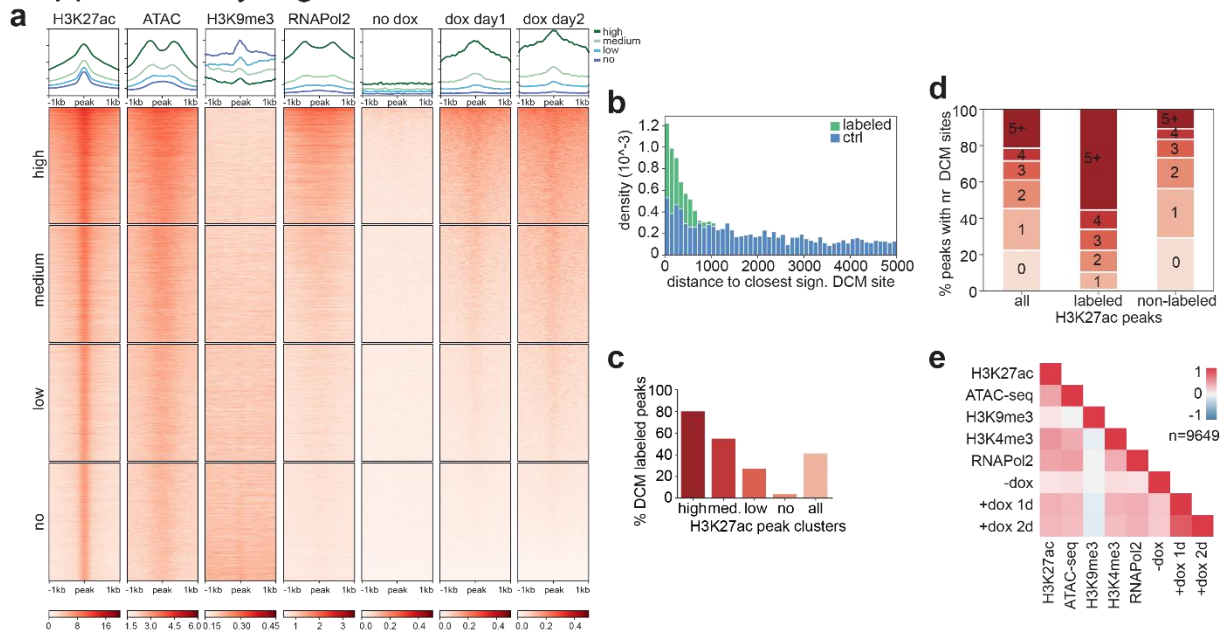
Supplementary Figure 4
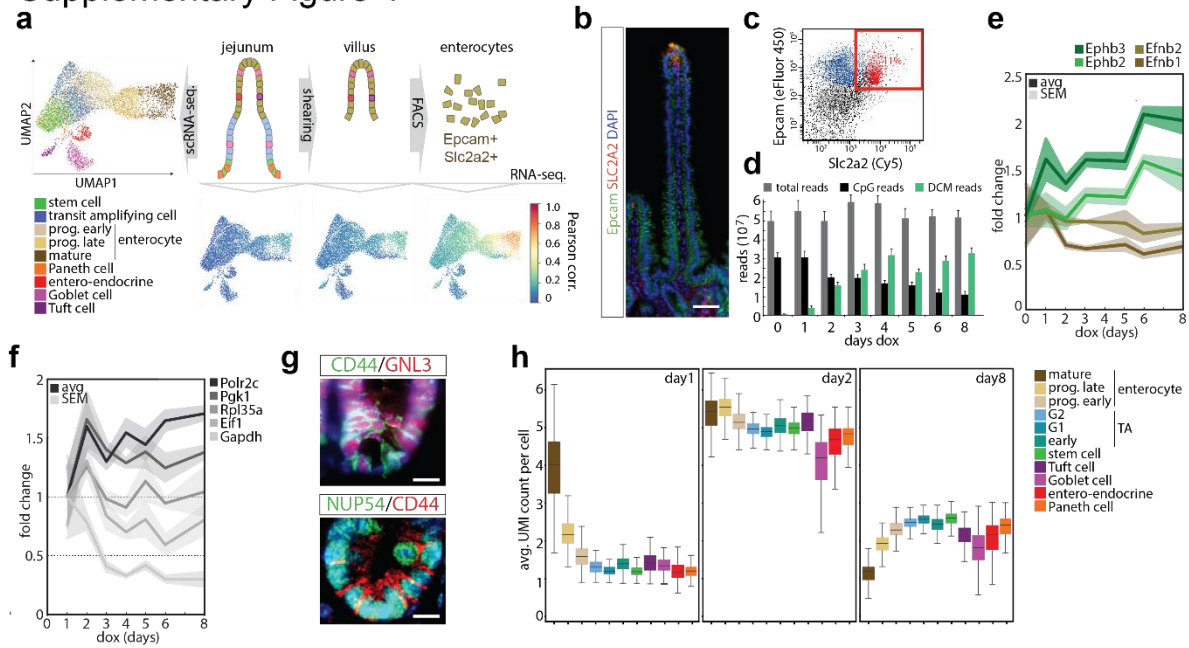
Supplementary Figure 4

(a) Overview of the experimental procedure to isolate SLC2A2 expressing enterocytes. Villi were isolated from intestinal epithelium of jejunum followed by FACS isolation of SLC2A2 positive cells. Left panel shows UMAP of scRNA-seq data colored according to annotation as specific cell types, bottom panels show Pearson correlation analysis of bulk RNA-seq analysis on total epithelial, villi and SLC2A2 positive fractions with scRNA-seq data. (b) Immuno-cytochemistry detecting Epcam (FITC) and SLC2A2 (Texas red, DNA in DAPI, scale bar: 50μm). (c) FACS analysis of intestinal epithelial cells, SLC2A2 and Epcam positive cells are highlighted in red. (d) Genome wide DCM and CpG methylation level at different time points after start of dox induction. (e,f) DCM labelling (relative to total and normalized to T=1d) of *Ephb2, Ephb3, Efnb1*, and *Efnb2* genes (e), and the ubiquitously expressed *Polr2c, Pgk1, Rpl35a, Eif1* and *Gapdh* genes at different time points after start of dox treatment (f). (g) Immunocytochemistry detecting GNL3 and NUP54 in combination with CD44 (DNA is DAPI, scale bar: 16μm) (h) Normalized UMI count distribution per cell type on day 1, 2 and 8 DCM labelling peak time point, showing increased expression of genes at day 2.

Supplementary Figure 5

Supplementary Figure 5

(a) Heatmap showing the overlap of villi H3K27ac peaks with several villi ChIP-seq and DCM datasets. All ChIP-seq datasets were generated from villi samples [Saxena et al. 2017] and the DCM data from –dox samples and the day 1 and day 2 +dox samples are shown. H3K27ac peaks are ordered according to overlapping DCM signal ±1kb of the peak center and split in four equally sized clusters based on this ordering. Each profile plot has the same y-axis range as its corresponding heatmap. (b) Histogram with the distance to the closest significant DCM site for both the DCM labeled H3K27ac peaks (i.e. peaks with ≥ 1 significant DCM site) and random controls based on 100 sets of reshuffled H3K27ac peaks. Density for each 100 bp bin up to 5kb is shown. (c) Percentage of H3K27ac peaks that are labelled by DCM (i.e. peaks with ≥ 1 significant DCM site <750 bp from peak). The H3K27ac peaks are split in four clusters based on H3K27ac intensity and related to all peaks. (d) Barplot showing the number of DCM sites overlapping each peak. The percentages with each number of sites are shown for all peaks, the labeled peaks (i.e. peaks with ≥ 1 significant DCM site) and non-labeled peaks (i.e. peaks wihout significant DCM site). (e) Correlation heatmap showing the spearman correlation at the H3K27ac peaks between the different ChIP-seq and DCM datasets shown in (a). The number of reads overlapping each

532   H3K27ac peak with ≥ 3 DCM sites were normalized for the peak length or the number of DCM sites for
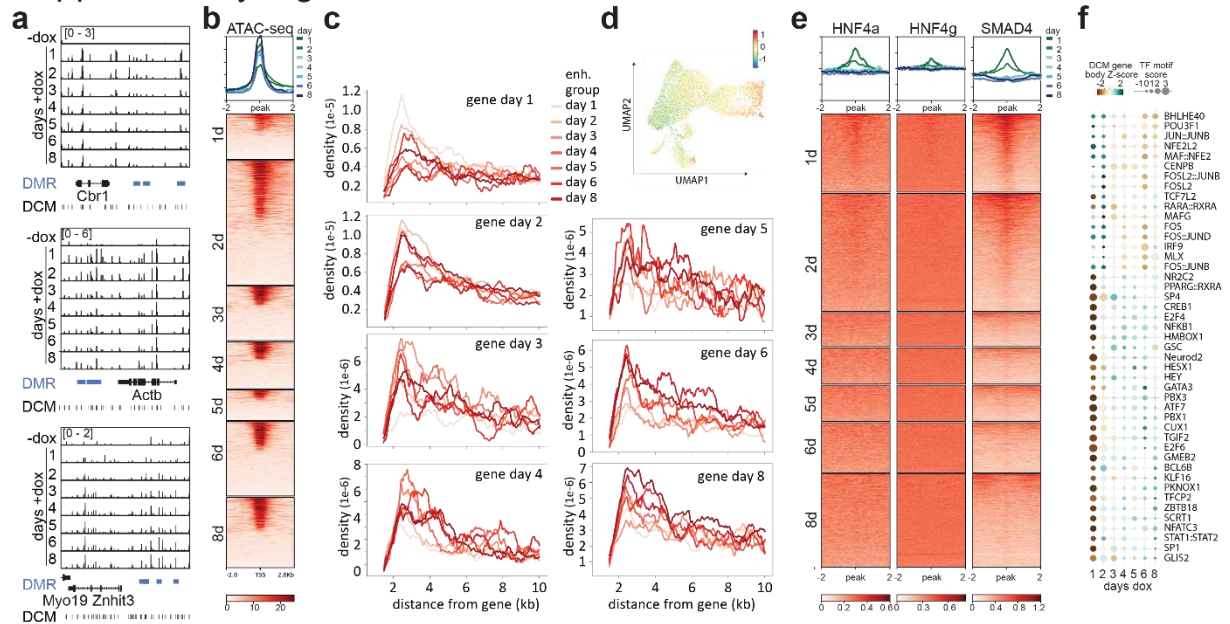
533   the ChIP-seq and DCM datasets, respectively.

534

535

536

537

33

Supplementary Figure 6

Supplementary Figure 6

(a) Genome browser view showing DCM labelling of enterocyte specific (*Cbr1*), ubiquitous (*Actb*) and ISC specific (*Znhit3*) genes with enhancers (in blue) nearby showing similar behaviour in time. (b) Heatmap showing ATAC-seq overlap with the regions around TSS of genes peaking at different days of dox induction. (c) Density plot showing the number of enhancer DMRs per peak day in the 10kb region around genes split in clusters based on peak timing of gene body DCM labelling. (d) Differential enrichment of genes peaking on day 2 split by having relatively more enhancers peaking on day 2 or day 8 nearby; difference in gene expression plot on UMAP showing increased expression of genes linked to day 2 enhancers in enterocytes. (e) Heatmap showing ChIP-seq overlap for HNFa, HNFg and SMAD4 with regions around enhancer DMRs. Enhancers were split in clusters based on the maximum day of DCM accumulation. (f) (f) Combined analysis of motif enrichment and DCM gene body labelling dynamics of TFs displaying a negative correlation in time (right).

34

Supplementary Figure 7
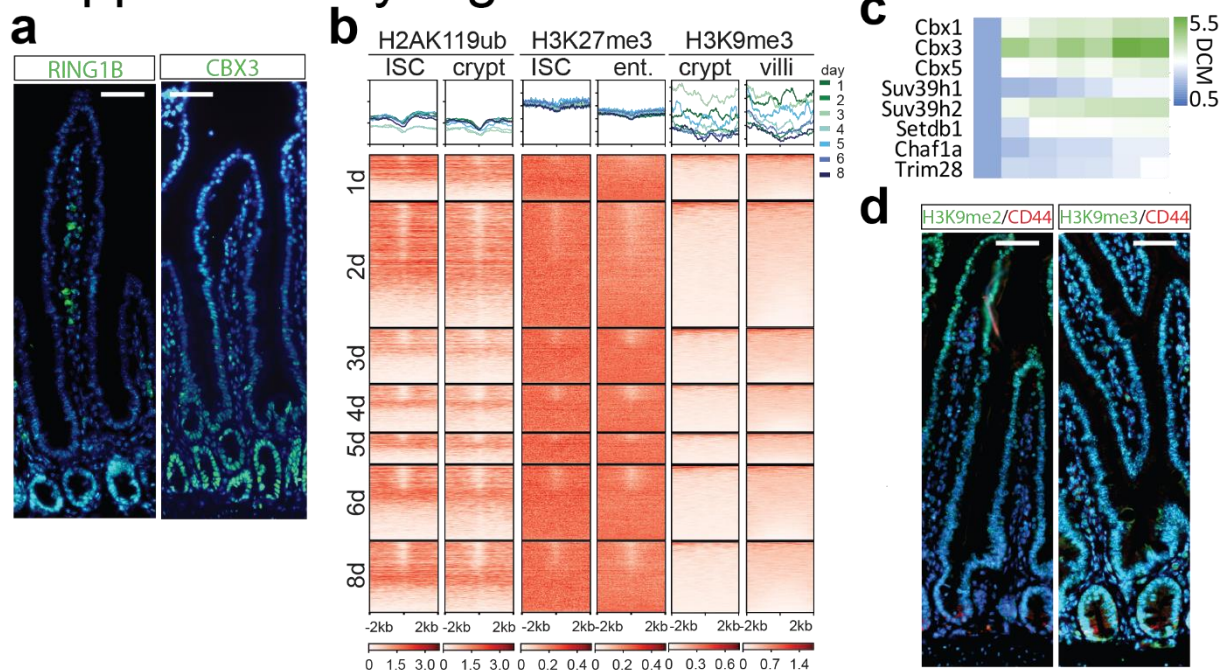
551

552 **Supplementary Figure 7**

553 (a) Immunocytochemistry detecting RING1B and CBX3 (FITC) (DNA=DAPI, scale bar: 50μm). (b)

554 Heatmap showing ChIP-seq overlap in enterocytes, ISC, crypt and villi for H2A119ub, H3K27me3, and

555 H3K9me3 with regions around the TSS DMRs. TSS were split in clusters based on the maximum day of

556 DCM accumulation and ordered according to H3K27ac and ATAC-seq enrichment. (c) Temporal
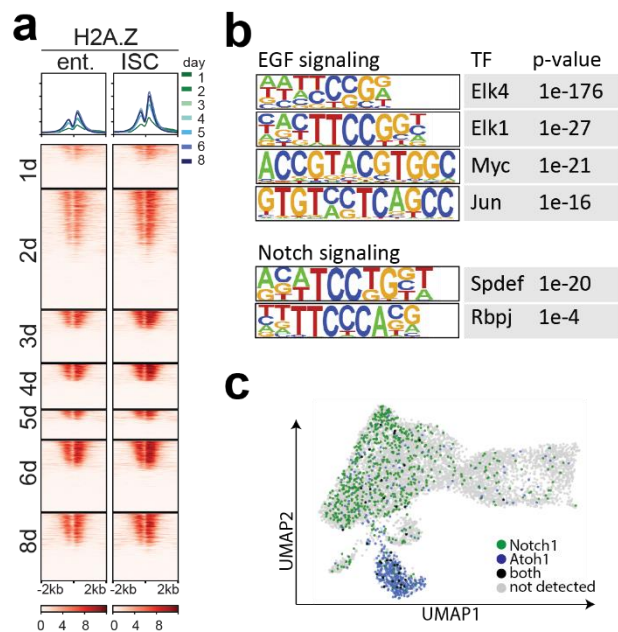
557 behaviour of DCM methylation (normalized to t=1 day) of genes encoding proteins involved in

558 establishment and maintenance of constitutive heterochromatin. (d) Immunocytochemistry detecting

559 H3K9me2 and H3K9me3 (FITC) in combination with CD44 (Texas Red, scale bar: 50μm).

560

Supplementary Figure 8

561

Supplementary Figure 8

(a) Heatmap showing H2A.Z ChIP-seq overlap in enterocytes and ISC with regions around the TSS DMRs. TSSs were split in clusters based on the maximum day of DCM accumulation and ordered according to H3K27ac and ATAC-seq enrichment. (b) HOMER motif analysis on H2A.Z enhancer peaks present in enterocytes revealing motif enrichment for TFs downstream of EGF and Notch signalling. (c) UMAP displaying cells in which only *Notch1* or *Atoh1* was detected and cells where both genes or no gene was detected (plotted in UMAP shown in Figure 4g).

569   **Supplementary Table 1**

570   Sequencing statistics.

571

572   **Supplementary Table 2**

573   Overview of gene body DCM counts in uninduced and induced ES cells and enterocytes.

574

575   **Supplementary Table 3**

576   Overview of intergenic differentially methylated DCM sites in uninduced and induced ES cells and

577   enterocytes.

578

579   **Supplementary Table 4**

580   Motif analysis on intergenic DMRs at different timepoints of dox induction in enterocytes.

581

582   **Supplementary Table 5**

583   KEGG pathway analysis on differential DCM labelling of genes upon dox induction.

584

585

586

587

References

Alemany, A., Florescu, M., Baron, C.S., Peterson-Maduro, J., and van Oudenaarden, A. (2018). Whole-organism clone tracing using single-cell sequencing. Nature *556*, 108-112.

Arand, J., Spieler, D., Karius, T., Branco, M.R., Meilinger, D., Meissner, A., Jenuwein, T., Xu, G., Leonhardt, H., Wolf, V.*, et al.* (2012). In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. PLoS Genet *8*, e1002750.

Aranda, S., Mas, G., and Di Croce, L. (2015). Regulation of gene transcription by Polycomb proteins. Sci Adv *1*, e1500737.

Barker, N., van Es, J.H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J.*, et al.* (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature *449*, 1003-1007.

Baron, C.S., and van Oudenaarden, A. (2019). Unravelling cellular relationships during development and regeneration using genetic lineage tracing. Nat Rev Mol Cell Biol *20*, 753-765.

Beard, C., Hochedlinger, K., Plath, K., Wutz, A., and Jaenisch, R. (2006). Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. Genesis *44*, 23-28.

Beumer, J., and Clevers, H. (2020). Cell fate specification and differentiation in the adult mammalian intestine. Nat Rev Mol Cell Biol.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L.P., Ito, S., Cooper, S., Kondo, K., Koseki, Y.*, et al.* (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell *157*, 1445-1459.

Boers, R., Boers, J., de Hoon, B., Kockx, C., Ozgur, Z., Molijn, A., van, I.W., Laven, J., and Gribnau, J. (2018). Genome-wide DNA methylation profiling using the methylation-dependent restriction enzyme LpnPI. Genome Res *28*, 88-99.

Bowling, S., Sritharan, D., Osorio, F.G., Nguyen, M., Cheung, P., Rodriguez-Fraticelli, A., Patel, S., Yuan, W.C., Fujiwara, Y., Li, B.E.*, et al.* (2020). An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. Cell *181*, 1693-1694.

Bradner, J.E., Hnisz, D., and Young, R.A. (2017). Transcriptional Addiction in Cancer. Cell *168*, 629-643.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol Cell *49*, 825-837.

Chen, L., Toke, N.H., Luo, S., Vasoya, R.P., Fullem, R.L., Parthasarathy, A., Perekatt, A.O., and Verzi, M.P. (2019). A reinforcing HNF4-SMAD4 feed-forward module stabilizes enterocyte identity. Nat Genet *51*, 777-785.

Chiacchiera, F., Rossi, A., Jammula, S., Piunti, A., Scelfo, A., Ordonez-Moran, P., Huelsken, J., Koseki, H., and Pasini, D. (2016a). Polycomb Complex PRC1 Preserves Intestinal Stem Cell Identity by Sustaining Wnt/beta-Catenin Transcriptional Activity. Cell Stem Cell *18*, 91-103.

Chiacchiera, F., Rossi, A., Jammula, S., Zanotti, M., and Pasini, D. (2016b). PRC2 preserves intestinal progenitors and restricts secretory lineage commitment. EMBO J *35*, 2301-2314.

Clark, S.J., Harrison, J., and Frommer, M. (1995). CpNpG methylation in mammalian cells. Nat Genet *10*, 20-27.

Ferrari, K.J., Amato, S., Noberini, R., Toscani, C., Fernandez-Perez, D., Rossi, A., Conforti, P., Zanotti, M., Bonaldi, T., Tamburri, S.*, et al.* (2021). Intestinal differentiation involves cleavage of histone H3 N-terminal tails by multiple proteases. Nucleic Acids Res *49*, 791-804.

Frieda, K.L., Linton, J.M., Hormoz, S., Choi, J., Chow, K.K., Singer, Z.S., Budde, M.W., Elowitz, M.B., and Cai, L. (2017). Synthetic recording and in situ readout of lineage information in single cells. Nature *541*, 107-111.

Giaimo, B.D., Ferrante, F., Herchenrother, A., Hake, S.B., and Borggrefe, T. (2019). The histone variant H2A.Z in gene regulation. Epigenetics Chromatin *12*, 37.

38

Haber, A.L., Biton, M., Rogel, N., Herbst, R.H., Shekhar, K., Smillie, C., Burgin, G., Delorey, T.M., Howitt, M.R., Katz, Y., *et al.* (2017). A single-cell survey of the small intestinal epithelium. Nature *551*, 333-339.

Herman, J.S., Sagar, and Grun, D. (2018). FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. Nat Methods *15*, 379-386.

Ishibashi, F., Shimizu, H., Nakata, T., Fujii, S., Suzuki, K., Kawamoto, A., Anzai, S., Kuno, R., Nagata, S., Ito, G., *et al.* (2018). Contribution of ATOH1(+) Cells to the Homeostasis, Repair, and Tumorigenesis of the Colonic Epithelium. Stem Cell Reports *10*, 27-42.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lonnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods *11*, 163-166.

Ito, N., Kii, I., Shimizu, N., Tanaka, H., and Takeda, S. (2017). Direct reprogramming of fibroblasts into skeletal muscle progenitor cells by transcription factors enriched in undifferentiated subpopulation of satellite cells. Sci Rep *7*, 8097.

Kaaij, L.T., van de Wetering, M., Fang, F., Decato, B., Molaro, A., van de Werken, H.J., van Es, J.H., Schuijers, J., de Wit, E., de Laat, W., *et al.* (2013). DNA methylation dynamics during intestinal stem cell differentiation reveals enhancers driving gene expression in the villus. Genome Biol *14*, R50.

Kazakevych, J., Sayols, S., Messner, B., Krienke, C., and Soshnikova, N. (2017). Dynamic changes in chromatin states during specification and differentiation of adult intestinal stem cells. Nucleic Acids Res *45*, 5770-5784.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., *et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182-187.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lonnerberg, P., Furlan, A., *et al.* (2018). RNA velocity of single cells. Nature *560*, 494-498.

Lee, B.K., Shen, W., Lee, J., Rhee, C., Chung, H., Kim, K.Y., Park, I.H., and Kim, J. (2015). Tgif1 Counterbalances the Activity of Core Pluripotency Factors in Mouse Embryonic Stem Cells. Cell Rep *13*, 52-60.

Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237-1251.

Liu, Y., Maekawa, T., Yoshida, K., Muratani, M., Chatton, B., and Ishii, S. (2019). The Transcription Factor ATF7 Controls Adipocyte Differentiation and Thermogenic Gene Programming. iScience *13*, 98-112.

Lopez-Arribillaga, E., Rodilla, V., Pellegrinet, L., Guiu, J., Iglesias, M., Roman, A.C., Gutarra, S., Gonzalez, S., Munoz-Canoves, P., Fernandez-Salguero, P., *et al.* (2015). Bmi1 regulates murine intestinal stem cell proliferation and self-renewal downstream of Notch. Development *142*, 41-50.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72-77.

Ng, A.Y., Waring, P., Ristevski, S., Wang, C., Wilson, T., Pritchard, M., Hertzog, P., and Kola, I. (2002). Inactivation of the transcription factor Elf3 in mice results in dysmorphogenesis and altered differentiation of intestinal epithelium. Gastroenterology *122*, 1455-1466.

Parker, A., Maclaren, O.J., Fletcher, A.G., Muraro, D., Kreuzaler, P.A., Byrne, H.M., Maini, P.K., Watson, A.J., and Pin, C. (2017). Cell proliferation within small intestinal crypts is the principal driving force for cell migration on villi. FASEB J *31*, 636-649.

Saxena, M., Roman, A.K.S., O'Neill, N.K., Sulahian, R., Jadhav, U., and Shivdasani, R.A. (2017). Transcription factor-dependent 'anti-repressive' mammalian enhancers exclude H3K27me3 from extended genomic domains. Genes Dev *31*, 2391-2404.

Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., *et al.* (2019). Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. Cell *176*, 1517.

688   Schmidt, F., Cherepkova, M.Y., and Platt, R.J. (2018). Transcriptional recording by CRISPR spacer
689   acquisition from RNA. Nature *562*, 380-385.
690   Stadhouders, R., Filion, G.J., and Graf, T. (2019). Transcription factors and 3D genome conformation
691   in cell-fate decisions. Nature *569*, 345-354.
692   Tetteh, P.W., Basak, O., Farin, H.F., Wiebrands, K., Kretzschmar, K., Begthel, H., van den Born, M.,
693   Korving, J., de Sauvage, F., van Es, J.H.*, et al.* (2016). Replacement of Lost Lgr5-Positive Stem Cells
694   through Plasticity of Their Enterocyte-Lineage Daughters. Cell Stem Cell *18*, 203-213.
695   Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J.,
696   Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed
697   by pseudotemporal ordering of single cells. Nat Biotechnol *32*, 381-386.
698   van Steensel, B., and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins
699   using tethered dam methyltransferase. Nat Biotechnol *18*, 424-428.
700   VanDussen, K.L., Carulli, A.J., Keeley, T.M., Patel, S.R., Puthoff, B.J., Magness, S.T., Tran, I.T., Maillard,
701   I., Siebel, C., Kolterud, A.*, et al.* (2012). Notch signaling modulates proliferation and differentiation of
702   intestinal crypt base columnar stem cells. Development *139*, 488-497.
703   Yan, K.S., Gevaert, O., Zheng, G.X.Y., Anchang, B., Probert, C.S., Larkin, K.A., Davies, P.S., Cheng, Z.F.,
704   Kaddis, J.S., Han, A.*, et al.* (2017). Intestinal Enteroendocrine Lineage Cells Possess Homeostatic and
705   Injury-Inducible Stem Cell Activity. Cell Stem Cell *21*, 78-90 e76.
706   Yu, S., Tong, K., Zhao, Y., Balasubramanian, I., Yap, G.S., Ferraris, R.P., Bonder, E.M., Verzi, M.P., and
707   Gao, N. (2018). Paneth Cell Multipotency Induced by Notch Activation following Injury. Cell Stem Cell
708   *23*, 46-59 e45.

709