

Motif distribution and DNA methylation underlie distinct Cdx2 binding during development and homeostasis

Received: 27 February 2024

Accepted: 7 January 2025

Published online: 22 January 2025

 Check for updatesAlireza Lorzadeh¹, George Ye¹, Sweta Sharma¹ & Unmesh Jadhav^{1,2} 

Transcription factors guide tissue development by binding to developmental stage-specific targets and establishing an appropriate enhancer landscape. In turn, DNA and chromatin modifications direct the genomic binding of transcription factors. However, how transcription factors navigate chromatin features to selectively bind a small subset of all the possible genomic target loci remains poorly understood. Here we show that Cdx2—a lineage defining transcription factor that binds distinct targets in developing versus adult intestinal epithelial cells—has a preferential affinity for a non-canonical CpG-containing motif *in vivo*. A higher frequency of this motif at embryonic Cdx2 targets and methylated state of the CpG during development enables selective Cdx2 binding and activation of developmental enhancers and genes. In adult cells, demethylation at these enhancers prevents ectopic Cdx2 binding, instead directing Cdx2 to its canonical motif without a CpG. This shift in Cdx2 binding facilitates Ctf and Hnf4 recruitment, establishing super-enhancers during development and homeostatic enhancers in adult cells, respectively. Induced DNA methylation in adult mouse epithelium or cultured cells recruits Cdx2 to developmental targets, promoting corecruitment of partner transcription factors. Thus, Cdx2's differential CpG motif preferences enable it to navigate distinct DNA methylation profiles, activating genes specific to appropriate developmental stages.

The establishment of functional cell types in adult tissues requires precise gene expression control throughout development. Transcription factors (TFs) recognize and bind to specific and short genomic sequences called target motifs within gene promoters and cis-regulatory enhancers giving rise to distinct transcriptional programs. Typically, TF motifs occur millions of times in the mammalian genome, however, most TFs only bind a fraction of those^{1,2}. How TFs achieve such cell type-specific binding is not fully understood. Multiple factors including DNA and histone modifications, nucleosome positioning, chromatin compaction, and motif frequency and variation may play a role^{3,4}. Many TFs can overcome barriers posed by chromatin features to help recruit other TFs, form and activate gene regulatory enhancers,

and promote gene transcription during tissue development and adult homeostasis⁵. To investigate such molecular controls *in vivo*, we utilized the intestinal epithelium as a model system. This single-cell thick inner lining of the intestine develops from a well-defined region of embryonic gut endoderm. In adults, intestinal stem cells (ISCs) residing in the crypts and continuously divide to produce terminally differentiated villus cells, which are predominantly absorptive enterocytes⁶. Cdx2, a homeodomain-containing TF, is a lineage-defining factor expressed in all intestinal epithelial cells from development through adulthood. It plays a pivotal role in establishing small intestinal cell types and promoting the maturation of functional cells in adult tissue. Developmental period around embryonic day 12.5 in

¹Department of Stem Cell Biology and Regenerative Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. ²Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. ✉e-mail: ujadhav@usc.edu

mouse (E12.5) forms a critical window for establishing intestinal identity and deletion of *Cdx2* before this time point results in ectopic expression of genes from other parts of the gut tube such as the esophagus and stomach^{7–9}. On the other hand, deletion of *Cdx2* after E12.5 during development or in the adult epithelial cells disrupts maturation of functional cell types and expression of genes required for homeostatic function^{10,11}. Thus, despite its consistent expression in developing and adult intestinal epithelium, *Cdx2* can activate different stage-specific genes. Prior studies have shown that *Cdx2* achieves such stage-specific gene control by binding to diverse genomic targets across developing epithelium and in adult cells and by establishing distinct chromatin accessibility patterns¹⁰. Thus, the developing intestinal epithelium offers a unique opportunity to explore the mechanisms that dynamically regulate TF interactions with epigenetic and chromatin features across the lifespan, establishing cell type- and developmental stage-specific enhancers and gene expression patterns.

Methylation at cytosine in the CpG dinucleotide is a well-studied epigenetic modification linked to transcriptional repression. Almost all CpGs across the vertebrate genome are methylated except for the ones at the active regulatory regions like promoters and enhancers^{12,13}. DNA methylation may further influence the formation and activity of these regulatory regions by impacting TF binding, which is central to establishing the open chromatin and facilitating epigenetic modifications. Recent high-throughput in vitro studies characterized the effect of CpG methylation on the binding of more than 500 human TFs to various DNA binding domains using SELEX (Systematic Evolution of Ligands by Exponential Enrichment), showing that more TFs prefer methylated motifs (34%) compared to those that are inhibited (23%) when containing a CpG in their motif¹⁴. These findings suggest that sensitivity to DNA methylation may be a major factor controlling TF binding in vivo. In addition to the *CDX2* binding at the canonical motif without a CpG (ATAAA), *CDX2* was shown to bind nucleotides with motifs containing CpGs with high affinity when the cytosine was methylated in vitro¹⁴. However, for most TFs (including *CDX2*), it remains unknown whether DNA methylation influences their genomic binding in vivo, where additional controlling factors may include DNA sequence surrounding the motif, nucleosome presence, and chromatin modifications. Our studies have previously revealed the genome-wide dynamics of DNA methylation in developing epithelial endoderm, where specific promoter and enhancer loci are demethylated stepwise to produce the adult chromatin landscape¹⁵. Here we asked how *Cdx2* navigates this chromatin landscape based on its capacity to bind variable motifs with and without DNA methylation and other chromatin controls. We demonstrate that *Cdx2* binding across development and in adult cells is influenced by the distribution of its CpG-containing and canonical (non-CpG-containing) motifs. The prevalence of the CpG-containing motif at *Cdx2*'s developmental targets, along with the methylation of the CpG within this motif, facilitates *Cdx2* binding during development, while the absence of CpG methylation at this motif helps avoid ectopic *Cdx2* recruitment at these loci in adult cells. Using two independent methods, we modulated DNA methylation through genetic elimination of Polycomb action in mouse intestinal epithelial cells and more directly using a chemical inhibitor in cultured cells. We show that induced methylation at the CpG-containing motifs causes ectopic *Cdx2* recruitment at developmental targets in adult cells. Based on this CpG methylation-dependent binding, *Cdx2* facilitates recruitment of the TFs, *Ctcf* and *Hnf4a* to distinct genomic targets to establish developmental versus adult enhancer patterns and gene activation in the mammalian intestinal epithelium.

Results

Distinct promoter- and enhancer-based gene control by *Cdx2* in developing and adult intestinal epithelium

Lineage-defining factors like *Cdx2* overcome obstacles posed by DNA and chromatin modifications to access their target motif sequences

across the genome^{16,17}. During mouse small intestinal development, where cell numbers are limiting, CUT&RUN assays allow precise detection of TF binding using just a few thousand cells as compared to the traditional ChIP-seq assays which require many times more cells as input¹⁷. We used 50,000 purified intestinal epithelial cells to determine *Cdx2* binding at embryonic day 12.5 (E12.5), fetal timepoint 16.5 (E16.5), and in adult differentiated villus cells (Supplementary Fig. 1a, b). CUT&RUN allowed us to identify a substantial number of *Cdx2* binding sites in the developing epithelium (20,880 and 24,119 at E12.5 and E16.5, respectively) as well as adult cells (27,737). Importantly, a large number of *Cdx2* binding events were specific to either development (7806 in E12.5 and E16.5) or adult villus cells (8901) (Fig. 1a). To further classify these loci, we used DESeq2¹⁸ based analysis to identify 2-fold changes in CUT&RUN signal of *Cdx2* binding from E12.5 or E16.5 to adult ($q < 0.01$) and divided the developmental and adult-specific *Cdx2* binding in 3 subgroups each (Dev 1–3 and Adult 1–3, respectively, Fig. 1a and Supplementary Fig. 1c). In line with previous findings¹⁰, 60% of all of the 8901 adult binding sites (Adult 1–3) showed *Cdx2* binding (MACS2 identified peak) at E16.5, a fetal timepoint representing advanced development. On the other hand, while the *Cdx2* binding at E12.5 was completely lost or reduced by 2-fold or more at all developmentally bound sites (Dev 1–2), we observed some signal at 8% of these sites in adult cells (Fig. 1a). This highlights the disparity in *Cdx2* binding in developing versus adult intestinal epithelial cells and suggests a need for distinct underlying molecular controls for temporally selective *Cdx2* recruitment. Genes linked to the developmental versus adult *Cdx2* binding loci (*Cdx2* peak ± 25 kb of the gene transcription start site, TSS) showed corresponding loss or gain of expression¹⁵ (Supplementary Fig. 1d), suggesting a direct role of *Cdx2* in regulating transcription of multiple stage-specific genes.

To test if *Cdx2* engages in distinct promoter- or enhancer-driven control of target genes in developing and adult epithelium, we classified the genomic loci with dynamic *Cdx2* binding (Dev 1–3 and Adult 1–3, Fig. 1a) based on their proximity to gene TSSs. At the loci with highly enriched *Cdx2* occupancy during embryonic development (E12.5, Dev1, Fig. 1a), 24% of the target genes had promoter-based *Cdx2* binding (2 kb upstream and 1 kb downstream of the TSS) as compared to all other groups (Dev 2–3, and Adult 1–3, Fig. 1b and Supplementary Fig. 1e), which had average 8% and maximum of 13% promoter occupancy. On the other hand, majority (>87%) of the *Cdx2* binding sites in fetal (E16.5) and adult cells lie within regulatory regions outside the promoters. We confirmed that the *Cdx2* binding loci were bona fide promoter or enhancers by overlapping them with publicly available ATAC-seq data and active chromatin-associated histone modification Histone 3 lysine 27 acetylation (H3K27ac, Supplementary Fig. 1f)^{15,19}. Notably, regions occupied by *Cdx2* in adult cells (Adult 1–3) show closed chromatin structure during early development (E12.5, Supplementary Fig. 1f) indicating that the temporal shift in *Cdx2* binding from E12.5 going into adult cells also resulted in altered chromatin accessibility. The prevalent promoter-based *Cdx2* binding in embryonic cells is quickly lost or reduced in fetal cells at E16.5 (90% of the 1346 promoters at E12.5, Fig. 1c, d), when significant morphological rearrangements, such as the formation of the crypt and villus structures, take place in the intestinal epithelium²⁰. Conversely, *Cdx2* binding is retained at 44% of the embryonic (E12.5) enhancers in fetal (E16.5) cells, and 49% of this fetal binding is further retained in adult cells. Additionally, 65% and 58% of the enhancers in E16.5 and adult cells are newly bound by *Cdx2* (Figs. 1c and S1g). This suggests a promoter-centric bias in gene targeting by *Cdx2* in embryonic cells while transitioning into a more enhancer-driven control in adult cells during development. Importantly, the embryonic (E12.5) promoter targets of *Cdx2* consist of genes critical for intestinal development such as *Sox4* and *Meis1*²¹ (Fig. 1d and Supplementary Fig. 1h). Interestingly, about 10% of promoter-based *Cdx2* targets in Dev 1–2 are other TFs, while only about 3% of the adult-specific *Cdx2* binding is at TF promoters. This

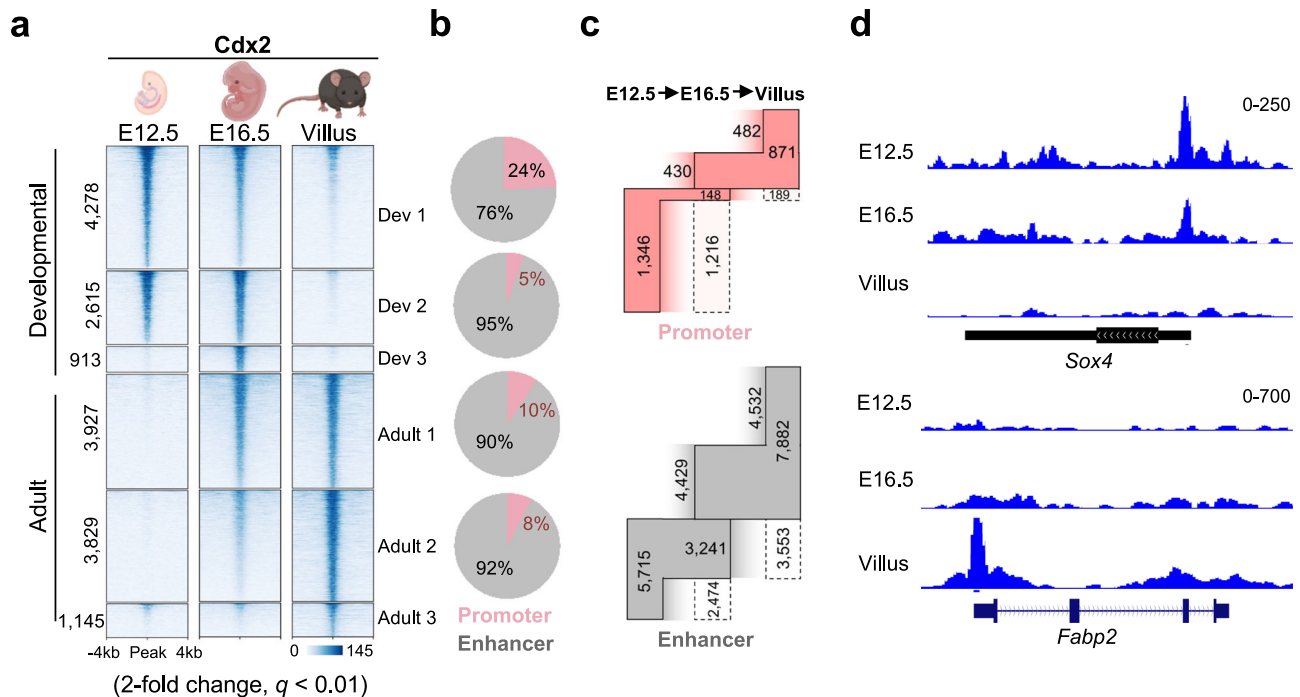


Fig. 1 | Evolution of Cdx2 binding across gene promoters and enhancers supports developmental and homeostatic functions in intestinal epithelium. **a** Heatmap showing Cdx2 CUT&RUN signal at its developmental and adult-specific binding sites in embryonic (E12.5), fetal (E16.5), and adult epithelial (villus) cells; Dev 1–3 and Adult 1–3 subgroups represent 2-fold change ($q < 0.01$) in Cdx2 signal among the three timepoints. **b** Pie charts showing percentages of Cdx2 binding sites shown in (a) (Dev 1, 2 and Adult 1, 2) at promoters (TSS ~ 2 kb to $+1$ kb) and

distal enhancers. **c** Gain, retention, and loss of Cdx2 binding at promoters (top panel) and distal enhancers (bottom panel) during development from E12.5 into adult epithelial cells; numbers outside the boxes and inside the dotted boxes represent gains and losses of Cdx2 binding at each of the timepoints, respectively. **d** Genome browser tracks showing progressive loss or recruitment of Cdx2 through epithelial development at *Sox4* and *Fabp2* gene promoters, respectively.

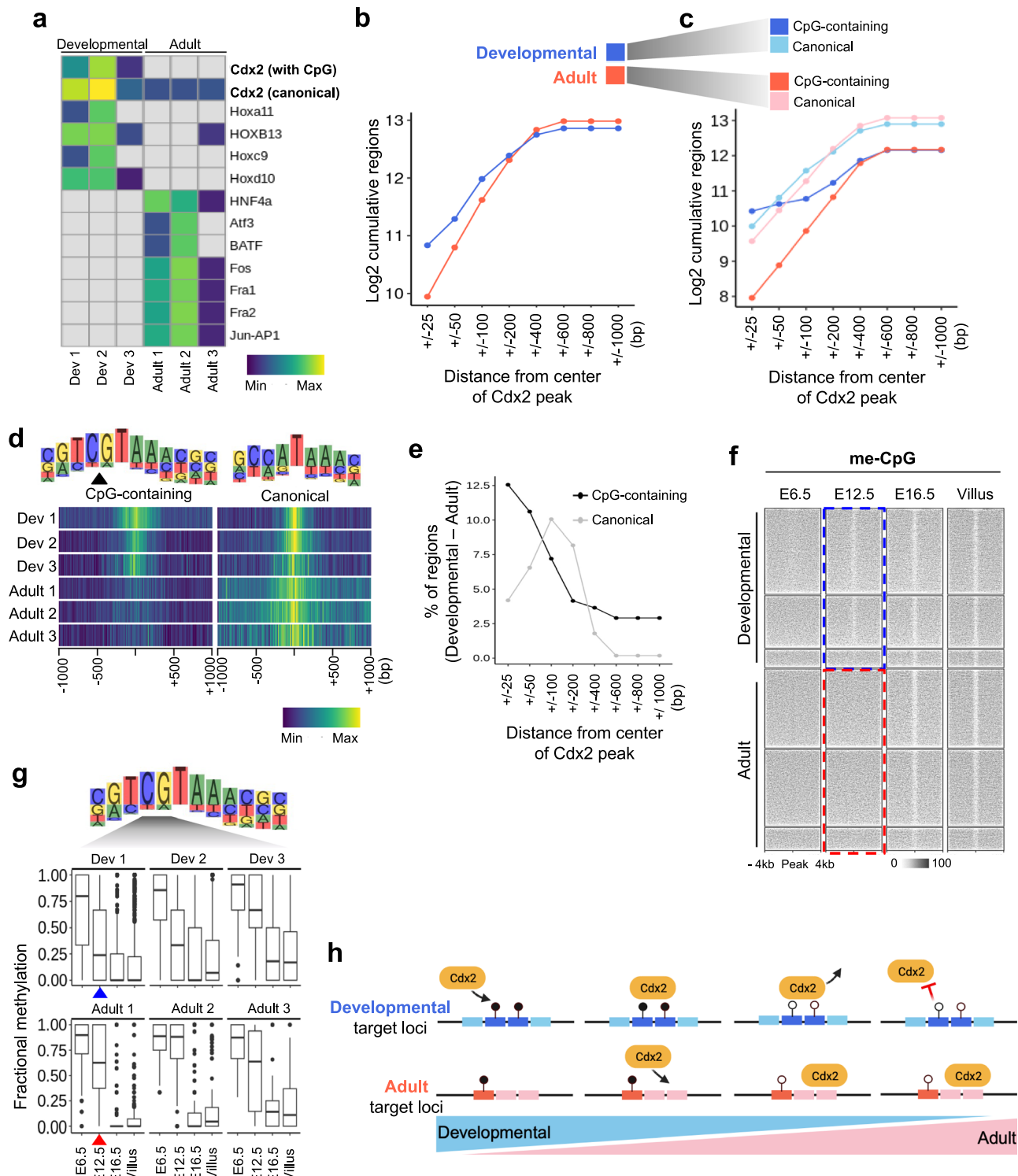
suggests that Cdx2 engages in directly activating other TFs critical for intestinal epithelial lineage commitment early on during development and engages in non-TF and functional gene activation through enhancers in adult tissue (Supplementary Fig. 1i).

Pronounced presence of the CpG-containing Cdx2 motif at developmental binding sites

How various TFs depend on or overcome chromatin modifications to establish temporally distinct binding patterns remains poorly understood. In addition to the canonical binding motif containing the sequence ATAAA, studies show that Cdx2 binds to a CpG-containing motif with high affinity in vitro when the CpG (at the 4th base in this motif) is methylated (Supplementary Fig. 2a). Based on these observations, we hypothesized that the differential Cdx2 sensitivity to distinct Cdx2 motifs might allow it to navigate and preferentially bind developmental or adult loci. To this end, we analyzed the Cdx2 binding dynamics across developmental and adult epithelial villus cells (Fig. 1a) in the context of Cdx2 motif distribution and the temporal changes in CpG methylation. First, we determined the relative enrichment of the canonical and the CpG-containing motifs of Cdx2 (Supplementary Fig. 2a) at the Cdx2 target loci Dev 1–3 and Adult 1–3 (Fig. 1a) using the motif enrichment analysis tool HOMER²². While the CpG-containing motif was significantly enriched at the developmental loci ($p < 10^{-92}$), it did not show enrichment at adult loci (Fig. 2a). On the other hand, the canonical Cdx2 motif was present in both developmental and adult sites. As expected, the motif for Hnf4A was only enriched at adult loci, while motifs of Hoxc9 and Hoxd10 (with high sequence similarity to the CpG-containing Cdx2 motif), were enriched in developmental loci. To dissect the Cdx2 motif distribution across developmental and adult binding sites in detail, we used the FIMO algorithm which identifies significant instances of specific motifs within a set of input regions²³. As seen in the HOMER analysis (Fig. 2a), FIMO detected Cdx2 motifs

(canonical or CpG-containing) more frequently at the developmental Cdx2 peaks in comparison to the adult loci, particularly near the center of the Cdx2 peaks (± 100 bp, $p < 0.005$, Fig. 2b); within 50 bp (± 25 bp) from the center of the Cdx2 peaks, almost twice as many of the developmental loci had a detectable Cdx2 motif as compared to the adult loci (1823 vs 986, Fig. 2b).

We next compared the distribution of the canonical versus CpG-containing motifs in development and adult cell-specific Cdx2 binding sites. While the canonical Cdx2 motif was present more frequently at both developmental and adult loci in comparison to the CpG-containing motif, developmental loci (Dev 1–3) showed higher prevalence of the CpG-containing motif in comparison to the adult loci (Adult 1–3), particularly within 50 bp of the center of Cdx2 peaks (± 25 bp, Fig. 2c, d). Among all Dev 1 loci, where Cdx2 binds early during development, the CpG-containing motif was represented 14% more frequently than in Adult 1 loci, where Cdx2 is first recruited to establish a stable binding that lasts in the adult villus epithelium (Supplementary Fig. 2b); collectively Dev 1–3 loci with the CpG-containing motif exceeded the number of adult sites (Adult 1–3) with this motif by 12.5% (Fig. 2e). When considering loci where motifs were detectable within 50 bp (± 25 bp) from the center of Cdx2 peaks, 67% of the developmental sites contained CpG-containing motif, which exceeded the number of adult sites with a detectable motif of this type by 42%. This suggested that Cdx2 recruitment to distinct loci in the embryonic and adult epithelium may be linked to the relative prevalence of different motifs. A higher prevalence of the CpG-containing motif towards the center of the developmental Cdx2 peaks and a relatively broader distribution of the canonical Cdx2 motif (Fig. 2d and Supplementary Fig. 2b) suggested a more focused targeting of Cdx2 at developmental loci. Indeed, the developmental Cdx2 peaks were significantly sharper (smaller genomic footprint) than the peaks in adult cells (Supplementary Fig. 2c). Interestingly, the developmental loci show more



conservation across species as compared to the adult loci, suggesting a conserved molecular mechanism based on motif distribution and DNA methylation controlling Cdx2 recruitment to developmental targets (Fig. S2d). To determine if Cdx2 binding at the developmental loci is facilitated by cooperative binding with other TFs, we identified all potential TF that may bind the Cdx2 target sites using HOMER ($p < 10^{-100}$), which revealed multiple Hox genes (Supplementary Fig. 2e). Lack of expression of all these TFs during the developmental timepoints precluded their potential role in co-recruiting Cdx2. To quantify the effect of Cdx2 binding through the CpG-containing or the canonical motif on gene expression, we identified genes with

transcription start site (TSS) within 25 kb of such loci (Fig. 2c). Six hundred seventy-nine genes linked to the Cdx2 bound loci with CpG-containing motif had significantly higher expression in developmental epithelium (E12.5 and E16.5, $p < 0.001$, ANOVA), while 710 genes linked to Cdx2 bound loci with the canonical motif showed bias towards regulating fetal (E16.5) and adult functions (Supplementary Fig. 2f).

These data suggest that the greater prevalence of the CpG-containing Cdx2 motif near the center of developmental loci (Fig. 2a–e) may selectively attract Cdx2 to these sites when the CpG in that motif is methylated, which happens only during development (Fig. 2f). On the other hand, the canonical Cdx2

Fig. 2 | Heightened presence of CpG containing Cdx2 motif at its developmental binding sites. **a** Heatmap showing relative prevalence of TF motifs at developmental (Dev 1–3) and adult (Adult 1–3) cell-specific Cdx2 binding sites as determined by HOMER. **b** Line plot showing number of Cdx2 binding loci containing Cdx2 motifs within an increasing distance (on X-axis) from center of the Cdx2 peak. **c** Line plot showing abundance of canonical and CpG-containing Cdx2 motifs at developmental (shades of blue) and adult (shades of red) cell-specific Cdx2 binding sites within an increasing distance (on X-axis) from center of the Cdx2 peak. **d** Density plots showing relative prevalence of CpG-containing motif near the center of developmental Cdx2 peaks (Dev 1–3) in comparison with the adult cell-specific Cdx2 peaks (Adult 1–3); canonical Cdx2 motif shows widespread presence near Adult 1–3 sites in contrast with the Dev 1–3 loci. **e** Higher percent of regions with development-specific Cdx2 binding contain CpG-containing motif as compared to regions with adult cell-specific Cdx2 binding near the center of Cdx2 peak (within ± 50 bp); beyond 50 bp of the peak center, canonical motif is more prevalent. **f** Heatmap showing DNA methylation at Cdx2 binding loci (as in Fig. 1a) in early endoderm (E6.5)²⁵, and embryonic (E12.5), fetal (E16.5), and adult epithelial

(villus) cells⁴⁵; Cdx2 binding in embryonic epithelium is associated with early loss of DNA methylation (blue dotted box), while sites first bound by Cdx2 in the fetal epithelium show higher methylation levels (red dotted box). **g** Fractional methylation change at CpG (4th base position) within the Cdx2 motif (For Dev1–3 and Adult 1–3, as in Fig. 1a) shows relatively early reduction in methylation at loci with developmental binding of Cdx2 (blue arrowhead) compared to binding in adult epithelium (red arrowhead) ($p < 2.2 \times 10^{-16}$, right-sided Wilcoxon test). The CpG remains hypomethylated in adult villus at both developmental and adult loci. The horizontal bars within the box plots represent medians, the tops and bottoms of the boxes represent the 75th and 25th percentiles, respectively, and the upper and lower whiskers extend to the maximum and minimum values, respectively. The outliers are indicated by dots. **h** A model for DNA methylation-based Cdx2 binding at developmental target loci; recruitment at adult target loci is driven by DNA methylation independent binding. While relative abundance of the CpG-containing motif and methylation at the CpG within the Cdx2 motif during development allows its early recruitment to these sites, demethylation at these loci protects them from ectopic Cdx2 binding in adult cells. Source data are provided as a Source Data file.

motif (non-CpG-containing, Supplementary Fig. 2a), which is more prevalent in adult Cdx2 target loci (Fig. 2a–e) may underlie the recruitment of Cdx2 in adult cells. To test these possibilities, we looked at the DNA methylation dynamics associated with altered Cdx2 binding between developmental and adult time-points (Fig. 1a). Pioneer TFs such as NRF1 and PU.1 that preferentially bind methylated DNA can recruit demethylases like Tet2 to demethylate their binding site and surrounding CpGs, which creates an opportunity for other DNA methylation-sensitive TFs to bind at the locus^{4,24}. Therefore, we considered the temporal changes in DNA methylation surrounding the Cdx2 binding peaks as well as at the CpG within the CpG-containing motif (4th base position, Supplementary Fig. 2a). In early embryonic endoderm (E6.5)²⁵, we see that all Cdx2 binding loci (Dev 1–3 and Adult 1–3) are fully methylated, including the CpG within the Cdx2 motif (Fig. 2f, g). During early embryonic development of the epithelium, only the loci with early development-specific binding of Cdx2 (Dev 1 and Dev 2) show average 23% and 32% drop in DNA methylation at E12.5 and methylation at the CpG within the Cdx2 motif is also reduced starting at E12.5. This is followed by further decrease in the DNA methylation at these loci by E16.5 (Dev 1: 30%, Dev 2: 48%, Dev 3: 49%) and in adult cells (Dev 1: 31%, Dev 2: 48%, and Dev 3: 50%, Fig. 2f). This progressive loss of DNA methylation at developmental loci and the CpG within the Cdx2 motif is accompanied by loss of Cdx2 binding at these sites starting at E16.5 and into adult cells (Fig. 1a), which suggests Cdx2's dependence on the methylated DNA for its sustained binding at the developmental sites. This may explain how Cdx2 achieves a development-specific binding at these loci as the lack of DNA methylation at these sites in adult cells may preclude its binding (Fig. 2g). On the other hand, the sites bound by Cdx2 specifically in adult cells (Adult 1–3) only show demethylation starting at E16.5, concordant with the beginning of Cdx2 binding (Figs. 1a, 2f), and progressive loss of DNA methylation at these loci is accompanied by increasing Cdx2 binding at the canonical motif (without CpG) in adult cells in contrast to the developmental sites.

Alongside the dependence of Cdx2 on methylation at the CpG within its CpG-containing motif for developmental stage-specific binding, we considered the possibility that other chromatin features, particularly repressive modifications, may inhibit its binding at these loci in adult cells. In this regard, we conducted native-ChIP-seq, an assay particularly suitable for detecting the genome-wide distribution of histone modifications using low cell number input²⁶. DNA methylation and Histone 3 lysine 27 trimethylation (H3K27me3), a modification deposited by the polycomb repressive complex 2 (PRC2), are known to have opposing genome-wide distribution^{27,28}. DNA

methylation and H3K27me3 form alternative modes of gene silencing as unmethylated genomic loci, particularly gene promoters, show widespread H3K27me3 deposition in mammalian cells^{29,30}. In ES cells and other in vitro systems, loss of DNA methylation or H3K27me3 causes alterations in the distribution of the other^{31,32}. Given this antagonistic relationship between DNA methylation and H3K27me3, we first looked at whether loss of DNA methylation at developmental loci in adult cells is accompanied by gain of H3K27me3. A small fraction of developmental sites (<10%) show deposition of H3K27me3 in adult cells (Supplementary Fig. 2g), suggesting no significant association with or impact on Cdx2 binding. H3K9me3, a modification associated with chromatin compaction, and H2AK119Ub, which is linked to active as well as inactive chromatin loci, showed very little concordance with Cdx2 binding in adult cells (Supplementary Fig. 2g). Adult loci (Adult 1–3) showed higher levels of H3K36me2, which is associated with active enhancers³³, as compared to the loci with developmental Cdx2 binding (Dev 1–3).

Together, these data strongly suggest that the genomic targeting of Cdx2 in developing and adult intestinal epithelial cells is based on variability in its target motif sequence, motif distribution, and sensitivity to DNA methylation. While methylation of the CpG within the CpG-containing motif may help recruit Cdx2 to developmental loci, the demethylation of this CpG post-development may deter ectopic Cdx2 binding at these sites in the adult cells (Fig. 2h).

Progressive activation of enhancers by Cdx2 through clustered binding and corecruitment of Ctcf and Hnf4

Given the gradual increase of enhancer-based Cdx2 binding through development (Fig. 1c), we examined the grouping of its binding and asked if Cdx2 facilitates corecruitment of TFs with different sensitivities to chromatin features. For the development and adult cell-specific target sites of Cdx2, (Dev 1–3 and Adult 1–3, Fig. 1a), we asked if there were other Cdx2 binding instances within 50 kb. While 60% of the embryonic sites (Dev 1 and Dev 2) had no other Cdx2 binding within 50 kb, 71% of the adult bound sites (Adult 1 and Adult 2) had at least one other site nearby (Fig. 3a), and this clustering was apparent when considering Cdx2 binding only at the enhancers or when considering Cdx2 peaks only within 10 kb distance from each other (Supplementary Fig. 3a, b). Cdx2 binding around genes expressed in adult homeostatic cells such as *Krt19* was accompanied by significantly more clustering of Cdx2 in adult villus cells as compared to developmental sites (groups of median 3 Cdx2 peaks vs 2 peaks at adult and developmental loci, respectively, Fig. 3c). Accordingly, 6492 adult loci had 2 or more Cdx2 binding events surrounding them as compared to 3283 developmental loci ($p < 0.001$) and genes near such sites with clustered Cdx2 binding (TSS within 25 kb) showed high expression in adult or developmental cells, respectively (Supplementary Fig. 3c). In

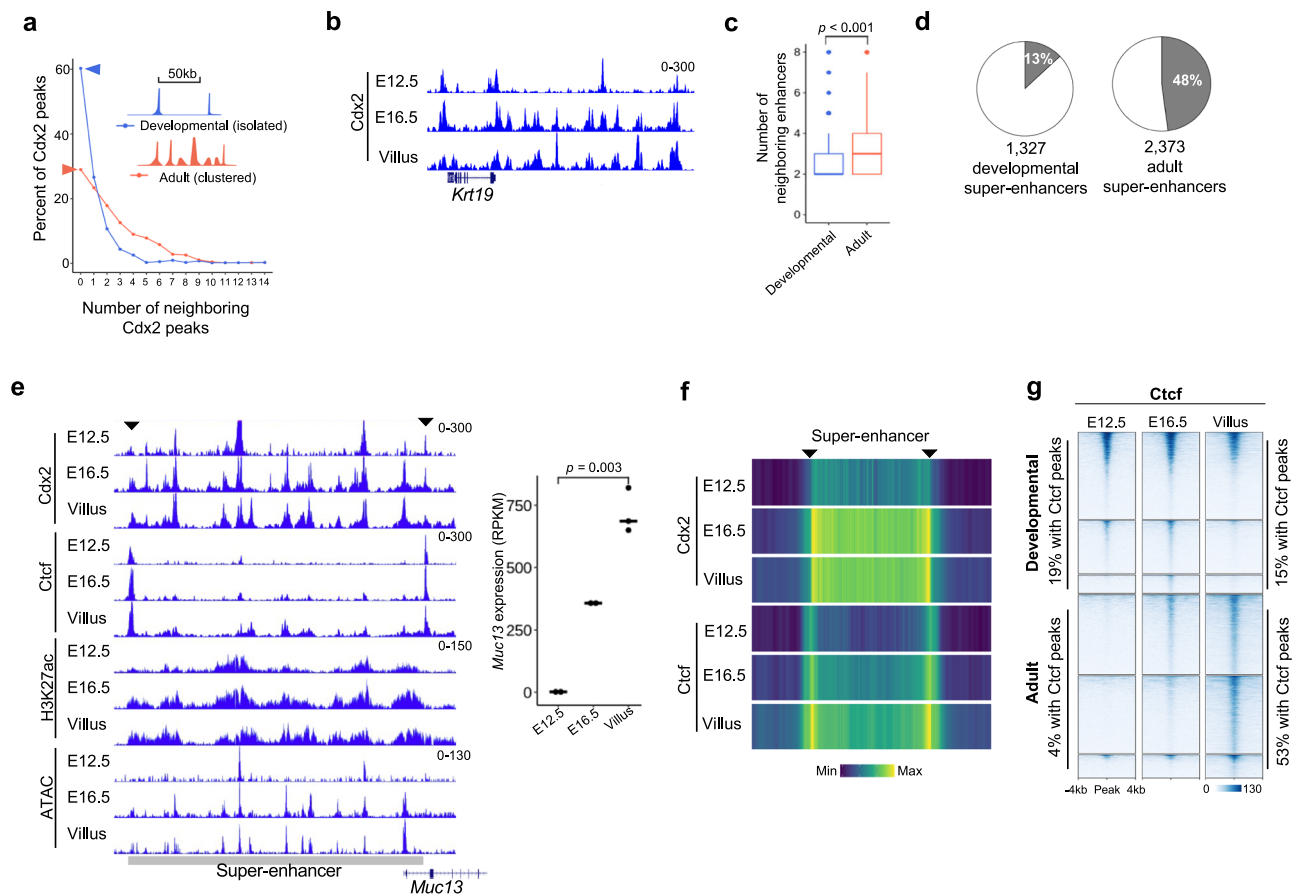


Fig. 3 | Cdx2 facilitates establishment of adult homeostatic super-enhancers by directing Ctcf recruitment. **a** Percentage of developmental (blue) or adult (red) cell-specific Cdx2 bound loci (as in Fig. 1a) with designated number of other Cdx2 peaks nearby (± 50 kb); drawing in the inset represents isolated Cdx2 peaks (not within 50 kb of each other) in developing cells, while adult cells show multiple neighboring peaks in the same area. Large number of (60%) of dynamic Cdx2 binding events are solitary in developing epithelium (blue arrowhead) as opposed to only 30% in adult cells (red arrowhead). **b** Representative genomic tracks showing clustered binding of Cdx2 that grows through development near gene *Krt19*. **c** boxplot showing number of neighboring (± 50 kb) dynamic Cdx2 bound enhancers (sites in Fig. 1a) surrounding 7806 developmental (blue) or 8901 adult (red) cell-specific Cdx2 binding loci ($p < 0.001$, right-sided t -test). The horizontal bars within the box plots represent medians, the tops and bottoms of the boxes represent the 75th and 25th percentiles, respectively, and the upper and lower whiskers extend to the maximum and minimum values, respectively. The outliers are indicated by dots. **d** Percentage of 1327 developmental (E12.5 and E14.5) and

2373 adult super-enhancers that have dynamic Cdx2 binding during epithelial maturation. **e** Genome browser view of Cdx2, Ctcf, H3K27ac, and ATAC-seq signals at adult-specific Cdx2 cluster at a super-enhancers near *Muc13* gene; black arrowheads indicate the super-enhancer borders with Cdx2 and Ctcf cobinding. Plot on the right shows gain in expression of *Muc13* through development. The horizontal bars represent medians, each dot represents individual values ($n = 2$ for E12.5 and E16.5, $n = 3$ for Villus) ($p = 0.003$, right-sided Wilcoxon test). **f** Heatmap showing Cdx2 and Ctcf signals at super-enhancers in E12.5, E16.5 and adult villus. Super-enhancers boundaries are shown with black arrowheads. **g** Heatmap showing Ctcf signal at developmental and adult-specific Cdx2 binding sites (as in Fig. 1a) in E12.5, E16.5, and adult villus cells; the regions are displayed in decreasing order of Ctcf signal. 19% and 15% loci with development-specific Cdx2 binding show Ctcf occupancy at E12.5 and adult villus samples, respectively. 4% and 53% loci with adult-specific Cdx2 binding show Ctcf binding in E12.5 and adult villus samples, respectively. Source data are provided as a Source Data file.

combination with the increased enhancer-based binding of Cdx2 in the adult cells (Fig. 1c), these data suggest that Cdx2 binding in groups or clusters of target loci drives the establishment of adult enhancer patterns.

Super-enhancers are known to be hubs for coordinate binding of multiple TFs that facilitate wide-scale deposition of active enhancer mark H3K27ac to induce linked gene expression³⁴. Using published data, we identified super-enhancers at developmental time points (1327 super-enhancers combined at E12.5 and E14.5) and adult villus cells (2373, super-enhancers)^{15,19}. While 13% of the developmental Cdx2 binding sites (Dev 1–3, Fig. 1a) were located within super-enhancers, 48% of the loci with adult cell-specific Cdx2 binding (Adult 1–3, Fig. 1a) overlapped with super-enhancers (Fig. 3d). Cdx2 showed binding at many of the super-enhancer boundaries at E12.5 in comparison to the binding within the body of super-enhancers. This binding at the boundaries showed rapid gain at E16.6 that was sustained in adult cells,

while Cdx2 binding within the body of super-enhancers increased gradually from E12.5 to adult cells, suggesting that the binding progressively spread into the body of super-enhancers during development (Figs. 3e, f and S3d, e). Super-enhancers near genes *Krt19* and *Tmc5*, which already have Cdx2 binding at E12.5, show spreading of Cdx2 binding at different times, E16.5 and adult cells, respectively (Fig. S3f). This indicates a role of Cdx2 in defining adult super-enhancer borders through binding during early development.

Ctcf is a TF involved in chromatin organization and functions to form topologically associated domains (TADs) for genome compartmentalization and promoter-enhancer interactions using cohesion-based DNA looping³⁵. Various chromatin controls of Ctcf recruitment are being discovered, which include sensitivity to DNA methylation at CpGs within its motif (Supplementary Fig. 3g), and recruitment to its genomic targets through interaction with other TFs^{14,35–37}. As Ctcf is often bound at the super-enhancer boundaries³⁸, we conducted

CUT&RUN assays to determine the dynamics of Ctf binding during intestinal development and its dependence on Cdx2-based corecruitment. While only 4% of the adult cell-specific Cdx2 sites (Adult 1–3) had Ctf binding at E12.5, 53% of these loci gained Ctf binding during development (Fig. 3g). On the other hand, about 19% percent of the developmental Cdx2 target sites at E12.5 had Ctf co-binding and 77% of this binding was maintained in adult cells with 15% of all developmental sites (Dev 1–3) showing co-binding with Cdx2 (Fig. 3g). These data show a more prevalent interaction between Ctf and Cdx2 at adult enhancers that temporally grows through development as compared to limited cooperation at early developmental target sites starting at E12.5, which remains stable in the adult cells. Moreover, Ctf binding at super-enhancer boundaries follows Cdx2 binding temporally as it starts from either or both sides of the super-enhancers in embryonic cells (E12.5) and spreads across the body of super-enhancers (Fig. 3e, f and Supplementary Fig. 3f). This suggests that Cdx2 binding may be required to prime these loci for Ctf binding in order to establish the adult super-enhancers during development. On the other hand, when we conducted CUT&RUN for Hnf4, an intestinal TF important for homeostatic gene activation in the adult tissue, it showed high cobinding with Cdx2 only at the adult villus specific Cdx2 binding sites (Supplementary Fig. 3h). Moreover, at super-enhancers, Hnf4 shows high signal through the body of the super-enhancers (Supplementary Fig. 3i, j) in contrast to the stronger Ctf binding at the boundaries (Figs. 3e, f, and S3f, j). As Hnf4a also binds many genomic targets in the intestinal epithelium without Cdx2³⁹, we used published data to ask if its cobinding with Cdx2 at super-enhancers would be preferentially compromised in absence of Cdx2. Indeed, in *Cdx2*^{-/-} mouse villus cells Hnf4a binding was reduced in comparison to *Cdx2*^{+/+} cells and Hnf4a signal was significantly low at super-enhancers as compared to the sites where its binding is independent of Cdx2 (Supplementary Fig. 3j, k). Thus, temporal dynamics of Cdx2 binding in intestinal epithelium may facilitate step-wise establishment of developmental and adult enhancers through systematic corecruitment of two different TFs, Ctf and Hnf4a, respectively.

Many master transcription factors bind in a clustered pattern, forming genomic sites with broad chromatin features like histone marks and accessibility, such as the super-enhancers³⁴. These sites often display cell cycle-dependent changes in chromatin accessibility, likely linked to the dynamic binding of tissue-specific and chromatin-organizing TFs such as CTCF, which may regulate the balance between cell cycle progression and exit to facilitate terminal differentiation^{40,41}. While intestinal villi represent post-mitotic (non-cycling) cells, the intestinal epithelium consists of ISCs that divide every 3–5 days and fast-cycling transit amplifying (TA) cells that divide 2–6 times every day (Supplementary Fig. 1a)^{42,43}. We used a genetically engineered mouse model that labels all ISCs with green fluorescence (EGFP) based on the expression of the marker gene *Lgr5* (*Lgr5*^{EGFP-IRES-DTR}, Supplementary Fig. 1a)⁴⁴, and purified ISCs and TA cells from the crypts using flow cytometry-based sorting (FACS). Cell cycle analysis of the various cell types showed that only 13% of villus cells were in synthesis and mitosis (S and G2/M) compared to 39% and 35% of the ISCs and TA cells, respectively (Supplementary Fig. 4a). To determine if the clustered Cdx2 binding in the villus cells (Fig. 3a–e) is linked to cell cycle changes in the epithelium, we performed Cdx2 CUT&RUN on cycling ISCs and TA cells and studied changes across the cell types (Supplementary Fig. 4b, c). At the loci with Cdx2 clustered binding, we see a small and gradual increase in Cdx2 signal from ISC to TA cells and further into villus cells. As expected, this clustered binding is absent from E12.5 epithelium and is detected in E16.5 (Supplementary Fig. 4b, c). Notably, villus super-enhancers called independently of the Cdx2 clustered binding (Fig. 3f) show consistent gain of Cdx2 binding going from stem cells to fully differentiated villus cells (Supplementary Fig. 4d). These observations suggest that clustered binding of Cdx2 is

not dynamic across cells with altered cycling properties but may instead control genes facilitating homeostatic cell differentiation and gene activation. Accordingly, genes linked to these sites (TSS within 25 kb of Cdx2 peaks) are involved in metabolic processes and intestinal absorption (Supplementary Fig. 4e). Moreover, CUT&RUN for Ctf in ISC and TA cells showed that Ctf may be corecruited to many such clustered Cdx2 bound loci and super-enhancers engaged in intestinal functions, where its binding is not dynamic across cells with varied cell cycle patterns (Supplementary Fig. 4f, g).

Loss of PRC2 action in adult epithelium leads to altered DNA methylation and Cdx2 binding

Although in vitro studies have noted Cdx2's propensity to bind the methylated CpG-containing motif⁴, its relevance to Cdx2 recruitment in cells and functional implications for gene expression or cellular identity remains unknown. To test this critically, we used two experimental systems—a mouse model and a cell line—in which we determined if induced DNA methylation at the CpG-containing Cdx2 motifs in adult cells could promote ectopic Cdx2 recruitment to development-specific target sites. In mammalian tissues, many developmental genes lose DNA methylation at promoters when they are expressed during embryonic growth and H3K27me3 deposition in adult cells protects these genes from ectopic expression in homeostatic conditions^{29,30}. Interestingly, while loss of DNA methylation in mouse ES cells causes wide-scale redistribution of H3K27me3, loss of PRC2 action and H3K27me3 due to deletion of *Eed*, a protein that stabilizes the complex, results in relatively modest change in DNA methylation^{30–32}. We and others have shown that genetic deletion of *Ezh1/2*, the enzymatic subunits of the PRC2 complex, or *Eed* cause loss of H3K27me3 in various adult tissues including intestinal epithelium, leading to reactivation of developmental genes^{29,30,45}. Accordingly, we have previously used induced deletion of *Eed* in the entire mouse intestinal epithelium or specifically in the intestinal stem cells (ISCs) and characterized the temporal impact of H3K27me3 loss on promoter and enhancer activity and gene expression^{15,29,43}. Upon deletion of *Eed*, adult intestinal epithelial cells derepressed many developmental genes due to loss of H3K27me3-based suppression at their promoters²⁹. Many adult enhancers marked by hypomethylated DNA gained enhancer-associated histone modifications H3K4me1 and H3K27ac, further leading to expression of linked genes¹⁵. Notably, we observed that many TFs, which are exclusively expressed during development and repressed by H3K27me3 in adult cells, were re-expressed in adult *Eed*-null cells and recruited to the developmental enhancers, potentially facilitating ectopic enhancer activation and gene expression¹⁵. On the other hand, how loss of H3K27me3 alters DNA methylation in vivo and if it further influences TF action and promoter or enhancer function remains relatively unstudied. We used the well-established in vivo model system of intestinal epithelium-specific *Eed* deletion to investigate how altered chromatin features, particularly DNA methylation, may influence Cdx2 binding in adult epithelial cells.

We deleted *Eed* across the adult intestinal epithelium (2–6-month-old mice) using 5 intraperitoneal injections of tamoxifen (1 mg/dose) on consecutive days in *Eed*^{fl/fl}; *Villin-Cre*^{ER-T2} mice as before^{15,29,46} (Fig. 4a) and analyzed the genome-wide changes in DNA methylation using whole genome bisulfite sequencing (WGBS, Fig. 4b). By comparing the methylation at 16,392,619 consensus CpGs between *WT* and the *Eed*^{-/-} cells, we determined differentially methylated regions (DMRs, minimum 40% change in DNA methylation and $q < 0.001$); 82% of the DMRs (23,978 sites) showed gain of DNA methylation. Importantly, 95% of these DNA methylation gains were at enhancers, while 1222 promoters showed hypermethylation in *Eed*^{-/-} cells (Fig. 4b). In a parallel approach, in *WT* cells we identified unmethylated regions (UMRs, with CpG methylation <10%), which mostly represent gene promoters, and low methylated regions (LMRs, with CpG methylation >10% and <60%) that represent enhancers^{15,47,48}. While many

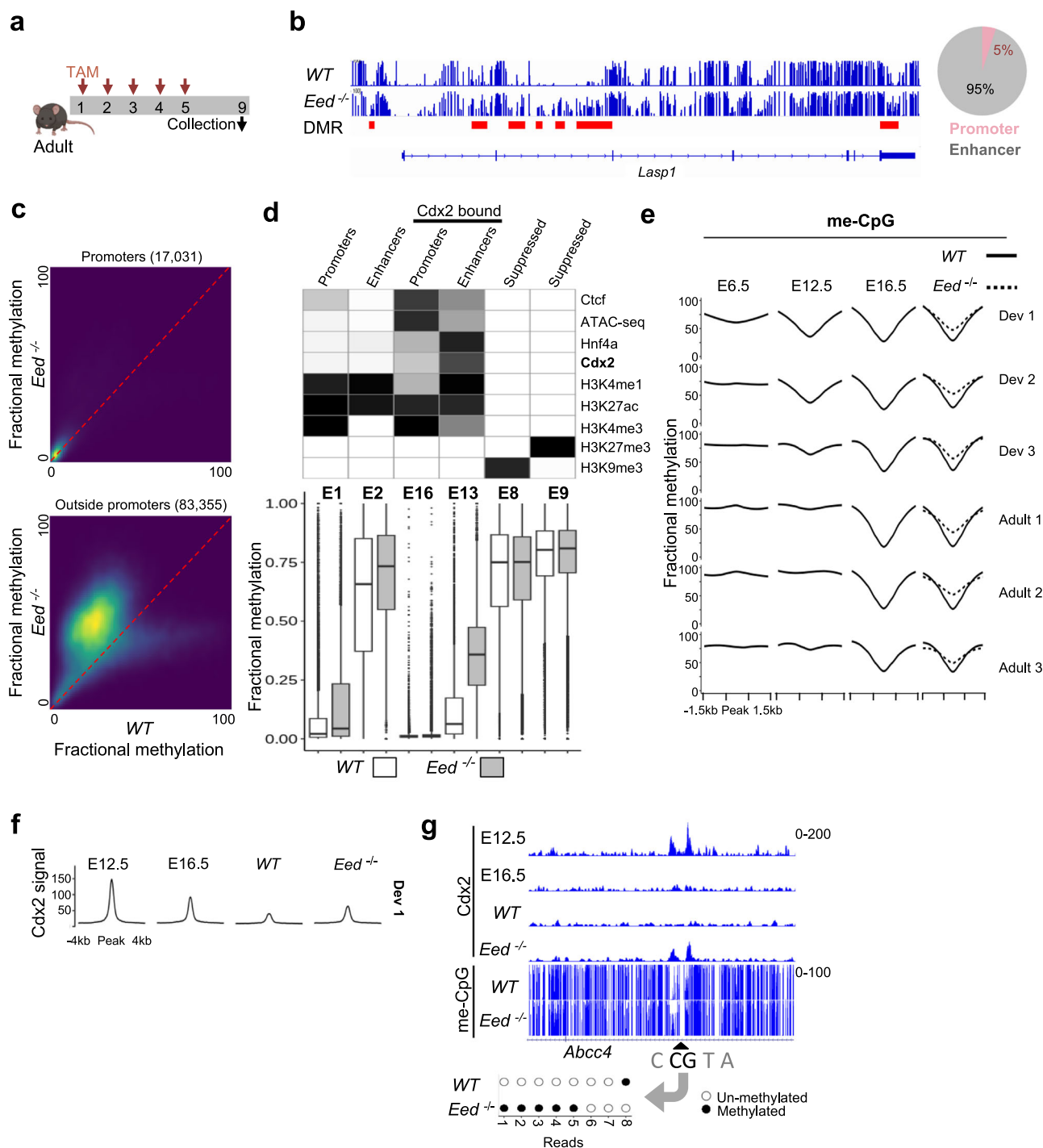


Fig. 4 | Loss of PRC2 activity causes gain of DNA methylation at enhancers leading to Cdx2 recruitment. **a** Experimental schematic showing 5 intraperitoneal injections of tamoxifen (TAM) on consecutive days cause deletion of *Eed* across the intestinal epithelium in adult mouse (2–6 month old); epithelial cells are collected 4 days after the last injection (experimental day 9). **b** Genome browser view showing gain of CpG methylation in *Eed*^{-/-} epithelium at multiple enhancer loci near *Lasp1* gene that are identified as DMRs; promoter region remains unmethylated. Pie chart showing percentages of DNA methylation gains at promoters and enhancers. **c** Density plot showing DNA methylation at promoter (top panel) and non-promoter (bottom panel) linked UMRs and LMRs; LMRs (mostly representing enhancers) have the most prevalent gain of methylation in *Eed*^{-/-} cells. **d** Heatmap showing genomic states identified by CHROMHMM based on TF binding and histone modifications corresponding to promoter, enhancer, and repressed chromatin; Cdx2 occupied enhancers (H3K4me1+ and H3K27ac+) particularly show gain in DNA methylation in *Eed*^{-/-} cells, while promoters (H3K4me3+) and repressed

chromatin regions (H3K27me3+ or H3K9me3+) show relatively stable unmethylated and methylated status, respectively. The horizontal bars within the box plots represent medians, the tops and bottoms of the boxes represent the 75th and 25th percentiles, respectively, and the upper and lower whiskers extend to the maximum and minimum values, respectively. The outliers are indicated by dots. **e** DNA methylation profiles at developmental and adult specific Cdx2 binding sites (Dev 1–3 and Adult 1–3, respectively) showing the gain of methylation upon loss of PRC2 activity (*Eed*^{-/-}). **f** Profile plot of Cdx2 signal show ectopic recruitment of Cdx2 at developmental loci (Dev 1) in adult *Eed*^{-/-} cells in concordance with the gain of DNA methylation in (e). **g** Genomic tracks showing *Abcc4* gene locus with CpG-containing Cdx2 motif bound by the TF only during development, where gain of methylation at the Cdx2 binding locus in *Eed*^{-/-} leads to recruitment of the TF; dot plot at the bottom shows methylated (solid black dots) or unmethylated (white dots) status of the CpG in Cdx2 motif (4th position) in 8 independent sequencing reads from WT and *Eed*^{-/-} WGBS data. Source data are provided as a Source Data file.

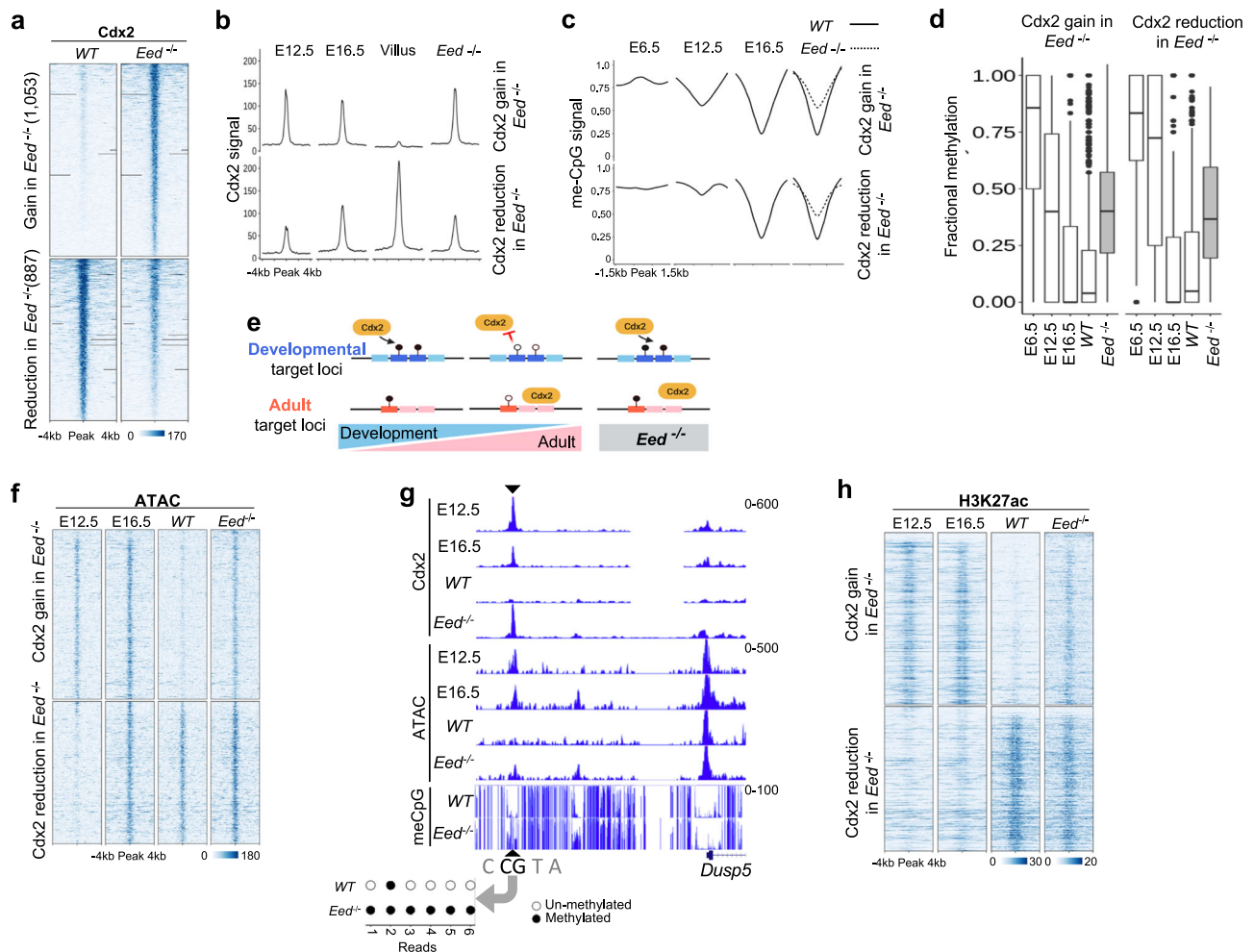


Fig. 5 | Altered Cdx2 binding in adult cells revives developmental Cctf recruitment. **a** Heatmap showing significant ($q < 0.05$, $1.5\times$) gain or reduction of Cdx2 binding in *Eed*^{-/-} cells as compared to the native villus cells (WT). **b** Profile plots showing Cdx2 signal at loci with gain and loss of Cdx2 binding in *Eed*^{-/-} cells from embryonic (E12.5), fetal (E16.5), adult villus, and *Eed*^{-/-} epithelium. **c** Profile plots showing DNA methylation signal at loci with gain or reduction of Cdx2 binding in *Eed*^{-/-} cells from embryonic (E6.5 and E12.5), fetal (E16.5), adult villus, and *Eed*^{-/-} epithelium. **d** Fractional methylation change at CpG at the 4th position within Cdx2 motif shows gain of methylation in *Eed*^{-/-} cells at the sites (panel a) with altered Cdx2 binding. The horizontal bars within the box plots represent medians, the tops and bottoms of the boxes represent the 75th and 25th percentiles, respectively, and the upper and lower whiskers extend to the maximum and minimum values, respectively. The outliers are indicated by dots. **e** A model showing DNA methylation-based Cdx2 binding at developmental target loci; Cdx2 is recruited to developmental sites upon gain of methylation of CpG (at 4th

position) within the Cdx2 motif in adult *Eed*^{-/-} cells. **f** Heatmap showing chromatin accessibility dynamics at sites with gain or reduction of Cdx2 binding in *Eed*^{-/-} cells as compared to WT cells (as in **a**); chromatin accessibility at developmental Cdx2 target loci is lost with lack of Cdx2 binding in adult cells (as seen in **a**) and rebinding of Cdx2 to such loci in *Eed*^{-/-} cells reinduces open chromatin formation. **g** Genomic tracks showing recruitment of Cdx2 to developmental site in *Eed*^{-/-} cells accompanied by gain of chromatin accessibility and increase in DNA methylation at the Cdx2 binding locus; dot plot at the bottom shows methylated (solid black dots) or unmethylated (white dots) status of the CpG in Cdx2 motif (4th position) in 6 independent sequencing reads from WT and *Eed*^{-/-} WGBS data. **h** Heatmap showing alterations in active enhancer mark H3K27ac at the sites with Cdx2 gain or reduction in *Eed*^{-/-} cells as compared to WT cells (as in **a**); enhancer activity from developmental Cdx2 target loci is lost with lack of Cdx2 binding in adult cells (as seen in **a**) and rebinding of Cdx2 to such loci in *Eed*^{-/-} cells reactivates these enhancers. Source data are provided as a Source Data file.

promoters maintained an unmethylated state upon loss of *Eed* (UMRs), we observed a gain of methylation at many enhancers (LMRs) (Fig. 4c). Additionally, we divided the genome into different states based on various chromatin features including activating and repressive histone modifications and TF binding using ChromHMM algorithm⁴⁹. This revealed a gain of DNA methylation at enhancers in *Eed*^{-/-} cells, particularly ones with Cdx2 binding (Fig. 4d and Supplementary Fig. 5a, b). Together these data show that the gain of DNA methylation upon loss of *Eed* is focused at hypomethylated adult enhancers rather than a global hypermethylation of the genome. The gain of DNA methylation at enhancers was apparent at developmental Cdx2 binding loci (Dev 1–3), recapitulating a developmental CpG methylation state in adult cells (Figs. 4e and Supplementary Fig. 5c). To examine the effect of this

remethylation at developmental enhancers on Cdx2 binding, we conducted CUT&RUN for Cdx2 using adult *Eed*^{-/-} villus cells. Cdx2 binding at developmental loci (Dev 1–3) was increased in *Eed*^{-/-} villus cells (Fig. 4f, g). Comparison of Cdx2 binding in *Eed*^{-/-} cells with that in WT cells identified 1053 Cdx2 binding events unique to *Eed*^{-/-} cells ($q < 0.01$, $1.5\times$ gain, Fig. 5a). A significant number of these Cdx2 gains in *Eed*^{-/-} overlapped with developmental sites (74%, hypergeometric p value < 0.001) and the Cdx2 signal at these loci in *Eed*^{-/-} cells was equivalent to that in the embryonic cells (E12.5), suggesting reinstatement of the developmental binding of Cdx2 (Fig. 5b). Gene ontology analysis showed that the sites with Cdx2 recruitment in *Eed*^{-/-} cells were associated with developmental genes regulating tissue patterning, further underscoring the reestablishment of

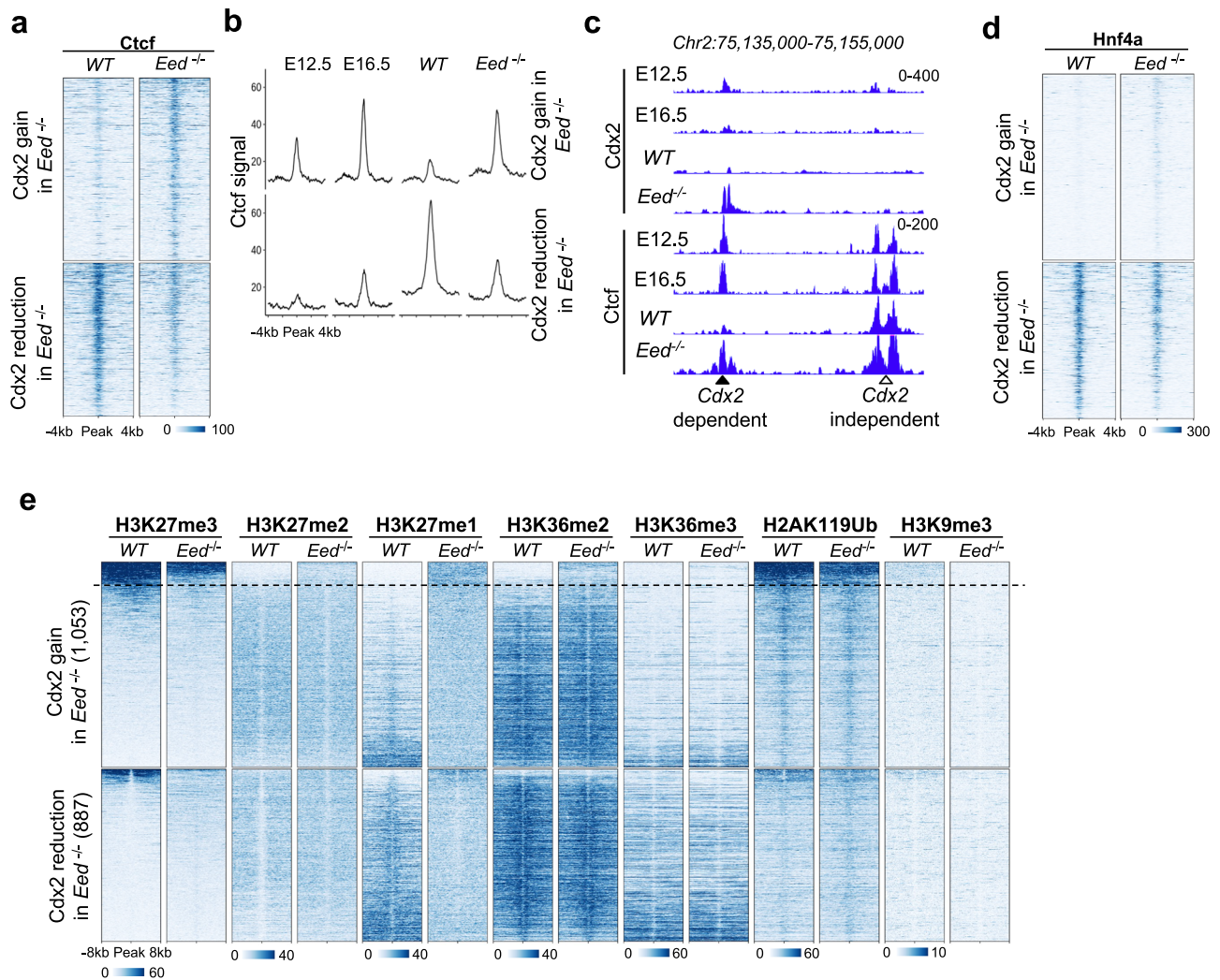


Fig. 6 | Perturbed Cdx2 binding resulting from DNA methylation changes leads to altered cobinding of TFs. a Heatmap showing Ctf signal in adult villus and *Eed*^{-/-} epithelium at loci with gain or reduction of Cdx2 binding (as in Fig. 5a). **b** Profile plots showing Ctf signal from embryonic (E12.5), fetal (E16.5), adult villus, and *Eed*^{-/-} epithelium at the loci with gain or reduction of Cdx2 binding in *Eed*^{-/-} cells. **c** Genomic tracks showing corecruitment of Ctf with Cdx2 in adult *Eed*^{-/-} cells reviving the early embryonic binding at E12.5; a nearby locus shows

stable binding of Ctf independent of Cdx2 occupancy. **d** Heatmap showing Hnf4a signal in adult villus and *Eed*^{-/-} epithelium at loci with gain or reduction of Cdx2 binding (as in Fig. 5a). **e** Heatmaps showing signal of various histone modifications in adult epithelial cells at sites with Cdx2 gain or reduction in *Eed*^{-/-} cells (as in Fig. 5a); regions in each cluster are arranged in decreasing order of the repressive H3K27me3 modification signal.

developmental Cdx2 binding patterns in adult cells. Contrary to this, site with reduced Cdx2 binding were associated with genes involved in intestinal homeostatic functions (Supplementary Fig. 6a). Moreover, the DNA methylation level in *Eed*^{-/-} cells at these sites was reminiscent of embryonic methylation level (E12.5) (Fig. 5c and Supplementary Fig. 6b). We could detect the CpG-containing motif at twice as many loci (37%) with Cdx2 gain in the *Eed*^{-/-} cells as compared to sites with reduced Cdx2 binding (17%), and the CpG at the 4th base position in the Cdx2 motif showed gain of methylation (Fig. 5d). Thus, our results strongly argue that the demethylation at developmental Cdx2 binding sites in adult cells may be sufficient to avoid its binding to these sites post development, and replenishing the embryonic DNA methylation state in adult cells can recruit Cdx2 to the developmental enhancers in adult epithelial cells (Fig. 5e). Importantly, the Cdx2 binding is necessary for open chromatin status at the developmental loci, as we see loss of ATAC-seq signal at these loci in adult cells, while Cdx2 binding in adult *Eed*^{-/-} cells results in re-opening the chromatin at these sites (Fig. 5f, g), highlighting Cdx2's pivotal function in developmental enhancer reactivation. This central role of Cdx2 is further

corroborated by the gain of H3K27ac signal indicating ectopic activation of the developmental enhancers in adult *Eed*^{-/-} cells (Fig. 5h)¹⁹.

To understand the effect of this flux in DNA methylation and Cdx2 binding on TF corecruitment, we conducted CUT&RUN for Ctf. This showed that Cdx2 promoted Ctf binding to developmental enhancers in *Eed*^{-/-} cells (Fig. 6a). Indeed, Ctf signal at the loci with Cdx2 gains in *Eed*^{-/-} cells reached almost the levels of its binding in native epithelial cells during developmental stages (Fig. 6b). Moreover, 1410 sites with significant gain and 511 with reduction of Ctf binding in *Eed*^{-/-} cells ($q < 0.05$, 1.5 \times) were associated with corresponding gain and reduction of Cdx2 binding ($p < 0.001$, Fig. 6c, Supplementary Fig. 7a, b). The sites where Ctf is corecruited with Cdx2 in adult *Eed*^{-/-} cells were highly linked to genes related to embryogenesis and pattern formation as compared to the loci that Ctf binds without Cdx2, which showed association with genes involved in immune response and other homeostatic functions (Supplementary Fig. 7c). Together, these data suggest that Cdx2 based priming and corecruitment controls Ctf genomic binding and sustained presence at specific genomic sites in intestinal epithelial cells. On the other hand, while Hnf4a gets

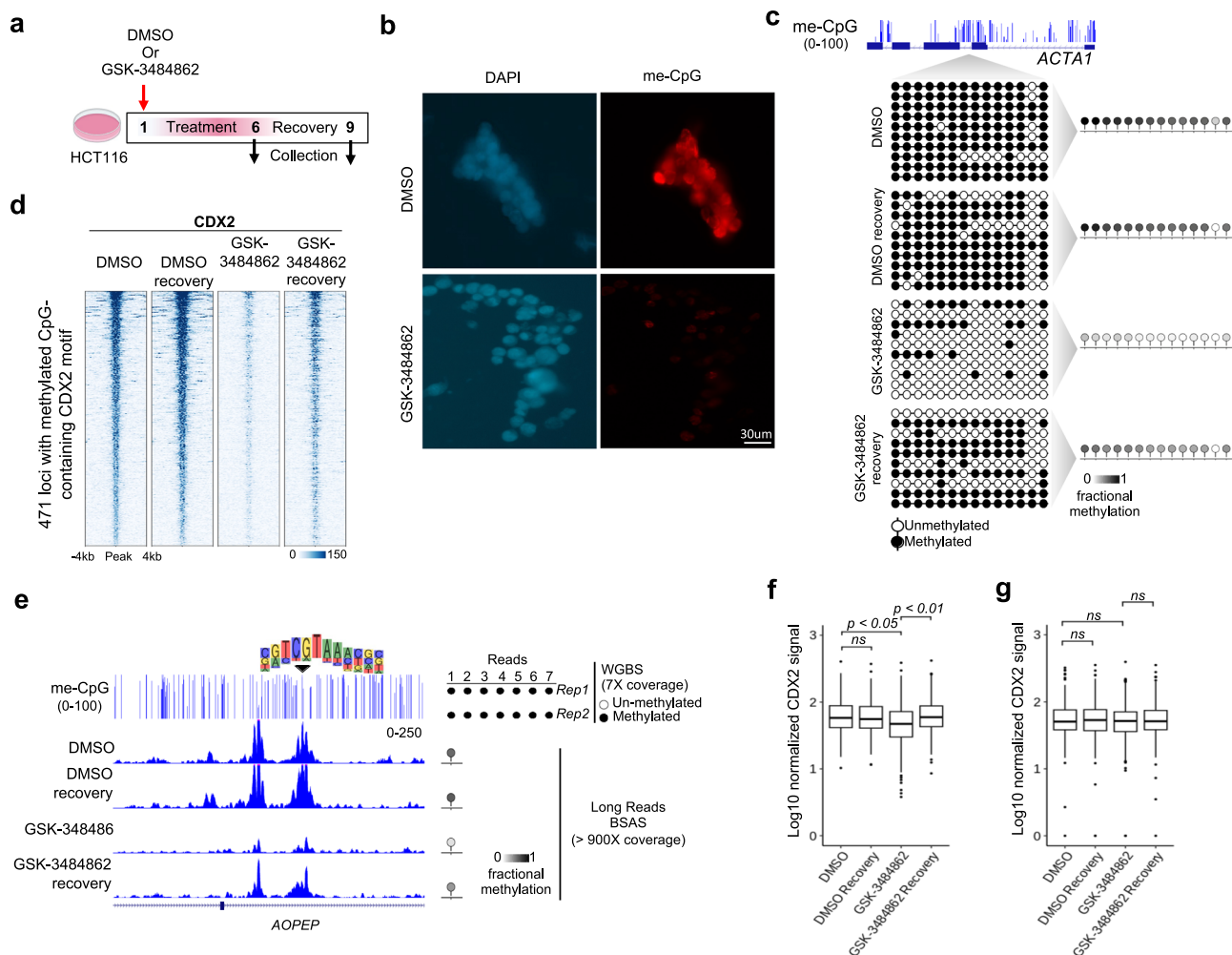


Fig. 7 | Direct inhibition of DNA methylation causes loss of Cdx2 binding and remethylation at binding loci allows its recruitment. **a** Experimental schematic showing treatment of HCT116 cells with DNA methylation inhibitor GSK-3484862 or DMSO (control) for 6 days in order to eliminate CpG methylation, followed by 3 days of recovery without treatment to allow remethylation. **b** Fluorescence micrographs showing loss of DNA methylation in cells upon treatment with inhibitor GSK-3484862 for 6 days, (4 independent experiments). **c** Representative genomic track showing fractional methylation of CpGs within *ACTA1* gene body in HCT116 cells. Lollipop plots show 14 consecutive CpGs within 10 random representative amplicon sequences (from >100× coverage on nanopore amplicon bisulfite sequencing) and their methylation status upon DMSO or GSK-3484862 treatment and 3 days of recovery; average methylation of all sequenced amplicons (>100× coverage) is represented for all treatment conditions in the right panel. **d** Heatmap showing loss of CDX2 binding at loci with CpG-containing CDX2 motif in HCT116 cells treated with DNA methylation inhibitor as compared to DMSO (control); upon removal of the inhibitor, CDX2 is recruited back to these loci within 3 days. **e** Representative genomic tracks showing loss of CDX2 binding upon

elimination of DNA methylation in HCT116 cells through treatment with GSK-3484862 as opposed to the stable binding upon treatment with DMSO (control); removing the inhibitor from the cell culture media causes recovery of the CDX2 binding. Lollipop plots (right panel) represent methylation level at the CpG within CDX2 motif in HCT116 cells (WGBS, 7 reads from 2 replicates) and cells treated with DMSO or GSK-3484862 treated and after recovery (average methylation from >900 reads, nanopore amplicon bisulfite sequencing). **f** Box and whisker plots showing significant loss of CDX2 binding at 471 loci with CpG-containing CDX2 binding motif upon treatment with DNA methylation inhibitor and recovery of this binding upon removal of the inhibitor (right-sided Wilcoxon test; ns , $p > 0.05$); there is no significant change in CDX2 binding at 500 randomly chosen loci without the CpG-containing motif (**g**) (right-sided Wilcoxon test; ns , $p > 0.05$). The horizontal bars within the box plots represent medians, the tops and bottoms of the boxes represent the 75th and 25th percentiles, respectively, and the upper and lower whiskers extend to the maximum and minimum values, respectively. The outliers are indicated by dots. Source data are provided as a Source Data file.

corecruited with Cdx2 at a few loci in *Eed*^{-/-} cells, its binding shows some reduction in signal at the loci with reduced Cdx2 binding (Fig. 6d), suggesting a role of Cdx2 in maintenance of Hnf4a binding in adult cells. To explore whether chromatin features beyond CpG methylation influence Cdx2 recruitment at developmental loci in adult *Eed*^{-/-} cells, we examined histone modification changes in these cells. We see no significant change in H3K9me3, H2AK119ub, or H3K36me3, and a small number of loci (156 regions) with high H3K27me3 signal in the *WT* adult cells showed a gain of H3K36me2 signal in the *Eed*^{-/-} cells (Fig. 6e and Supplementary Fig. 7d), suggesting that remethylation of CpG at the 4th base position in the Cdx2 motif (Supplementary Fig. 2a) is a major driver of its recruitment to developmental target loci.

Biochemical modulation of DNA methylation causes altered CDX2 binding

To further establish the effect of DNA methylation on CDX2 binding, we altered CpG methylation more directly using the well-established and specific inhibitor GSK-3484862⁵⁰ in HCT116 cells, a colorectal carcinoma cell line with a well-characterized epigenome including DNA methylation (Fig. 7a)⁵¹. We first analyzed CDX2 binding in duplicate experiments using CUT&RUN and observed Cdx2 occupancy at 10,197 loci (MACS2 peaks, $q < 0.01$), with the CpG-containing motif detected at 2464 of these (FIMO, $p < 0.005$). Using triplicate WGBS data and a stringent cut-off of 5× coverage at the CpG in 4th position of the CDX2 bound motifs in

untreated cells⁵¹. Exposure to 10 μ M GSK-3484862 for 6 days caused global loss of DNA methylation as detected by immunofluorescence for 5-methyl Cytosine (Fig. 7b). Additionally, we conducted bisulfite conversion of the genomic DNA and quantified DNA methylation changes at multiple genomic loci using nanopore sequencing of amplicons, which provided coverage greater than 100X for CpGs allowing precise determination of CpG methylation. Sequencing at CpG islands, including one within the *ACTA1* gene, validated loss of DNA methylation upon GSK-3484862 treatment and its recovery 3 days after drug removal from the cell culture media⁵² (Fig. 7c and Supplementary Fig. 8a). CUT&RUN assays revealed a significant loss of CDX2 binding at sites with the CpG-containing motif in cells treated with GSK-3484862 and amplicon sequencing confirmed the loss of DNA methylation at CpG within the CDX2 motif (Fig. 7d, e and Supplementary Fig. 8b). Importantly, removal of the inhibitor from culture media and the resulting gain in the CpG methylation caused rebinding of CDX2, demonstrating a DNA methylation-based recruitment of CDX2 (Fig. 7d–f); 500 randomly chosen Cdx2 peaks without a CpG within the motifs showed no significant alteration in CDX2 binding upon inhibitor treatment or its removal (Fig. 7g). These results demonstrate that the recruitment and dynamics of CDX2 binding is controlled by its motif distribution and alterations of CpG methylation.

To address the role of cooperative TF binding in Ctf recruitment, we examined the impact of Cdx2 loss on Ctf binding using CUT&RUN for CTCF in HCT116 cells. In cells treated with the DNA methylation inhibitor, loss of CDX2 binding at CpG-containing loci corresponded with CTCF loss, while CTCF binding persisted at CDX2-independent sites (Supplementary Fig. 8c, d). These data reveal that co-recruitment of Ctf with Cdx2 is prevalent at select intestinal enhancers.

Discussion

Lineage-defining TFs establish tissue patterning and function through the establishment of tissue-specific gene expression networks. To do this, they rapidly activate other TFs, which in turn allows the expansion of the lineage and cell type-specific enhancer landscape over time^{3,4}. Our data show that Cdx2 activates a proportionally high number of TF genes such as *Sox4* during early development (Fig. 1d), and it binds directly to promoters of such TF genes and other developmental targets. By activating other TFs through simple promoter-based control during development, TFs may rapidly increase the potential to establish lineage-specific enhancers necessary for more complex gene control required by adult homeostatic tissue functions. Indeed, recent studies looking at TF action in developmental and adult erythroblast cells show that simple promoter-centric regulation through Gata1 targets embryonic-specific genes, while increasing combinatorial enhancer-driven gene control through Myb in adult cells⁵³. We show that such promoter-based targeting of Cdx2 during intestinal development may be programmed in elevated presence of the CpG-containing motif at the developmental target sites in comparison to the canonical motif without CpG. Thus, for lineage-defining TFs like Cdx2 with variable affinity for different motifs, the relative abundance of such motifs may be a potential way to shift from rapid promoter-based regulation in development to a more nuanced enhancer-driven control in adult tissue. Interestingly, the Cdx2 loci specifically bound in development show more conservation across species as compared to the loci bound only in adult cells. Together, these data suggest a conserved molecular mechanism based on DNA methylation for TF recruitment to developmental targets.

Enhancers are cis-regulatory elements that control the spatio-temporal and quantitative expression of target genes and are believed to act through physical contact with the target gene promoters. Most enhancers consist of dense clusters of TF binding sites⁵⁴, where cell-type-specific TFs help assemble complexes containing coregulators, architectural proteins like CTCF and Cohesin, chromatin modifiers,

and RNAPII. Super-enhancers show clustered binding of many TFs and significantly impact transcription of genes that are highly expressed in tissue-specific manner⁵⁵. In this light, it is possible that the clustered binding of Cdx2, which grows during development as the adult epithelial cells are formed and are also observed at super-enhancers (Fig. 3a, b, f), may allow Cdx2 to recruit cofactors such as Ctf and contribute to the formation of enhancer-promoter contacts and the establishment of adult cell-specific gene control. Cdx2 is consistently expressed across stem cells, TA, and villus cells, and its clustered binding is consistently maintained during stem cell differentiation. Thus, Cdx2 may allow early formation of gene regulatory loops that facilitate rapid activation of homeostatic genes during stem cell differentiation in this fast-dividing tissue system. The use of new technologies such as 4-C and HiChIP^{56,57}, along with genetic manipulation of Cdx2 in mouse intestinal cells, may allow testing of such contributions of Cdx2 in future.

Along with DNA sequence, CpG methylation and chromatin structure can affect TF binding⁵⁸. Ctf has a relatively large motif with multiple CpGs, which makes it sensitive to DNA methylation³⁶. A small proportion of all known Ctf binding sites are highly conserved across species and show high occupancy of Ctf across tissues, while the tissue-specific Ctf binding sites show low occupancy⁵⁹. Although DNA methylation and chromatin remodelers such as ISWI are involved in Ctf recruitment^{60,61}, the role of cobinding TFs, which may allow Ctf to overcome the DNA methylation barrier and facilitate its tissue-specific binding, remains understudied. Interestingly, the Ctf motif is not highly detectable in the developmental clusters Dev 1–3 (Fig. 2a). In this light, cobinding of Ctf with Cdx2 in the developing intestinal epithelium (Fig. 2e) suggests a developmental stage-specific mechanism where the lineage-defining TF Cdx2 and its chromatin sensitivity allows Ctf occupancy through corecruitment. This Cdx2-based recruitment of Ctf is further highlighted in adult cells when gain of DNA methylation at developmental Cdx2 binding loci results in ectopic recruitment of Cdx2 and cobinding of Ctf. On the other hand, the loci occupied by Cdx2 specifically in adult cells (Adult 1–3) show strong presence of Hnf4a motif and co-binding of Hnf4a, and reduction of Cdx2 at these loci decreases Hnf4a occupancy. These results suggest that Cdx2, through its motif variety and distribution as well as chromatin sensitivities, may control access and maintenance of Ctf and Hnf4a at distinct target sites.

Although in vitro studies suggest that many TFs have DNA methylation sensitivity, how DNA methylation changes during tissue development direct TFs to their cell type and developmental stage-specific target sites is not well understood. In the in vivo context, both DNA methylation and TF binding may be influenced by nucleosome distribution, histone modifications, motif variability and frequency, and other cobinding TFs. Moreover, these factors may have a distinct impact on TF recruitment versus their sustained binding. Many TFs that prefer DNA methylation-mediated recruitment seem to mediate demethylation at their target loci upon binding⁴⁷. This is also true for Cdx2, as it binds its developmental targets in the intestinal epithelium (E12.5 and E16) when the CpG in Cdx2 motif is uniquely methylated, which then gets demethylated in adult cells. This suggests that Cdx2 may recruit demethylases at these loci through direct or indirect interactions. Indeed, Ctf, which shows cobinding with Cdx2 at many such loci (Fig. 3e) can inhibit DNA methylation by locally suppressing the activity of the ubiquitously expressed DNA (cytosine-5)-methyltransferase 1 (DNMT1) through cobinding with poly(ADP-ribose) polymerase 1 (PARP1), an enzyme that adds ADP-ribose groups to DNMT1 to inactivate it^{47,62,63}.

In multiple studies, we have previously shown that embryonic and fetal enhancers remain hypomethylated in adult cells and that they are ectopically activated upon loss of PRC2 action and the repressive H3K27me3 modification^{15,43}. In the same mouse model system, we now

find that induced loss of PRC2 activity in adult cells causes remethylation of CpG within the Cdx2 motif and induces Cdx2 recruitment to developmental enhancers. In this light, our results reveal a dual-purpose molecular mechanism where CpG methylation facilitates Cdx2 recruitment at developmental loci, while subsequent demethylation at these sites precludes Cdx2 binding in adult cells. Interestingly, adult cells show low levels of H3K36me₂, a histone modification known to recruit the de novo DNA methylase DNMT3A, at the developmental loci despite the hypomethylated state of the enhancers⁶⁴ (Figs. S2g, S5c), which may aid DNA methylation gains in cells with *Eed* loss.

Overall, our data unveil that through the capability to bind motifs with and without CpGs distributed unequally at developmental and adult target sites, respectively, Cdx2 navigates distinct DNA methylation profiles and recruits other TFs to establish lineage-specific enhancer patterns.

Methods

Experimental model and subject details

Mouse models. Mice were handled and treated according to IACUC protocol approved by the Department of Animal Resources at the University of Southern California. Mice were housed in a facility maintained at 23 ± 1 °C, 50 ± 10% humidity, and 12 h light/dark cycles. Animals that showed, hunched posture, ruffled fur, reduced movement, or acute loss of 20% of body weight were removed from the study and humanely euthanized. Euthanasia was conducted using CO₂ inhalation for 5 min with a flow rate of average 20% of chamber volume of CO₂ delivery per minute in accordance with the American Veterinary Medical Association's Guidelines for the Euthanasia of Animals (2013). Cervical dislocation was used as a secondary method of euthanasia. Embryos delivery by cesarean section were euthanized using decapitation with a scalpel. *ROSA26^{LSL-tdTomato}* mice were purchased from The Jackson Laboratories; *Eed^{fl/fl}*⁶⁵, *Lgr5^{EGFP-IRES-DTR}*⁴⁴, and *Villin^{CreER-T2}*⁴⁶ mice were generous gifts from S. Orkin (Boston Children's Hospital), F. de Sauvage (Genetech) and S. Robine (Institut Pasteur, France), respectively. Genetically altered alleles and were confirmed by genotyping mice using PCR at weaning and during experiments. For induced deletion of *Eed*, adult male and female mice with age 8 weeks or older were given 2 mg tamoxifen using intraperitoneal (i.p.) injections on 5 consecutive days, and tissues were harvested at times as indicated for individual experiments.

Isolation of intestinal epithelial cells

Adult villus cells were procured using proximal 1/3rd small intestine of adult male and female mice with age 8 weeks or older. Immediately after harvesting, the tissue was cut along the length to expose epithelium and washed with cold phosphate-buffered saline (PBS). Tissue was rotated in 5 mM EDTA in PBS (pH 8) at 4 °C for 30 min, with vigorous manual shaking every 10 min. The recovered epithelium was filtered using 70 µm filters and the villi retained on the filter were collected in ice-cold PBS. Villi were collected by centrifugation at 300 × g for 5 min and single-cell suspension was generated by rotating in 20 ml of 4% TrypLE solution (Invitrogen) in DMEM at 37 °C for 30 min; 15 ml DMEM was added to neutralize the TrypLE solution and cells were collected using centrifugation. For the isolation of epithelial cells at prenatal time points E12.5 and E16.5, embryos were harvested from pregnant mice 12 and 16 days after identification of copulation plug, respectively (*n* = 4). The small intestine (between stomach and cecum) was chopped into small pieces with razor and digested in 500 µl of 0.25% trypsin for 20 min at 37 °C to generate single-cell suspension. Reaction was stopped by addition of 500 µl of DMEM with 2% FBS. Cells were spun down at 500 × g for 5 min at 4 °C and resuspended in PBS with 2% BSA. EpCam⁺ epithelial cells were isolated by EasySep™ Release Mouse Biotin Positive Selection Kit according to manufacturer protocol and EpCam ab (Invitrogen 13-5791-82) for both adult and developmental tissues.

Isolation of intestinal crypt TA and stem cells

The proximal 1/3 of mouse intestine (duodenum) from *Lgr5^{EGFP-IRES-DTR}* mouse⁴⁴ was used to collect cells for CUT&RUN studies. Intestines harvested immediately after euthanasia were washed with cold phosphate-buffered saline (PBS), followed by rotation for 40 min in 5 mM EDTA in PBS (pH 8) at 4 °C, with manual shaking every 10 min. Crypt epithelium was recovered by filtering the resulting suspension over 70-µm filters (B-D Falcon). Crypts in the 70-µm filtrate were dissociated into single cells by treatment with 4% TrypLE solution (Invitrogen) at 37 °C for 30 min and EGFP^{hi} ISCs and EGFP negative crypt cells (TA cell population) were isolated from the viable (PI) and EpCam⁺ cell fraction by flow cytometry on a BD SORP FACSsymphony S6 instrument.

Native-ChIP-seq

Native-ChIP-seq was performed as previously described²⁶. In brief, cells were lysed in 0.1% Triton X-100 and 0.1% Sodium Deoxycholate with protease inhibitor cocktail. Chromatin in the cell lysate was digested using Micrococcal nuclease (MNase, New England BioLabs, M0247S) at room temperature for 5 min, and 5 µl 0.25 mM EDTA was added to stop the reaction. The digested chromatin was incubated with anti-IgA magnetic beads (Dynabeads, Thermo Fisher, 10001D) for 2 h for pre-clearing and then incubated overnight with antibody-bead complexes with 0.5 µg of antibodies against H3K27me₃, H3K27me₂, H3K27me₁, H3K36me₂, H3K36me₃, and H2AK119Ub (Diagenode, C15410195, C15410046, C15410045, C15310127, C15410192 8240S, respectively) in immunoprecipitation (IP) buffer (20 mM Tris-HCl pH 7.5, 2 mM EDTA, 150 mM NaCl, 0.1% Triton X-100, 0.1 % Sodium Deoxycholate) at 4 °C. IPs were washed 2 times by Low Salt (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.1% SDS) and High Salt (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 500 mM NaCl, 1% Triton X-100, 0.1% SDS) wash buffers. IPs were eluted in elution buffer (1% SDS, 100 mM Sodium Bicarbonate) for 1.5 h at 65 °C. Histones were digested by Protease (Qiagen 19155) for 30 min at 50 °C and DNA fragments were purified using Sera Mag magnetic beads in 30% PEG. Illumina sequencing libraries were generated as previously described⁶⁶ by end repair, 3' A-addition, and Illumina sequencing adaptor ligation (New England BioLabs, E6000B-10). Libraries were then indexed and PCR amplified (10 cycles) and sequenced on Illumina HiSeq 2500 sequencing platform following the manufacturer's protocols (Illumina).

Whole genome bisulfite sequencing (WGBS)

MasterPure DNA purification kit (Epicenter MCD85201) was used to purify genomic DNA from the cells followed by bisulfite conversion using 50 ng DNA along with the EZ DNA Methylation-Gold kit (Zymo Research D5005). 10 ng of bisulfite-converted DNA was used to generate the whole genome bisulfite sequencing (WGBS) libraries with the EpiGenome Methyl-Seq kit (Epicenter EGMK81312). We used AMPure beads (Beckman Coulter) to purify the library material and confirmed that the library DNA was in the size range of 200–800 bp using high-sensitivity DNA Chip detection (Bioanalyzer 2100, Agilent Genomics), and sequenced the libraries on a NextSeq 500 instrument (Illumina) to generate 150-bp paired-end reads; up to 50% PhiX phage DNA (Illumina) was mixed with the libraries to assist the sequencing.

GSK-3484862 treatment

HCT116 cells were seeded at 20 K per well in a 24-well plate in DMEM containing 10% FBS and 1% Penicillin-Streptomycin. Cells were treated with 1.25% DMSO or 20 µM GSK-3484862 (Med Chem Express 2170136-65-7) for 6 days with daily media change. Cells were collected at the end of this treatment or after 3 days of culture without GSK-3484862 or DMSO treatment (recovery condition). HCT116 was a gift from Dr. Heinz-Josef Lenz (University of Southern California) and was originally procured from ATCC (CCL-247).

Bisulfite amplicon sequencing

DNA was purified with DNeasy DNA Isolation kit (Qiagen 69504). 300 ng of DNA was bisulfite converted with EZ DNA Methylation-Gold kit (Zymo Research D5005). 5 ng of DNA was used for amplicon generation using TaKaRa EpiTaq™ HS (Takara Bio R110A). Primer Sequences are provided in Supplementary Table 1. Amplicons were pooled together and sequenced using Oxford Nanopore Technologies long read sequencing at PLASMIDSAURUS (v14 library prep chemistry, R10.4.1 flow cells).

Cell cycle analysis

Adult villus and crypt epithelial cells were procured using proximal 1/3rd small intestine of *Lgr5^{EGFP-ires-DTR}* mice⁴⁴. Intestinal stem cells were FACS sorted using endogenous GFP signal. Cells were spun down at 500 × *g* for 5 min at 4 °C and fixed in 70% EtOH for 30 min at 4 °C. Cells were washed twice with PBS and resuspended in DMEM and 2% FBS. DAPI was added to the solution and cells were processed on BD SORP FACSsymphony S6 instrument. Proportion of cells in each cell cycle stage was determined by Dean-Jett Fox model using FlowJo (version 10.10) with default parameters.

CUT&RUN

CUT&RUN was performed using EpiCypher CUTANA kit (14-1048) according to manufacturer's protocol with the following changes. Freshly isolated epithelial cells were resuspended in the permeabilization buffer (CUTANA Wash Buffer, 0.01% Digitonin, and 1X protease inhibitor) and incubated on ice for 5 min and captured on Concanavalin A beads at room temperature for 10 min. Captured cells were resuspended in 180 µl of antibody buffer (permeabilization buffer with 2 mM EDTA) with 0.5 µg of antibody against Cdx2 (Cell Signalling, D11D10), Ctf (Diagenode, C15410210), or Hnf4a (Santa Cruz Biotechnology, sc-374229) and incubated overnight. Antibody-bound chromatin was digested at 4 °C for 30 min on a nutator. Fragment release postdigestion was done at 4 °C for 30 min and purified using SeraMag beads with 30% PEG-8000. Library construction was done as described in the Native-ChIP-seq section.

ATAC-seq

Fifty thousand *Eed*^{-/-} adult villus cells were lysed in 50 µl cold ATAC lysis buffer (10 mM Tris-Cl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% (v/v) Igepal CA-630) followed by centrifugation at 500 × *g* at 4 °C to isolate nuclei. Nuclei were resuspended in 50 µl transposition reaction mix with Nextera Tn5 Transposase (Illumina, FC-121-1030) and incubated for 30 min at 37 °C. Transposed DNA was purified using columns (Qiagen, 28004) and amplified using high-fidelity 2X PCR Master Mix (New England Biolabs) using primers with standard ATAC-seq barcodes. AMPure beads Beckman Coulter, A63880 were used to remove primer dimers, and libraries were sequenced on a NextSeq 500 instrument (Illumina) to generate 75 bp single-end reads.

Cross-linked ChIP-seq

Approximately 1 × 10⁶ wild-type and *Eed*^{-/-} villus cells were fixed in 1% formaldehyde for 25 min at room temperature immediately after isolation. The reaction was stopped by addition of Glycine to a final concentration of 0.13 M. Cells were lysed using a buffer containing 30 mM Tris-HCl (pH 8), 1% SDS, 10 mM EDTA, and protease inhibitors (Roche), and the chromatin was sonicated for 50 min using a Covaris sonicator (5 min on/off cycles at 4 °C). Debris were removed using centrifugation and the chromatin was incubated overnight at 4 °C with antibody against H3K9me3 (Abcam, Ab8898, 1 µg). Antibody-bound chromatin was captured using magnetic beads (Thermo Fisher, 10001D) and washed using low-salt (20 mM Tris-HCl pH 8.1, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% TritonX-100), high-salt (20 mM Tris-HCl pH 8.1, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% TritonX-100), and lithium chloride (10 mM Tris pH 8.1, 0.25 M LiCl,

1 mM EDTA, 1% NP-40, 1% deoxycholate) buffers. Chromatin was treated with 1% SDS and 0.1 M NaHCO₃ for 6 h at 65 °C to reverse the cross-links and DNA was purified using columns (Qiagen) followed by ChIP-seq library preparation using ThruPLEX kit (Rubicon, R400427) and sequenced using NextSeq 500 instrument (Illumina) to produce 75 bp single-end reads.

Quantification and statistical analysis

Computational analyses

RNAseq. Raw RNA-seq reads were aligned to the mouse genome (GRCm38, GENCODEvM25) using STAR aligner v2.7.8a⁶⁷ in 2-passed mode followed by assessment to determine per-base sequence quality, per-read GC content (~50%), comparable read alignments to ± strands, exon vs intron read distributions, and 3' bias. Transcript levels were expressed as read counts using HTSeq v0.13.5⁶⁸ and normalized across libraries using DESeq2¹⁸, followed by conversion into reads per kb of transcript length per 1 M mapped reads (RPKM). We determined differential expression between samples using DESeq2, with false-discovery rate (FDR) as indicated in the text. All alignments and initial processing steps were run using the Snakemake workflow management system (v6.6.1). Cartoon Illustrations were created with BioRender.com.

ChIP-seq and ATAC-seq. ChIP-seq and ATAC-seq data were trimmed using trim-galore (v0.6.6) to remove adapters and low-quality reads. Trimmed sequences were aligned using bwa-mem (v0.7.17) with default parameters. Non-uniquely mapped reads, duplicate reads, and reads with a MAPQ score of less than 5 were removed using Samtools (v1.17). Peak calling was done using MACS (v2.2.7.1) with a *q*-value cutoff of 0.01 in sharp mode for H3K27ac and ATAC, and broad mode for H3K27me3,2 and 1, H3K36me3 and 2, H2AK119Ub and H3K9me3. BigWig files were generated using deepTools (v3.5.1) with a bin size of 10, smooth length of 30, and quantile normalized using Haystack. Bedtools (v2.30.0) was used to removed blacklisted regions defined by ENCODE blacklist.

Analysis of DNA methylation data. Whole genome bisulfite data were aligned to mm10 in non-strand specific mode using Bismark v0.23.0⁶⁹. Coverage from both strands for each CpG was combined and fractional methylation was calculated for each CpG using custom AWK command. Differential single CpGs were identified using Fisher exact test (*p* < 0.001). Differential CpG with at least 20% difference in fractional methylations and within 300 bp of one another were merged using custom AWK command. Hypermethylated regions with less than 40% methylation were discarded.

Analysis of bisulfite amplicon sequencing. Sequences for each amplicon were demultiplexing using forward and reverse primer sequences (Supplementary Table 1). Fractional methylation for CpG's in each amplicon was calculated by identifying converted (T) and unconverted Cs on both strands using R.

Motif enrichment analysis. Motif enrichment comparisons between developmental and adult Cdx2 sites were conducted using HOMER findMotifGenome with default parameters across the whole Cdx2 binding site. Instances of the motif were identified as previously described⁶⁶. In brief, significant instances of the Cdx2 motif (*p* < 0.005) were identified using MEME-SUIT fimo command⁷⁰. Instances of CpG containing Cdx2 motif were further filtered to include motif with CpG at the 4th position and CGTA sequence at 4–7th positions. Both 10 bp and 12 bp CpG containing Cdx2 motif (Fig. S2a) were considered together for counting of CpG containing motifs⁶⁶. In brief, significant instances of the Cdx2 motif (*p* < 0.005) was identified using MEME-SUIT fimo command⁷⁰. Instances of CpG containing Cdx2 motif were further filtered to include motif with CpG at the 4th position and CGTA

sequence at 4 to 7th positions. Both 10 bp and 12 bp CpG containing Cdx2 motif (Fig. S2a) were considered together for counting of CpG containing motifs.

Identification of unmethylated (UMR) and low-methylated (LMR) regions. The MethylSeekR package⁴⁸ was used along with the mouse bisulfite-converted genome (BSgenome.Mmusculus.UCSC.mm10) from Bioconductor (www.bioconductor.org) as the reference to determine UMRs and LMRs. We only used CpGs with coverage five or more and eliminated C nucleotides overlapping known SNPs between the reference strains (C57BL/6J and 129/S5) to avoid any effects of polymorphism. Methylation levels were smoothed over 5 consecutive CpG dinucleotides hypomethylated regions were identified as UMRs (unmethylated and minimum 5 CpGs and less than 10% fractional methylation) or LMRs (minimum 5 CpGs with fractional methylation between 10% and 60%).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The sequencing data reported in this study is accessible through Gene Expression Omnibus (GEO). The accession number for the data generated in this study is [GSE253736](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE253736). Additional published data used: ERP001953 [<https://www.ebi.ac.uk/ena/browser/view/PRJEB3376>] (WGBS DNA methylation data for E6.5 endoderm). [GSE115541](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115541) (mRNA-seq and ATAC-seq data for intestinal epithelial cells during endoderm development). [GSE11024](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11024) (WGBS DNA methylation and H3K27ac CHIP-seq data for developing and adult intestinal epithelial cells). [GSE60106](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60106) (WGBS DNA methylation data for HCT-116 cells). Source data are provided with this paper.

References

- D'haeseleer, P. What are DNA sequence motifs? *Nat. Biotechnol.* **24**, 423–425 (2006).
- Hansen, L., Mariño-Ramírez, L. & Landsman, D. Differences in local genomic context of bound and unbound motifs. *Gene* **506**, 125–134, (2012).
- Adams, C. C. & Workman, J. L. Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol. Cell Biol.* **15**, 1405–1421, (1995).
- Domcke, S. et al. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528**, 575–579 (2015).
- Slattery, M. et al. Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* **39**, 381–399 (2014).
- Gehart, H. & Clevers, H. Tales from the crypt: new insights into intestinal stem cells. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 19–34 (2019).
- Gao, N., White, P. & Kaestner, K. H. Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. *Dev. Cell* **16**, 588–599, (2009).
- Tamai, Y. et al. Colonic hamartoma development by anomalous duplication in Cdx2 knockout mice. *Cancer Res.* **59**, 2965–2970 (1999).
- Grainger, S., Savory, J. G. & Lohnes, D. Cdx2 regulates patterning of the intestinal epithelium. *Dev. Biol.* **339**, 155–165, (2010).
- Kumar, N. et al. The lineage-specific transcription factor CDX2 navigates dynamic chromatin to control distinct stages of intestine development. *Development* **146**, dev172189 (2019).
- San Roman, A. K., Tovaglieri, A., Breault, D. T. & Shivdasani, R. A. Distinct processes and transcriptional targets underlie CDX2 requirements in intestinal stem cells and differentiated villus cells. *Stem Cell Rep.* **5**, 673–681 (2015).
- Bird, A., Taggart, M., Frommer, M., Miller, O. J. & Macleod, D. A fraction of the mouse genome that is derived from islands of non-methylated, CpG-rich DNA. *Cell* **40**, 91–99 (1985).
- Schmitz, R. J., Lewis, Z. A. & Goll, M. G. DNA methylation: shared and divergent features across eukaryotes. *Trends Genet.* **35**, 818–827 (2019).
- Yin, Y. et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017).
- Jadhav, U. et al. Extensive recovery of embryonic enhancer and gene memory stored in hypomethylated enhancer DNA. *Mol. Cell* **74**, 542–554 e545 (2019).
- Mingay, M. et al. Vitamin C-induced epigenomic remodelling in IDH1 mutant acute myeloid leukaemia. *Leukemia* **32**, 11–20 (2018).
- Meers, M. P., Janssens, D. H. & Henikoff, S. Pioneer factor-nucleosome binding events during differentiation are motif encoded. *Mol. Cell* **75**, 562–575.e565 (2019).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- Kazakevych, J., Sayols, S., Messner, B., Krienke, C. & Soshnikova, N. Dynamic changes in chromatin states during specification and differentiation of adult intestinal stem cells. *Nucleic Acids Res.* **45**, 5770–5784 (2017).
- Shyer, A. E., Huycke, T. R., Lee, C., Mahadevan, L. & Tabin, C. J. Bending gradients: how the intestinal stem cell gets its home. *Cell* **161**, 569–580 (2015).
- Francis, R. et al. Gastrointestinal transcription factors drive lineage-specific developmental programs in organ specification and cancer. *Sci. Adv.* **5**, eaax8898 (2019).
- Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
- de la Rica, L. et al. PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. *Genome Biol.* **14**, R99 (2013).
- Seisenberger, S. et al. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48**, 849–862 (2012).
- Lorzadeh, A. et al. Nucleosome density ChIP-Seq identifies distinct chromatin modification signatures associated with MNase accessibility. *Cell Rep.* **17**, 2112–2124 (2016).
- Lorzadeh, A., Romero-Wolf, M., Goel, A. & Jadhav, U. Epigenetic regulation of intestinal stem cells and disease: a balancing act of DNA and histone methylation. *Gastroenterology* **160**, 2267–2282 (2021).
- Meehan, R. R. & Pennings, S. Shoring up DNA methylation and H3K27me3 domain demarcation at developmental genes. *EMBO J.* **36**, 3407–3408 (2017).
- Jadhav, U. et al. Acquired tissue-specific promoter bivalency is a basis for PRC2 necessity in adult cells. *Cell* **165**, 1389–1400 (2016).
- Xie, W. et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
- Reddington, J. P. et al. Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biol.* **14**, R25 (2013).

32. Hagarman, J. A., Motley, M. P., Kristjansdottir, K. & Soloway, P. D. Coordinate regulation of DNA methylation and H3K27me3 in mouse embryonic stem cells. *PLoS ONE* **8**, e53880 (2013).
33. Bannister, A. J. et al. Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J. Biol. Chem.* **280**, 17732–17736 (2005).
34. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
35. Liu, Y. et al. CTCF coordinates cell fate specification via orchestrating regulatory hubs with pioneer transcription factors. *Cell Rep.* **42**, 113259 (2023).
36. Maurano, M. T. et al. Role of DNA methylation in modulating transcription factor occupancy. *Cell Rep.* **12**, 1184–1195 (2015).
37. Nanavaty, V. et al. DNA methylation regulates alternative polyadenylation via CTCF and the cohesin complex. *Mol. Cell* **78**, 752–764.e756 (2020).
38. Islam, R. et al. RUNX1 colludes with NOTCH1 to reprogram chromatin in T cell acute lymphoblastic leukemia. *iScience* **26**, 106795 (2023).
39. Verzi, M. P., Shin, H., San Roman, A. K., Liu, X. S. & Shivdasani, R. A. Intestinal master transcription factor CDX2 controls chromatin access for partner transcription factor binding. *Mol. Cell Biol.* **33**, 281–292, (2013).
40. Oomen, M. E., Hansen, A. S., Liu, Y., Darzacq, X. & Dekker, J. CTCF sites display cell cycle-dependent dynamics in factor binding and nucleosome positioning. *Genome Res.* **29**, 236–249 (2019).
41. Zhang, H. et al. Chromatin structure dynamics during the mitosis-to-G1 phase transition. *Nature* **576**, 158–162 (2019).
42. Parker, A. et al. Cell proliferation within small intestinal crypts is the principal driving force for cell migration on villi. *FASEB J.* **31**, 636–649 (2017).
43. Jadhav, U. et al. Replicational dilution of H3K27me3 in mammalian cells and the role of poised promoters. *Mol. Cell* **78**, 141–151.e145 (2020).
44. Tian, H. et al. A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature* **478**, 255–259 (2011).
45. Ezhkova, E. et al. Ezh2 orchestrates gene expression for the step-wise differentiation of tissue-specific stem cells. *Cell* **136**, 1122–1135 (2009).
46. el Marjou, F. et al. Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* **39**, 186–193 (2004).
47. Stadler, M. B. et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495 (2011).
48. Burger, L., Gaidatzis, D., Schübeler, D. & Stadler, M. B. Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res.* **41**, e155 (2013).
49. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216, (2012).
50. Azevedo Portilho, N. et al. The DNMT1 inhibitor GSK-3484862 mediates global demethylation in murine embryonic stem cells. *Epigenetics Chromatin* **14**, 56 (2021).
51. Blattler, A. et al. Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biol.* **15**, 469 (2014).
52. Butz, S. et al. DNA sequence and chromatin modifiers cooperate to confer epigenetic bistability at imprinting control regions. *Nat. Genet.* **54**, 1702–1710 (2022).
53. Cai, W. et al. Enhancer dependence of cell-type-specific gene expression increases with developmental age. *Proc. Natl. Acad. Sci. USA* **117**, 21450–21458 (2020).
54. Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626, (2012).
55. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
56. Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).
57. Huang, Z., Wang, C., Treuter, E. & Fan, R. An optimized 4C-seq protocol based on cistrome and epigenome data in the mouse RAW264.7 macrophage cell line. *STAR Protoc.* **3**, 101338 (2022).
58. Crispatzu, G. et al. The chromatin, topological and regulatory properties of pluripotency-associated poised enhancers are conserved in vivo. *Nat. Commun.* **12**, 4344 (2021).
59. Essien, K. et al. CTCF binding site classes exhibit distinct evolutionary, genomic, epigenomic and transcriptomic features. *Genome Biol.* **10**, R131 (2009).
60. Wang, H. et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680–1688 (2012).
61. Wiechens, N. et al. The chromatin remodelling enzymes SNF2H and SNF2L position nucleosomes adjacent to CTCF and other transcription factors. *PLoS Genet.* **12**, e1005940 (2016).
62. Guastafierro, T. et al. CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. *J. Biol. Chem.* **283**, 21873–21880 (2008).
63. Zampieri, M. et al. ADP-ribose polymers localized on Ctcf-Parp1-Dnmt1 complex prevent methylation of Ctcf target sites. *Biochem. J.* **441**, 645–652 (2012).
64. Weinberg, D. N. et al. The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* **573**, 281–286 (2019).
65. Xie, H. et al. Polycomb repressive complex 2 regulates normal hematopoietic stem cell function in a developmental-stage-specific manner. *Cell Stem Cell* **14**, 68–80 (2014).
66. Lorzadeh, A. et al. Polycomb contraction differentially regulates terminal human hematopoietic differentiation programs. *BMC Biol.* **20**, 104 (2022).
67. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
68. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169, (2015).
69. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
70. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018, (2011).

Acknowledgements

This work was supported by NIH grants KO1DK113067, R03DK134799, and Donald E. and Delia B. Baxter Foundation Fellowship Award to U.J. A.L. was supported by Canadian Institute of Health Research (CIHR) Post-Doctoral Fellowship and California Institute of regenerative Medicine (CIRM) Post-Doctoral Fellowship.

Author contributions

A.L. and U.J. conceptualized the study design. A.L. and G.Y. performed the analysis. A.L. and S.S. performed experimental work. A.L. and U.J. wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-56187-0>.

Correspondence and requests for materials should be addressed to Unmesh Jadhav.

Peer review information *Nature Communications* thanks Timothy Downing, and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025