In Press, *Developmental Cell* 2010

Differentiation-specific histone modifications reveal dynamic chromatin interactions and alternative partners for the intestinal transcription factor CDX2

Michael P. Verzi[1,2,*], Hyunjin Shin[3,*], H. Hansen He[1,3], Rita Sulahian[1,2], Clifford A. Meyer[3], Robert K. Montgomery[1,4], James C. Fleet[5], Myles Brown[1,2], X. Shirley Liu[3], Ramesh A. Shivdasani[1,2,‡]

Department of [1]Medical Oncology and [3]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA; [2]Departments of Medicine, Brigham & Women's Hospital and Harvard Medical School, Boston, MA;
[4]Gastroenterology and Nutrition Division, Children's Hospital, Boston, MA;
[5]Department of Foods and Nutrition, Purdue University, West Lafayette, IN

*Equal contribution

‡**Corresponding author:**
Ramesh A. Shivdasani, M.D., Ph.D.
Dana-Farber Cancer Institute
44 Biney Street, Boston, MA 02115
Tel. 617-632-5746 Fax 617-582-7198
ramesh_shivdasani@dfci.harvard.edu

Running title: Dynamic Binding and Functions of CDX2

SUMMARY

Cell differentiation requires remodeling of tissue-specific gene loci and activities of key transcriptional regulators, which are recognized for their dominant control over cellular programs. Using epigenomic methods, we characterized enhancer elements specifically modified in differentiating intestinal epithelial cells and found enrichment of transcription factor-binding motifs corresponding to CDX2, a critical regulator of the intestine. Directed investigation revealed surprising lability in CDX2 occupancy of the genome, with redistribution from hundreds of sites occupied only in proliferating cells to thousands of new sites in differentiated cells. Knockout mice confirmed distinct Cdx2 requirements in dividing and mature adult intestinal cells, including responsibility for the active enhancer configuration associated with maturity. Dynamic CDX2 occupancy corresponds with condition-specific gene expression and, importantly, to differential co-occupancy with other tissue-restricted transcription factors: HNF4A in mature cells and GATA6 in progenitors. These results reveal dynamic, context-specific functions and mechanisms of a prominent transcriptional regulator within a cell lineage.

HIGHLIGHTS

- A histone modification screen flagged CDX2 as a regulator of gut differentiation
- CDX2 interactions with the genome change materially as cells differentiate
- Adult knockout mice reveal distinct Cdx2 functions in progenitor and mature cells
- Context-specific CDX2 binding occurs with specific partner proteins GATA6 or HNF4A

INTRODUCTION

Thousands of transcripts are coordinately regulated in differentiating cells, in part as a result of stable, heritable, cell type-specific changes in chromatin structure and in part through the actions of selected transcription factors (TFs) at many gene loci (Struhl, 1999). TFs that are highly restricted in their expression, regulate large numbers of genes within a cell lineage, and may hence confer a cell's distinctive properties, are often considered "master regulators." Examples of such TFs include PHA4 in the worm pharynx (Gaudet and Mango, 2002) and mammalian myogenic basic helix-loop-helix proteins (Molkentin and Olson, 1996). Although these TFs may function in both progenitor and terminally differentiated cells of a lineage, their interactions with alternative chromatin states are insufficiently characterized. Nor is it clear if they occupy *cis*-regulatory DNA elements stably throughout differentiation and whether distinct TF combinations control genes in different cell states within a lineage. We addressed these questions on a genome scale in the context of the self-renewing intestinal epithelium.

The adult intestine is a site of continual cell differentiation (Potten, 1998). Stem and dividing progenitor cells in the small bowel mucosa are confined to the crypts of Lieberkühn, whereas the differentiated post-mitotic cells that serve essential absorptive and secretory functions reside along villous projections. Gene expression differs substantially in crypts and villi (Tremblay et al., 2006) and, as with other self-renewing tissues, constitutive proliferation and arrested differentiation underlie intestinal tumorigenesis (van de Wetering et al., 2002). Gut epithelial cells hence serve as an excellent model to investigate gene regulation during differentiation.

Recent delineation of chromatin modifications at activated gene loci provides new tools to investigate the epigenomic basis of cell differentiation (Barski et al., 2007; Bernstein et al., 2006; Pokholok et al., 2005). Mono- and di-methylation of Lys4 (K4Me2) and acetylation of Lys27 (K27Ac) residues on Histone 3 (H3) are particularly associated with distant transcriptional enhancers and gene activation in cultured mammalian cells (Bernstein et al., 2005; He et al., 2010; Heintzman et al., 2009). Monitoring these chromatin modifications as cells transition from one state to another could identify the *cis*-elements responsible for differentiation in individual tissues, and characterization of such elements should help elucidate underlying transcriptional mechanisms. To identify *cis*-regulatory elements that are differentially active in proliferating and

terminally mature intestinal epithelial cells, we mapped genome-wide histone H3 modifications. Our subsequent detailed analysis uncovered a prominent role and mechanisms for the homeodomain protein CDX2.

CDX2 expression is restricted at different developmental stages and its ability to specify cell fates or axial position in each context makes it a strong candidate "master regulator." Expression of Cdx2 in mouse embryonic stem cells induces trophoblasts (Niwa et al., 2005) and expression in stomach or esophageal cells confers intestinal properties (Liu et al., 2007; Silberg et al., 2002). Conversely, Cdx2 loss in mouse embryos disrupts trophoblast, axial skeleton, and intestine development (Chawengsaksophak et al., 1997; Gao et al., 2009; Grainger et al., 2010). In adult mammals, Cdx2 expression is confined to the intestinal epithelium, where it is expressed in both progenitor and differentiated cells (James et al., 1994; Silberg et al., 2000). To examine if CDX2 behaves as a key regulator and to investigate its potentially diverse functions in dividing and mature cells, we assessed CDX2 binding properties during differentiation. We find that CDX2 interactions with the genome and with other sequence-specific TFs are surprisingly fluid; it interacts dynamically with active chromatin and co-occupies DNA differentially with other intestinal TFs, GATA6 or HNF4A, to regulate distinct genes in dividing and differentiated intestinal cells, respectively. A new conditional knockout mouse line revealed distinct Cdx2 functions in progenitor and differentiated cells *in vivo* and its requirement for differentiation-specific chromatin modifications. These results illustrate the dynamic functions and mechanisms of a critical TF in distinct cellular contexts within a continually differentiating tissue.

RESULTS

Epigenomic analysis of intestinal epithelial cells implicates CDX2 in cell maturation

Caco-2 human intestinal cells are widely used to investigate epithelial functions, intestinal gene expression, and transcriptional mechanisms of differentiation (Fleet et al., 2003; Halbleib et al., 2007; Soutoglou and Talianidis, 2002). These cells proliferate rapidly in sparse cultures but stop dividing at confluence (Fig. 1A) and develop morphologic features of mature enterocytes; the transcriptional changes that accompany cell differentiation mirror differences in gene expression between intestinal crypts and villi (Saaf et al., 2007; Tremblay et al., 2006). Accordingly, they serve as an ideal model to study pivotal cellular transitions. To identify *cis*-regulatory sequences associated with Caco-2 cell differentiation, we evaluated a distinct pattern of histone modifications attributed to active enhancer elements (He et al., 2010; Heintzman et al., 2009): abundant H3K4Me2 and H3K27Ac marks, with depletion of labile nucleosomes within an altered chromatin configuration (Fig. 1B). We digested chromatin from proliferating and post-confluent Caco-2 cells with micrococcal nuclease (MNase), precipitated the mononucleosome fraction with H3K4Me2- or H3K27Ac antibodies (Ab), and characterized the chromatin immunoprecipitates (ChIP) using Illumina high-throughput sequencing (ChIP-seq). Subsequent analysis of the two H3 modifications gave similar results (Pearson correlation coefficient 0.4 (*P*-value $<10e^{-22}$), Supplemental Fig. S1); for simplicity, here we show the analysis of H3K4Me2 ChIP.

We analyzed 14.5 million and 13 million mappable sequence tags for H3K4Me2 ChIP from proliferating and mature cells, respectively, using published methods (He et al., 2010), and identified 219,266 and 208,762 positioned nucleosomes under the two conditions (Supplemental Fig. S2). H3K4Me2 marks were restricted either to transcriptional start sites (TSSs) or to dispersed short intergenic and intronic regions, with little background signal across the genome (examples in Fig. 1C). Approximately 60% of nucleosomes bearing the H3K4Me2 mark in each condition were present more than 5 kb from a TSS and hence represent putative distant *cis*-regulatory elements. We focused on these distal elements because previous work indicates that enhanced H3K4Me2 signals at flanking nucleosomes and a corresponding diminution of intervening signal typifies sites where nucleosomes are depleted or destabilized to allow TF binding (He et al., 2010; Heintzman et al., 2007; Jin et al., 2009). A computational method to

infer increased H3K4Me2 signals at flanking nucleosomes relative to the intervening region uncovered thousands of individual sites that acquire or lose this signature during Caco-2 cell differentiation (sample tracings in Fig. 1D). We used the magnitude of difference in signal between replicating and mature cells to assign Nucleosome Stability-Destability (NSD) scores (He et al., 2010) and to rank genomic regions. The resulting patterns, displayed in composite plots of 1,000 sites with the highest NSD scores in proliferating (Fig. 1E) and differentiated (Fig. 1F) cells, underscore the considerable genome-wide modulation of chromatin during cell differentiation.

To identify proteins that might associate with labile regions, we searched for TF-binding sequence motifs that are over-represented at the centers of the most differentially modified chromatin domains (Supplemental Table S1). In regions active exclusively in proliferating cells, the most enriched motif matched the GATA protein family (Fig. 1E). Motifs enriched in chromatin that is active mainly in differentiated cells corresponded to the transcription factors HNF1, HNF4A and CDX2 (Fig. 1F, Supplemental Table S1), all of which are tissue-restricted proteins previously implicated in controlling genes expressed in mature enterocytes (Beck, 2004; D'Angelo et al., 2010; Mouchel et al., 2004; Stegmann et al., 2006). The selective over-representation of these sequences within differentially active enhancers reinforces the idea that NSD scores signify a regulatory function.

In vitro studies suggest that CDX2 controls intestinal genes like *SI*, *LCT* and *CDH17* through their proximal promoters (Fang et al., 2000; Hinoi et al., 2002; Traber and Silberg, 1996), contributing to the idea of a "master" function. Indeed, Cdx2 is required to specify embryonic intestinal epithelium *in vivo* (Gao et al., 2009; Grainger et al., 2010) but in the adult intestine it is abundantly and equally expressed in crypt progenitors and differentiated villus cells (Silberg et al., 2000) and its function is unknown. The enrichment of CDX2 motifs at differentiation-associated enhancers suggested broad and distinct functions in controlling adult intestinal differentiation through distant *cis*-elements and provided fresh impetus to study its mechanisms.

CDX2 interacts dynamically with the genome of differentiating cells

CDX2 is expressed in both crypt and villus cells (Silberg et al., 2000) and could in principle regulate different loci in these two contexts. To verify CDX2 binding at enhancer regions, as the

H3K4Me2 and H3K27Ac ChIP results suggested, and to determine if CDX2 interacts differently with the genome in progenitor and differentiated cells, we used ChIP-seq to identify CDX2 binding sites. We detected CDX2 occupancy with high confidence ($p \leq 10^{-10}$) at 3,122 regions in sub-confluent, proliferating cells and at 16,198 sites in terminally differentiated cells (Fig. 2A). 679 sites were unique to proliferating cells and 13,755 were unique to differentiated cells (examples in Fig. 2B and Supplemental Fig. S3). CDX2-bound regions in both states were strongly enriched for the consensus CDX2 recognition motif and showed high, centered evolutionary conservation; most sites were far from TSSs (Supplemental Fig. S4), similar to findings with other TFs (Carroll et al., 2006). Thus, although CDX2 binds many regions common to the two states, its occupancy across the genome is surprisingly fluid, with hundreds of distinct progenitor- and thousands of mature cell-specific binding sites. As CDX2 protein levels increase no more than 2- to 3-fold in differentiated cells (data not shown), changes in CDX2 binding are unlikely to reflect only the protein concentration; furthermore, occupancy is selectively diminished at many progenitor sites.

To test the significance of condition-specific CDX2 occupancy, first we defined condition-specific binding rigorously, considering only high-stringency sites from one condition ($p \leq 10^{-10}$) that were unoccupied in the other, even at lower stringency ($p \leq 10^{-3}$, Supplemental Fig. S3). To determine how these high-confidence, condition-specific CDX2 sites relate to H3K4Me2 modifications, we mapped binding sites onto the genome-wide distribution of nucleosomes that are differentially marked in replicating and mature Caco-2 cells. Progenitor-specific CDX2 binding occurred more frequently at enhancers that carry the active chromatin signature in proliferating cells, whereas differentiation-specific CDX2 binding predominated in regions of active chromatin in mature cells (Fig. 2C). CDX2 occupancy at proliferation-specific sites was also strongly associated with transcripts whose expression is enriched selectively in dividing Caco-2 cells ($p<7.3e^{-6}$, Fisher's exact test), whereas CDX2 binding at regions specific to differentiated cells correlated better, and strongly, with transcripts that increase in the latter ($p<1.6e^{-23}$, Fig. 2D). Moreover, loci selectively occupied and expressed in dividing cells were enriched for Gene Ontology (GO) functions related to progenitors, including Embryonic, Organ and Tissue Development, whereas GO terms associated with differentiated enterocytes, such as Lipid Metabolism and Digestion, predominated among genes selectively occupied by CDX2 and expressed in terminally mature cells (Fig. 2E, Supplemental Table S2). Thus, labile CDX2

interactions with distant *cis*-elements in dividing and mature intestinal cells have observable counterparts in differential chromatin structure and gene expression. These results provided strong rationale to investigate the dynamic actions of a transcriptional regulator in the course of cell differentiation.

Cdx2 is required for both intestinal epithelial differentiation and proliferation in vivo

Whereas knockout mice have revealed Cdx2 functions in specifying fetal intestinal epithelium (Gao et al., 2009), its roles in the adult organ remain untested *in vivo*. To circumvent neonatal lethality associated with embryonic Cdx2 deficiency, we generated a new conditional null allele, $Cdx2^{fV}$, with LoxP sites flanking exon 2 (Supplemental Fig. S5A-F), and crossed targeted mice with the intestine-specific, tamoxifen-inducible Cre recombinase-expressing transgenic strain $Villin\text{-}Cre^{ER(T2)}$ (el Marjou et al., 2004). Tamoxifen-treated adult $Cdx2^{fV/fV};Villin\text{-}Cre^{ER(T2)}$ mice showed loss of Cdx2 protein in >99% of small intestine villi and absence of *Cdx2* mRNA (Fig. 3A,B; Supplemental Fig. S5G-H); expression of putative Cdx2 target genes, *Sucrase Isomaltase* and *Intestine Specific Homeobox* (Choi et al., 2006; Traber and Silberg, 1996), was profoundly reduced (Fig. 3B). These mice developed chronic diarrhea, rapidly lost weight (Fig. 3C) and became moribund, requiring humane euthanasia within 21 days. At necropsy they uniformly showed small intestine dilation and a dilated, fluid-filled cecum (Fig. 3D). Tissue architecture was overtly preserved in $Cdx2^{fV/fV};Villin\text{-}Cre^{ER(T2)}$ intestines (Fig. 3A,E). However, alkaline phosphatase activity was notably reduced along a proximal-distal gradient in mutant small bowel compared to tamoxifen-treated sibling controls (Fig. 3F), indicating enterocyte immaturity; enzyme activity was lowest in the ileum, where Cdx2 levels are the highest in wild-type mice (Silberg et al., 2000). The numbers of periodic acid Schiff-stained goblet (increased 2.1%, Fig. 3G), lysozyme$^+$ Paneth (unchanged, Fig. 3H) and ChromograninA$^+$ endocrine cells (decreased by 40%, P=0.0895, Fig. 3I) were altered subtly or not at all.

Although cell state-specific profiles in Caco-2 cells revealed distinct CDX2-occupied sites in proliferating and mature cells (Fig. 2A), the defects in adult $Cdx2^{fV/fV};Villin\text{-}Cre^{ER(T2)}$ intestines seemed confined to villus enterocytes. Crypt progenitors were also overtly normal and measures of cell proliferation - expression of Ki67 or Proliferating Cell Nuclear Antigen (Pcna) - were intact (Fig. 3J). Because Cdx1, a homologous factor normally expressed in crypt progenitors

(Silberg et al., 2000; Subramanian et al., 1998) and present in tamoxifen-exposed mutant intestines (Fig. 3B), might compensate for Cdx2 loss in dividing cells, we crossed mice carrying the conditional *Cdx2*^{fV} allele with *Cdx1*$^{-/-}$ mice, which have skeletal but no intestinal defects (Subramanian et al., 1995). Tamoxifen-treated *Cdx1*$^{-/-}$*;Cdx2*$^{fV/fV}$*;Villin-Cre*$^{ER(T2)}$ mice became markedly sicker than mice lacking only *Cdx2* and survived fewer than 8 days after the first dose of hormone. Villi were stunted compared to single mutant littermates and the proliferation markers Ki67 and Pcna were expressed in significantly fewer crypt cells (Fig. 3J), revealing a redundant requirement for Cdx proteins in adult gut epithelial cell proliferation.

Consistent with the histologic analysis, mRNA profiling of isolated *Cdx2*$^{fV/fV}$*;Villin-Cre*$^{ER(T2)}$ jejunal epithelium revealed lower levels of 991 transcripts (dChip analysis (Li and Wong, 2001), $p<0.01$) and Gene Ontology term analysis of affected genes pointed overwhelmingly to mature enterocyte functions, including solute transport and lipid and organic acid metabolism (Fig. 4A). Death from starvation and diarrhea in the face of overtly intact intestinal villus structure is best explained by malabsorption. Indeed, enterocyte products required for terminal digestion and nutrient absorption were prominent among transcripts that drop after tamoxifen treatment (Supplemental Table S3). Representative examples such as *Dpp4*, *Lipe*, and *Abca1* showed lower expression throughout the small intestine (Fig. 4B). Thus, in agreement with the functions suggested by CDX2 occupancy in differentiated Caco-2 cells (Fig. 2E), Cdx2 is necessary for survival of adult mice and its absence profoundly affects gene expression in mature enterocytes. Moreover, genes affected by Cdx2 loss in mice are enriched in CDX2 occupancy near their human orthologs in differentiated Caco-2 cells, as shown in a representative example (Fig. 4C) and whole-genome analysis (Fig. 4D). These data further justify the use of Caco-2 cells to study gut epithelial differentiation and indicate conservation of Cdx2 function across species.

Cdx2 is required for the chromatin configuration associated with active enhancers

The *in vivo* requirement for Cdx2 in differentiated and proliferating cells provides a strong functional correlate for the distinctive site occupancy in different cell states. To investigate Cdx2 requirements in gene regulation, we examined its role in chromatin structure. We and others define active enhancers as distal chromatin elements carrying two strongly positioned, H3K4Me2-marked nucleosomes around a central region with destabilized or absent nucleosomes

(Fig. 1B). This designation facilitated genome-wide mapping of condition-sensitive epigenomic marks and triggered elucidation of CDX2 requirements *in vivo*. To determine if Cdx2 is necessary for the active chromatin configuration in differentiated mouse intestinal epithelium, we studied the consequence of its absence in tamoxifen-treated *Cdx2*$^{fV/fV}$*;Villin-Cre*$^{ER(T2)}$ mouse villi. We used MNase digestion followed by quantitative real-time PCR to interrogate CDX2-bound regions located near genes that are dysregulated in the mutant mice. We designed oligonucleotide primers to detect nucleosome-protected DNA at representative CDX2-binding sites near differentiation-associated genes and both flanking nucleosomes. Villus cells isolated from tamoxifen-treated sibling control mice gave the pattern predicted for active nucleosome positioning at all 9 tested regions (Fig. 4E and data not shown). This active nucleosome signature was not apparent at 5 of the 9 regions in Cdx2-depleted villus cells, which showed relatively higher MNase sensitivity at flanking nucleosomes and greater protection in the center. This labile, Cdx2-dependent pattern most likely reflects displacement of a center nucleosome when Cdx2 is available and its reappearance when Cdx2 is absent. Our results thus indicate that Cdx2 is necessary to produce or maintain key chromatin modifications at loci associated with mature intestinal epithelium, an activity that may account for its control over a large number of genes required for intestinal function.

CDX2 partners with different transcription factors in different cell states

The demonstration of distinct CDX2 binding sites and functions in intestinal progenitors and mature cells suggested the possibility of distinct classes of regulatory *cis*-elements. To address the particular possibility of condition-specific partnership with other TFs, we searched CDX2-occupied regions in the two different cell states for other consensus sequence motifs. At CDX2-occupied regions in proliferating Caco-2 cells, the SeqPos motif analysis tool (Lupien et al., 2008) detected significant over-representation of GATA binding sites (Fig. 5A); thus, in dividing cells GATA sequences preferentially abound not only in chromatin with an active enhancer configuration (Fig. 1E) but also very close to CDX2-occupied sites. By contrast, CDX2-occupied regions specific to differentiated cells showed notable enrichment of a canonical HNF4A motif, the same sequence that is enriched among enhancers activated in mature Caco-2 cells (Fig. 1F) but not among proliferation-specific CDX2 sites (Fig. 5A). Adding credence to these results, the FOXA motif, matching a factor previously implicated in regulating endoderm-derived tissues

(Sekiya et al., 2009; Zaret, 1999), was over-represented at CDX2 sites in both dividing and differentiated cells, whereas Wnt-responsive TCF/LEF motifs, associated with intestinal cell proliferation (Clevers, 2006), were enriched selectively in progenitors. Mirroring the sequence motif enrichments, *GATA6* mRNA levels decline with Caco-2 cell differentiation, as do those of the proliferation marker *MYC*; as noted previously (Soutoglou and Talianidis, 2002), *HNF4A* transcripts and the representative differentiation marker *CDH17* increase markedly over the same period (Fig. 5B). Thus, CDX2 may partner with specific TFs in different cellular states; in particular, with a GATA protein in replicating and with HNF4A in differentiated cells.

To test this possibility, we conducted ChIP-seq analysis for GATA6 and HNF4A in mature and proliferating Caco-2 cells. In agreement with the motif search results, we identified 10,594 GATA6 binding sites ($p \leq 10^{-10}$) in dividing cells and fewer than 500 sites in differentiated cells (Fig. 5C). Conversely, HNF4A bound more than twice as many sites in mature cells than in progenitors (Fig. 5D; 28,187 *vs* 12,119 sites, $p \leq 10^{-10}$). Regions occupied by GATA6 and HNF4A were highly conserved across species (Supplemental Fig. S6) and showed substantial enrichment of the expected sequence motifs (Fig 5C-D). Because CDX2 binding to DNA in replicating and mature Caco-2 cells correlates with active chromatin in the respective cell states (Fig. 2C), we asked if binding of these candidate partners is also associated with the chromatin state. Indeed, GATA6 preferentially occupies sites that carry the active-enhancer mark in proliferating but not in differentiated cells, whereas HNF4A binding shows a marked preference for chromatin that is selectively active in mature cells (Fig. 5E). Thus, although GATA6 occupancy is more state-selective than HNF4A binding, H3 modifications are tightly coupled to experimentally determined binding of all three proteins and our results identify putative co-regulatory complexes for CDX2 function. Corroborating the existence of such complexes, immunoprecipitation of CDX2 from Caco-2 nuclear extracts revealed association with GATA6 and HNF4A (Fig. 5F). These associations were not intrinsically cell state-specific, probably reflecting interactions at constitutively occupied regions (represented in the intersection set in Fig. 2A).

Importantly, in analysis of empirical binding, CDX2-binding sites specific to dividing cells were more likely to be identified as GATA6-binding sites in ChIP-seq on dividing cells, whereas CDX2-binding sites in differentiated cells were more likely to be identified as HNF4A binding sites in ChIP-seq on differentiated cells. Individual examples of this duality are illustrated in Fig.

6A and Supplemental Fig. S7. Furthermore, CDX2 and its condition-specific partners typically occupy DNA within at most a few hundred bp of each other (Fig. 6B), suggesting they target the same regulatory elements. These relationships are reflected well in heat maps for genome-wide binding of each factor in relation to CDX2 occupancy (Fig. 6C). These data collectively indicate that CDX2 interacts with intestinal DNA in state-specific pairings with GATA6 and HNF4A at regions of differentially modified chromatin.

DISCUSSION

Progenitor cells and their mature progeny elicit substantially different properties from genomes that are usually identical. Analysis of gene knockouts *in vivo* and lineage-specific *cis*-elements *in vitro* support the idea that key regulatory TFs participate materially in differentiation, using mechanisms that are diverse and incompletely understood. Our identification of the CDX2 recognition motif at enhancers active in differentiated Caco-2 cells prompted detailed investigation of a TF whose highly restricted tissue expression, activity at intestine-specific gene promoters *in vitro*, and embryonic requirement encompass the properties of a "master regulator" (Beck, 2004; Gao et al., 2009). Findings in Cdx2-null mice support this designation in the adult intestine and reveal its requirement for cellular functions, gene expression, and active enhancer chromatin in distinct states. We also present evidence for dynamic CDX2 associations in a regenerative tissue, including labile interactions with chromatin and other TFs that correlate strongly with gene expression in dividing and mature cells (Fig. 7). The sum of these results provides a comprehensive view of the actions and mechanisms of a critical lineage-specific TF.

Dual activity in progenitor and differentiated cells is a feature of several lineage-restricted TFs. For example, Pu.1 functions in blood progenitor commitment and proliferation as well as terminal macrophage differentiation (DeKoter et al., 1998), much as MITF does in melanocyte proliferation, survival, and maturation (McGill et al., 2002). In development, the worm FoxA homolog PHA4 functions throughout pharynx formation, as do Pax6 and its homologs in the metazoan eye (Ashery-Padan and Gruss, 2001). Even in the face of such examples, it is unclear if these factors directly control tissue-specific genes throughout a cell's ontogeny, if their genome associations are stable across cellular transitions, and if Cdx2's distinct activities in dividing and differentiated cells reflect a general property of lineage-determining TFs. The muscle regulator MyoD, for example, functions in both myoblast progenitors and mature myotubes, and its binding to DNA was recently reported to vary little during differentiation (Cao et al., 2010). In contrast, redistribution of CDX2 during intestinal cell maturation exposes diversity among mechanisms of key TFs. Investigation of other regulators, ideally coupled with delineation of chromatin modifications, will resolve whether their associations with DNA are largely invariant, as reported for MyoD, or dynamic, as we observe with CDX2.

In *C. elegans* pharynx development, PHA4 controls early and late genes by binding unique target sites with different affinities. Early genes contain high-affinity sites and bind first; late targets carry sites with distinct sequences and lower affinity, and engage only after PHA4 levels increase late in development (Gaudet and Mango, 2002). CDX2 levels change little during intestinal differentiation and the consensus CDX2 motifs we identified in different conditions are almost identical (Supplemental Fig. S4), indicating a limited role for primary DNA sequence or TF concentrations in directing occupancy. Hence, post-translational modifications, alternative partners, chromatin structure, or a combination of these factors, likely underlie CDX2's cell state-specific complexes. Known post-translational alterations of CDX2 do not affect DNA affinity *in vitro* (Gross et al., 2005; Rings et al., 2001) but could in principle affect alternative complexes. Biochemical analysis of modifications on CDX2 and its partner proteins might advance understanding of the mechanisms underlying nucleosome lability at target enhancers. In identifying GATA6 and HNF4A as state-specific CDX2 partners at differentially modified regions of chromatin, we take an important step toward characterizing tissue-restricted TFs that produce or interact with altered chromatin states to enable intestinal cell differentiation. Future work might determine if the factors are co-dependent or restructure chromatin in an ordered hierarchy; understanding these mechanisms will help define the networks that control intestinal cell transitions (Davidson and Levine, 2008). Although findings in mutant mice corroborate our conclusions, the bulk of the present analysis was conducted in a tumor cell line that replicates the distinction between progenitor and differentiated states only partially. Primary intestinal crypt and villus cells would provide a physiologic model to test these ideas.

Studies in cancer cell lines paradoxically implicate CDX2 as both an oncogene and a tumor suppressor (Aoki et al., 2003; Guo et al., 2004); context-dependent CDX2 occupancy and function may account for distinct roles in cell proliferation and differentiation and help explain the discrepancy. Contextual genome occupancy and cooperation with other TFs might also underlie CDX2 requirements in embryonic axial patterning and trophectoderm formation. Indeed, CDX2 binding in the corresponding tissues suggests remarkable diversity in target loci (our unpublished data), underscoring the idea that TF activities are strongly influenced by cellular context. Another TF with many target genes, PPARG, was recently shown to utilize different binding sites and partners in two cell types, adipocytes and macrophages (Lefterova et al., 2010).

The continually renewing gut epithelium, a frequent target of malignant transformation, is an ideal model system to study transcriptional mechanisms of differentiation. Our integrated approach toward genome-wide chromatin analysis, TF binding, and mRNA expression can also be extended to uncover mechanisms of lineage-specific gene regulation in other tissues.

EXPERIMENTAL PROCEDURES

Nucleosome-resolution ChIP and identification of mononucleosomes. Cells were collected in a digestion buffer (50 mM Tris-HCl, pH7.6, 1 mM $CaCl_2$, 0.2% Triton X-100, 5 mM Na butyrate, complete protease inhibitors) and treated with 0.2 U micrococcal nuclease (MNase, Sigma Aldrich) for 8 min at 37°C. The reaction was terminated by adding 5 mM EDTA in 10 mM Tris, pH7.6, and samples were dialyzed in chromatin RIPA buffer (10 mM Tris, pH7.6, 1 mM EDTA, 0.1% SDS, 0.1% Na deoxycholate, 1% Triton X-100) before overnight IP at 4^{o}C with Histone H3K27Ac (Abcam ab4729) or H3K4Me2 (Millipore 07-030) antibodies. IP material was subsequently handled as described below for conventional ChIP of sonicated genomic DNA. Mononucleosomes carrying H3K4Me2 or H3K27Ac marks were identified using Nucleosome Positioning from Sequencing (NPS) with default parameters (Zhang et al., 2008b), as described (Extended Experimental Procedures).

Detection of state-specific open chromatin structures and identification of transcription factor motifs. NPS was used to identify mononucleosomes with H3K4Me2 in progenitor cells. Two adjacent nucleosomes were declared a pair if the distance between their center locations was between 250 and 400 nucleotides. For each nucleosome pair, relative depletion of a nucleosome in the inter-nucleosomal region in one state against the other was estimated using the Nucleosome Stability-Destability (NSD) scoring scheme defined previously (He et al., 2010). In this context, the NSD score of a nucleosome pair indicates the degree of differential chromatin structure between proliferating and mature intestinal cells by quantifying the relative change in nucleosome signal between cell states. Thus, nucleosome pairs with larger NSD scores can be interpreted to represent genomic regions more open to TF binding in one cell state compared to static regions (i.e., regions with NSD scores ~0). Conversely, negative NSD scores correspond to open chromatin structure in the other state (Supplemental Table S1). A hierarchical mixture model was used to search for DNA motifs enriched at the center of nucleosome pairs (the region of depleted signal corresponding to the middle nucleosome). Enrichment of each motif represented in the TRANSFAC database was weighed by the NSD score distribution of other regions with similar enrichment levels of the same motif (He et al., 2010) (and Meyer CA, unpublished) and ranked using t-statistics.

Chromatin Immunoprecipitation (ChIP) and co-immunoprecipitation for transcription factors, ChIP-seq, and data analysis. Standard procedures were used for ChIP (Extended Experimental Procedures) on sub-confluent or 26-day post-confluent Caco-2 cells. For ChIP-seq, 10 ng each of ChIP and input DNA from up to 3 pooled experiments were processed for deep sequencing according to manufacturer's instructions (Illumina). Prior to sequencing, qPCR was used to verify that positive and negative control ChIP regions amplified in the linear range.

Binding peaks for CDX2, GATA6, and HNF4A were identified using Model-Based Analysis of ChIP-Seq (MACS) (Zhang et al., 2008a) with a p-value cutoff of 10^{-10} and default values for other parameters. Sequences were mapped to reference genomes using ELAND tools (Illumina software suite). All sequence information is relative to human genome build 18. CDX2 binding sites specific to proliferating or differentiated cells were delineated by eliminating sites that might potentially be occupied in the other condition had the much less stringent *p*-value cut-off of 10^{-3} been applied. Binding sites determined by ChIP-Seq were scanned for sequence motifs using the SeqPos algorithm (He et al., 2010) as described (Extended Experimental Procedures). Co-immunoprecipitation was done using standard techniques and anti-CDX2 (BD Pharmingen 560171; Extended Experimental Procedures). All ChIP-seq data have been deposited in GEO (Series GSE23436).

Association assessment of empirical transcription factor binding with paired nucleosomes and gene expression. To calculate the relative frequency of nucleosome pairs occupied by a TF (Figs. 2C, 5E), paired nucleosomes were first sorted according to their NSD scores and grouped into bins of 750. For each bin, the number of paired nucleosomes bound by the factor was counted and scaled with respect to binding in bins representing static chromatin (*i.e,*. paired nucleosomes with NSD score ~0). To determine association of protein binding with gene expression (Fig. 2D), Caco-2 expression microarray data (Fleet et al., 2003) (GEO accession number GSE 1614) were enumerated based on log-fold changes in differentiated over proliferating cells using SAM analysis and genes were binned into groups of 100. Similar to Heintzman *et al* (Heintzman et al., 2009), the ratio of the average number of CDX2 binding sites within 100 kb of genes belonging in the same bin and the average number of binding sites near all genes was calculated and transformed to a logarithmic scale.

Association assessment of transcription factor partners. Associations between binding partners, such as CDX2 and GATA6 in proliferating or CDX2 and HNF4A in mature cells (Fig. 6B), were interrogated using their ChIP-Seq signals. First, the union of binding sites of CDX2, GATA6, and HNF4A in both conditions was obtained and the ChIP signals of each individual protein at the union sites were computed using the equation $T = \frac{\sqrt{T_{TF} \times S_{input}} - \sqrt{T_{input} \times S_{TF}}}{\sqrt{S_{input}}}$. T and S stand for the maximum tag count for a binding site and the sequence depth, respectively; subscripts *TF* and *input* indicate a transcription factor or input. This equation quantifies relative signal strength over the input on account of the sequence depth of TF and input data sets.

Knockout mice. LoxP sites were placed flanking exon 2 of mouse *Cdx2* in a targeting construct containing an FRT-flanked Neomycin resistance (Neo^R) cassette immediately downstream of the 3' LoxP site. For ES cell screening, southern blots on HindIII-digested DNA were hybridized with 5' (chr5:148120111+148120519, mm9) and 3' (chr5:148107065-148107467, mm9) probes. Two properly targeted clones were injected into blastocysts. Chimeric progeny were crossed with R26R-FLP mice (Farley et al., 2000) to excise the Neo^R gene and resulting progeny were crossed with *Villin-Cre*$^{ER(T2)}$ transgenic mice (el Marjou et al., 2004), yielding Mendelian transmission of all alleles. Experimental and littermate control mice were treated with 1 mg tamoxifen by intraperitoneal injection for 5 consecutive days. Control mice included Cre^+ animals and no effects of Cre toxicity were observed. For RNA microarray analysis, mice were harvested on the 4th day after completion of the tamoxifen course. For MNase assays, mice were analyzed at least 1 week after conclusion of tamoxifen treatment. Histologic analysis was done 2 days after the last tamoxifen dose. *Cdx1*$^{-/-}$ mice were maintained and genotyped as described (Subramanian et al., 1998).

Expression analyses. Cell and tissue RNAs were harvested using Trizol reagent (Invitrogen) and reverse transcribed using SuperScript enzyme (Invitrogen) or labeled for microarray (Affymetrix). cDNA was assessed by qPCR using gene-specific primers (Extended Experimental Procedures) and SYBR green master mix (Applied Biosystems). Additional details are found in Extended Experimental Procedures. For immunostaining, routine histological techniques were

applied (Extended Experimental Procedures) using the following antibodies: CDX2 (Biogenex, 1:50), Ki67 (Novocastra Laboratories, 1:1000), Chromogranin A (Immunostar, 1:500), Lysozyme (Invitrogen, 1:50), and PCNA (NeoMarkers, 1:1000). All expression data have been deposited to GEO (Series GSE23436).

Gene association and GO analyses were performed using MetaCore and DAVID Gene Functional Analysis tools (http://david.abcc.ncifcrf.gov/) as described (Extended Experimental Procedures).

Association of CDX2 binding sites in Caco-2 cells with genes dysregulated in Cdx2-deficient intestine. To determine associations between transcripts that change upon Cdx2 loss in mouse intestine and CDX2 binding near their human orthologs in Caco-2 (Fig. 4D), we first used Ensembl BioMart (http://uswest.ensembl.org/index.html) to match mouse genes with human orthologs, which were then arranged according to expression levels of the corresponding mouse gene and grouped into bins of 100. For each such bin, the ratio between the average number of CDX2 binding sites within 100 kb of genes in the same bin and the average number of binding sites near all genes was estimated and displayed (Fig. 4D).

MNase protection. Cells were harvested from knockout or control mouse intestinal villi and treated with MNase as described above. Instead of incubation with antibody, DNA was purified after Proteinase K digestion and compared to control samples (no MNase treatment) by qPCR using the indicated oligonucleotides (Extended Experimental Procedures).

Cell proliferation. Proliferating (50% confluent) and 26-day post-confluent Caco-2 cells were treated for 1 h with EdU. DNA synthesis was measured by fluorescent detection of EdU according to manufacturer's protocols (Invitrogen).

ACKNOWLEDGMENTS

Supported by National Institutes of Health (NIH) grants RC2CA148222 and R01DK082889 (RAS), R01HG004069 (XSL), R01DK054111 (JCF), and P50CA127003 (Dana-Farber/Harvard Cancer Center), and a gift from the Caring for Carcinoid Foundation (RAS). MPV was supported by NIH training grant T32DK07477 and Fellowship No. 1987 from the Crohn's & Colitis Foundation of America. We thank Steve Krasinski for critical appraisal of the work; Madeline Lemieux and Andrew Kung for helpful discussions; Sylvie Robine for permission to use, and Eric Martin for providing, *Villin-Cre*$^{ER(T2)}$ transgenic mice; and Peter Gruss for permission to use, and Gary Gilliland and Sumin Koo for providing, *Cdx1*$^{-/-}$ mice.

AUTHOR CONTRIBUTIONS

MPV and RAS conceived the study, analyzed data, and wrote the manuscript; MPV, RS and RKM performed experiments; MPV, HS, HHH, CAM, JCF and XSL developed the algorithms and performed computational analyses; HS, JCF, MB and XSL interpreted data and edited the report.

REFERENCES

Aoki, K., Tamai, Y., Horiike, S., Oshima, M., and Taketo, M. M. (2003). Colonic polyposis caused by mTOR-mediated chromosomal instability in Apc+/Delta716 Cdx2+/- compound mutant mice. Nat Genet *35*, 323-330.

Ashery-Padan, R., and Gruss, P. (2001). Pax6 lights-up the way for eye development. Curr Opin Cell Biol *13*, 706-714.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Beck, F. (2004). The role of Cdx genes in the mammalian gut. Gut *53*, 1394-1396.

Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., 3rd, Gingeras, T. R.*, et al.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169-181.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Cao, Y., Yao, Z., Sarkar, D., Lawrence, M., Sanchez, G. J., Parker, M. H., MacQuarrie, K. L., Davison, J., Morgan, M. T., Ruzzo, W. L.*, et al.* (2010). Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. Dev Cell *18*, 662-674.

Carroll, J. S., Meyer, C. A., Song, J., Li, W., Geistlinger, T. R., Eeckhoute, J., Brodsky, A. S., Keeton, E. K., Fertuck, K. C., Hall, G. F.*, et al.* (2006). Genome-wide analysis of estrogen receptor binding sites. Nat Genet *38*, 1289-1297.

Chawengsaksophak, K., James, R., Hammond, V. E., Kontgen, F., and Beck, F. (1997). Homeosis and intestinal tumours in Cdx2 mutant mice. Nature *386*, 84-87.

Choi, M. Y., Romer, A. I., Hu, M., Lepourcelet, M., Mechoor, A., Yesilaltay, A., Krieger, M., Gray, P. A., and Shivdasani, R. A. (2006). A dynamic expression survey identifies transcription factors relevant in mouse digestive tract development. Development *133*, 4119-4129.

Clevers, H. (2006). Wnt/beta-catenin signaling in development and disease. Cell *127*, 469-480.

D'Angelo, A., Bluteau, O., Garcia-Gonzalez, M. A., Gresh, L., Doyen, A., Garbay, S., Robine, S., and Pontoglio, M. (2010). Hepatocyte nuclear factor 1alpha and beta control terminal differentiation and cell fate commitment in the gut epithelium. Development *137*, 1573-1582.

Davidson, E. H., and Levine, M. S. (2008). Properties of developmental gene regulatory networks. Proc Natl Acad Sci USA *105*, 20063-20066.

DeKoter, R. P., Walsh, J. C., and Singh, H. (1998). PU.1 regulates both cytokine-dependent proliferation and differentiation of granulocyte/macrophage progenitors. EMBO J *17*, 4456-4468.

el Marjou, F., Janssen, K. P., Chang, B. H., Li, M., Hindie, V., Chan, L., Louvard, D., Chambon, P., Metzger, D., and Robine, S. (2004). Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. Genesis *39*, 186-193.

Fang, R., Santiago, N. A., Olds, L. C., and Sibley, E. (2000). The homeodomain protein Cdx2 regulates lactase gene promoter activity during enterocyte differentiation. Gastroenterology *118*, 115-127.

Farley, F. W., Soriano, P., Steffen, L. S., and Dymecki, S. M. (2000). Widespread recombinase expression using FLPeR (flipper) mice. Genesis *28*, 106-110.

Fleet, J. C., Wang, L., Vitek, O., Craig, B. A., and Edenberg, H. J. (2003). Gene expression profiling of Caco-2 BBe cells suggests a role for specific signaling pathways during intestinal differentiation. Physiol Genomics *13*, 57-68.

Gao, N., White, P., and Kaestner, K. H. (2009). Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. Dev Cell *16*, 588-599.

Gaudet, J., and Mango, S. E. (2002). Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. Science *295*, 821-825.

Grainger, S., Savory, J. G., and Lohnes, D. (2010). Cdx2 regulates patterning of the intestinal epithelium. Dev Biol *In Press*.

Gross, I., Lhermitte, B., Domon-Dell, C., Duluc, I., Martin, E., Gaiddon, C., Kedinger, M., and Freund, J. N. (2005). Phosphorylation of the homeotic tumor suppressor Cdx2 mediates its ubiquitin-dependent proteasome degradation. Oncogene *24*, 7955-7963.

Guo, R. J., Suh, E. R., and Lynch, J. P. (2004). The role of Cdx proteins in intestinal development and cancer. Cancer Biol Ther *3*, 593-601.

Halbleib, J. M., Saaf, A. M., Brown, P. O., and Nelson, W. J. (2007). Transcriptional modulation of genes encoding structural characteristics of differentiating enterocytes during development of a polarized epithelium in vitro. Mol Biol Cell *18*, 4261-4278.

He, H. H., Meyer, C. A., Shin, H., Bailey, S., Wei, G., Wang, Q., Zhany, Y., Xu, K., Ni, M., Lupien, M*., et al.* (2010). Positioned nucleosomes flanking a labile nucleosome characterize transcriptional enhancers. Nat Genet *42*, 343-347.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W*., et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A*., et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Hinoi, T., Lucas, P. C., Kuick, R., Hanash, S., Cho, K. R., and Fearon, E. R. (2002). CDX2 regulates liver intestine-cadherin expression in normal and malignant colon epithelium and intestinal metaplasia. Gastroenterology *123*, 1565-1577.

James, R., Erler, T., and Kazenwadel, J. (1994). Structure of the murine homeobox gene cdx-2. Expression in embryonic and adult intestinal epithelium. J Biol Chem *269*, 15229-15237.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941-945.

Lefterova, M. I., Steger, D. J., Zhuo, D., Qatanani, M., Mullican, S. E., Tuteja, G., Manduchi, E., Grant, G. R., and Lazar, M. A. (2010). Cell-specific determinants of peroxisome proliferator-activated receptor gamma function in adipocytes and macrophages. Mol Cell Biol *30*, 2078-2089.

Li, C., and Wong, W. H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci USA *98*, 31-36.

Liu, T., Zhang, X., So, C. K., Wang, S., Wang, P., Yan, L., Myers, R., Chen, Z., Patterson, A. P., Yang, C. S., and Chen, X. (2007). Regulation of Cdx2 expression by promoter methylation, and effects of Cdx2 transfection on morphology and gene expression of human esophageal epithelial cells. Carcinogenesis *28*, 488-496.

Lupien, M., Eeckhoute, J., Meyer, C. A., Wang, Q., Zhang, Y., Li, W., Carroll, J. S., Liu, X. S., and Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell *132*, 958-970.

McGill, G. G., Horstmann, M., Widlund, H. R., Du, J., Motyckova, G., Nishimura, E. K., Lin, Y. L., Ramaswamy, S., Avery, W., Ding, H. F.*, et al.* (2002). Bcl2 regulation by the melanocyte master regulator Mitf modulates lineage survival and melanoma cell viability. Cell *109*, 707-718.

Molkentin, J. D., and Olson, E. N. (1996). Defining the regulatory networks for muscle development. Curr Opin Genet Dev *6*, 445-453.

Mouchel, N., Henstra, S. A., McCarthy, V. A., Williams, S. H., Phylactides, M., and Harris, A. (2004). HNF1alpha is involved in tissue-specific regulation of CFTR gene expression. Biochem J *378*, 909-918.

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell *123*, 917-929.

Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E.*, et al.* (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517-527.

Potten, C. S. (1998). Stem cells in gastrointestinal epithelium: numbers, characteristics and death. Philos Trans R Soc Lond B Biol Sci *353*, 821-830.

Rings, E. H., Boudreau, F., Taylor, J. K., Moffett, J., Suh, E. R., and Traber, P. G. (2001). Phosphorylation of the serine 60 residue within the Cdx2 activation domain mediates its transactivation capacity. Gastroenterology *121*, 1437-1450.

Saaf, A. M., Halbleib, J. M., Chen, X., Yuen, S. T., Leung, S. Y., Nelson, W. J., and Brown, P. O. (2007). Parallels between global transcriptional programs of polarizing Caco-2 intestinal epithelial cells in vitro and gene expression programs in normal colon and colon cancer. Mol Biol Cell *18*, 4245-4260.

Sekiya, T., Muthurajan, U. M., Luger, K., Tulin, A. V., and Zaret, K. S. (2009). Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. Genes Dev *23*, 804-809.

Silberg, D. G., Sullivan, J., Kang, E., Swain, G. P., Moffett, J., Sund, N. J., Sackett, S. D., and Kaestner, K. H. (2002). Cdx2 ectopic expression induces gastric intestinal metaplasia in transgenic mice. Gastroenterology *122*, 689-696.

Silberg, D. G., Swain, G. P., Suh, E. R., and Traber, P. G. (2000). Cdx1 and cdx2 expression during intestinal development. Gastroenterology *119*, 961-971.

Soutoglou, E., and Talianidis, I. (2002). Coordination of PIC assembly and chromatin remodeling during differentiation-induced gene activation. Science *295*, 1901-1904.

Stegmann, A., Hansen, M., Wang, Y., Larsen, J. B., Lund, L. R., Ritie, L., Nicholson, J. K., Quistorff, B., Simon-Assmann, P., Troelsen, J. T., and Olsen, J. (2006). Metabolome, transcriptome, and bioinformatic cis-element analyses point to HNF-4 as a central regulator of gene expression during enterocyte differentiation. Physiol Genomics *27*, 141-155.

Struhl, K. (1999). Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. Cell *98*, 1-4.

Subramanian, V., Meyer, B., and Evans, G. S. (1998). The murine Cdx1 gene product localises to the proliferative compartment in the developing and regenerating intestinal epithelium. Differentiation *64*, 11-18.

Subramanian, V., Meyer, B. I., and Gruss, P. (1995). Disruption of the murine homeobox gene Cdx1 affects axial skeletal identities by altering the mesodermal expression domains of Hox genes. Cell *83*, 641-653.

Traber, P. G., and Silberg, D. G. (1996). Intestine-specific gene transcription. Annu Rev Physiol *58*, 275-297.

Tremblay, E., Auclair, J., Delvin, E., Levy, E., Menard, D., Pshezhetsky, A. V., Rivard, N., Seidman, E. G., Sinnett, D., Vachon, P. H., and Beaulieu, J. F. (2006). Gene expression profiles of normal proliferating and differentiating human intestinal epithelial cells: a comparison with the Caco-2 cell model. J Cell Biochem *99*, 1175-1186.

van de Wetering, M., Sancho, E., Verweij, C., de Lau, W., Oving, I., Hurlstone, A., van der Horn, K., Batlle, E., Coudreuse, D., Haramis, A. P.*, et al.* (2002). The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell *111*, 241-250.

Zaret, K. (1999). Developmental competence of the gut endoderm: genetic potentiation by GATA and HNF3/fork head proteins. Dev Biol *209*, 1-10.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008a). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zhang, Y., Shin, H., Song, J. S., Lei, Y., and Liu, X. S. (2008b). Identifying positioned nucleosomes with epigenetic marks in human from ChIP-Seq. BMC Genomics *9*, 537.

FIGURE LEGENDS

Figure 1. An epigenomic screen identifies CDX2 as a candidate regulator of intestinal epithelial maturation. **A**, Caco-2 cells show progenitor properties while proliferating (left); after reaching confluence, the cells differentiate to resemble mature intestinal villus enterocytes (right). Histograms depict EdU incorporation, a marker of DNA replication, 1 h after exposure, as measured by fluorescence intensity. Diagrams in the upper right represent cell morphologies. **B**, Chromatin modifications associated with nascent (left) and active (right) enhancers are depicted, the latter characterized by relative increase in H3K4Me2 on flanking nucleosomes and depletion of signal in the nucleosome in between. **C**, H3K4Me2 ChIP-seq results in proliferating and differentiated Caco-2 cells at a representative region on chromosome 10. **D**, H3K4Me2 ChIP-seq tag counts across indicated 1-kb intervals in proliferating (left, blue) and differentiated (right, red) cells, including high- resolution image of the data shown in C. The top 2 panels show gain of active chromatin structure in mature cells; the lower panel shows a region selectively active in progenitors. **E**, Composite plots of H3K4Me2 ChIP-seq tag counts from the 1,000 regions with the greatest differential chromatin structure in proliferating cells (left, blue) and the same 1,000 regions in differentiated cells (right, red), aligned at center nucleosomes. TF motifs most significantly over-represented at the valleys of the composite plots are shown (details in Experimental Procedures). A complete list of identified motifs is found in Table S1. **F**, The same analysis as in E, but on the 1,000 regions with the greatest differential chromatin structure in differentiated cells.

Figure 2. CDX2 interacts dynamically with the genome during intestinal cell differentiation. **A,** Venn diagram representation of binding sites identified by CDX2 ChIP-seq in proliferating and mature Caco-2 cells. **B,** Histograms of normalized CDX2 ChIP-seq tags at representative sites bound exclusively in proliferating (left) or in differentiated cells (right). Additional examples appear in Figure S3. **C,** Condition-specific CDX2 binding corresponds to condition-specific active chromatin. H3K4Me2-marked putative *cis*-elements were ranked from most differentially active chromatin signatures in proliferating (left) to differentiated cells (right) and plotted along the x-axis in bins of 750. CDX2 occupancy was then plotted relative to occupancy in areas where active chromatin did not vary (center). CDX2 binding sites specific to

proliferating (blue triangles) or differentiated cells (red squares) correlated with proliferating and differentiated cell-specific active chromatin, respectively. **D,** Heat maps demonstrating correlation of stage-specific CDX2 occupancy with stage-specific gene expression. Top: Genes differentially expressed in proliferating cells (left, blue) versus differentiated cells (right, red) were ranked and binned in groups of 100; genes with unaltered expression are represented in the center. The color map (blue - red) presents the scale of $\log_2$ fold change (-1 to +1) in gene expression levels in the two stages. Bottom: The frequency of condition-specific CDX2 binding within 100 kb of each gene expression bin is indicated by the intensity of yellow shading. Numbers below the color map correspond to the minimum, mean, and maximum ratios (0.5, 1, and 2.4, respectively) between the average number of CDX2 binding sites within 100kb from the genes in the same bin over the average number of binding sites near all genes. **E,** Gene Ontology term analysis reveals that expressed genes with condition-specific CDX2 binding match functions classically attributed to differentiated (red) or proliferating (blue) cells. Complete analysis is found in Table S2.

Figure 3. *Cdx2* is required for proliferation and differentiation of the adult mouse intestinal epithelium. **A,** Immunostaining verifies absence of intestinal Cdx2 expression in tamoxifen-treated *Cdx2*$^{fV/fV}$*;Villin-Cre*$^{ER(T2)}$ mice (also see Supplemental Fig. S5). **B,** Reduced mRNA levels of *Cdx2* and candidate target genes *Sis* and *Isx* in tamoxifen-treated *Cdx2*$^{fV/fV}$*;Villin-Cre*$^{ER(T2)}$ intestines; *Cdx1* transcript levels are unaffected. **C,** *Cdx2*$^{fV/fV}$*;Cre*$^{ER(T2)}$ mice rapidly lose weight after tamoxifen treatment. **D,** Intestines of *Cdx2*$^{fV/fV}$*;Villin-Cre*$^{ER(T2)}$ mice are dilated and the cecum is particularly engorged with fluid. **E,** H&E staining reveals intact tissue architecture in the *Cdx2* mutant ileum. **F,** A profound deficiency is evident in Alkaline Phosphatase activity, a marker of enterocytes, especially in the distal small intestine. **G-I,** Periodic acid-Schiff (G), Lysozyme (H) and ChromograninA (I) were used as markers of goblet, Paneth, and enteroendocrine cells, respectively, in the ileum. **J,** In contrast to the respective single gene mutants, *Cdx1*$^{-/-}$*;Cdx2*$^{fV/fV}$*;Villin-Cre*$^{ER(T2)}$ intestinal crypts show markedly reduced expression of the proliferative markers Ki67 and Pcna. Quantitation of the Ki67^{+} fraction of crypt cells showed significant reduction in the double mutant ($P<0.0001$; two tailed t-test). Box plot indicate median Ki67^{+} counts per crypt, with lower and upper quartiles and whiskers

indicating minimum and maximum counts. Error bars are SEM for B and C. Scale bars, 33.3 μm except for F, 200uM.

Figure 4. Cdx2 is required for active chromatin structure *in vivo*. A, Gene Ontology terms associated with mature intestinal functions are notably enriched among transcripts reduced in *Cdx2$^{fV/fV}$;Villin-Cre$^{ER(T2)}$* intestinal mucosa. **B,** Examples of terminal digestive enzyme transcripts reduced in Cdx2-null duodenum, jejunum, and ileum and **C,** a representative example of binding data from Caco-2 cells showing differentiated cell-specific CDX2 occupancy (arrow). **D**, Top: Transcripts reduced or unchanged in *Cdx2$^{fV/fV}$;Villin-Cre$^{ER(T2)}$* mice were detected in triplicate samples and grouped into bins of 100 by a ratio relative to controls (-0.5 to 0). Bottom: Differentiated cell-specific CDX2 binding frequency within 100 kb of the corresponding orthologs in Caco-2 cells, indicated by intensity of yellow shading (a range of 0.5-2.4, similar to Fig. 2d). Binding is observed more frequently near loci with reduced expression, suggesting direct regulation of many targets and inter-species conservation. **E,** To determine the consequence of Cdx2 deletion on chromatin structure, oligonucleotides were designed to query the left, center, and right nucleosomes by PCR in MNase protection assays at nine CDX2-bound putative enhancers in control and tamoxifen-treated *Cdx2$^{fV/fV}$;Villin-Cre$^{ER(T2)}$* (KO) adult mice. MNase-protected DNA is shown relative to the central nucleosome at each region. In villi depleted of Cdx2, MNase protection was relatively enhanced at center nucleosomes, consistent with reappearance of nucleosomes and loss of active enhancer chromatin. Five of the 9 tested regions gave this pattern; 4 examples are shown. Results are depicted schematically on the right, together with 1 example where nucleosome protection was not significantly affected in KO mice. Reductions in transcript levels for each gene in the Cdx2-depleted tissue are indicated. All error bars are SEM.

Figure 5. Differential occupancy of proliferating and differentiated intestinal cell genomes by GATA6 and HNF4A. A, DNA motif analysis at condition-specific CDX2-binding regions shows, in addition to CDX2 motifs, substantial enrichment of GATA motifs in proliferating (left) and of HNF4A motifs in differentiated cells (right). **B,** Over the course of Caco-2 differentiation, *HNF4A* transcript levels increase and *GATA6* levels decrease; known markers of maturation (*CDH17*) or proliferation (*MYC*) increase or decrease, respectively, as expected. Error bars are

SEM. **C-D,** ChIP-seq for GATA6 and HNF4A in proliferating (blue) and differentiated (red) cells reveals differences in preferred binding contexts. Venn diagrams indicate the number of sites bound uniquely in proliferating or differentiated cells ($p<1e^{-10}$). HNF4A binding sites are more numerous in mature cells, whereas GATA6 binds predominantly in progenitors. For each occupancy dataset, the most enriched *de novo* binding motif is indicated (all within the statistical limit of $p<1e^{-30}$,) and each is similar to empirically defined TRANSFAC motifs (as shown in Fig. 3A). Identified binding regions were highly conserved and mainly found far from promoters (Figure S6). **E,** As in figure 2C, TF occupancy at regions of differentially active chromatin is compared to binding frequency at static chromatin regions. Top, GATA6 binding in progenitors is better associated with the active H3K4Me2 pattern in proliferating than in differentiated cells, whereas HNF4A occupancy (bottom) correlates better with active H3K4Me2 in differentiated cells. **F**, GATA6 and HNF4A co-immunoprecipitate with CDX2, but not with IgG controls, indicating their presence in CDX2 complexes.

Figure 6. Condition-specific CDX2 binding is accompanied by GATA6 in proliferating cells and HNF4A in differentiated cells. **A**, Individual examples of ChIP-seq in proliferating (blue) and differentiated (red) cells show condition-specific GATA6 and CDX2 occupancy at the same region in proliferating but not in differentiated cells; HNF4A and CDX2 co-occupy a region bound selectively in differentiated cells. Condition-specific H3K4me2 signals are evident at the regions selectively occupied by the TFs and demarcated here by the dotted lines. Additional examples appear in Figure S7. **B**, Frequency of co-occupancy among TFs in proliferating and differentiated cells as a function of the distance between peak factor-binding sites. Condition-specific CDX2 binding sites more frequently lie near GATA6-bound regions in proliferating cells and near HNF4A-occupied regions in mature cells. The frequency of overlap in binding peaks within a 300bp distance (dotted line) is indicated in pie charts (right). **C,** Heat maps of all condition-specific CDX2-bound sites, showing the extent of GATA6 and HNF4A co-occupancy at these regions and confirming the relationship between these factors in binding across the genome. Sites are ordered by robustness of CDX2 occupancy in each state of differentiation. The measure of each TF's occupancy was defined as the relative ChIP signal in relation to the input signal at binding sites (see Experimental Procedures for mathematical details) and distributed between -4 and 12 (color map scale).

Figure 7. Model for dynamic CDX2 function during intestinal differentiation. In proliferating cells, CDX2 binds with relative selectivity, frequently accompanied by GATA6, at H3K4Me2 regions with selectively open chromatin at loci active in progenitor cells. Genes associated with epithelial maturity are not expressed in these cells and their enhancers have a relatively closed structure with less H3K4Me2. As cells differentiate, CDX2 relocates from progenitor-specific target genes to those associated with maturity, now accompanied by HNF4A; the appearance of an active H3K4Me2 pattern at many loci depends on CDX2.

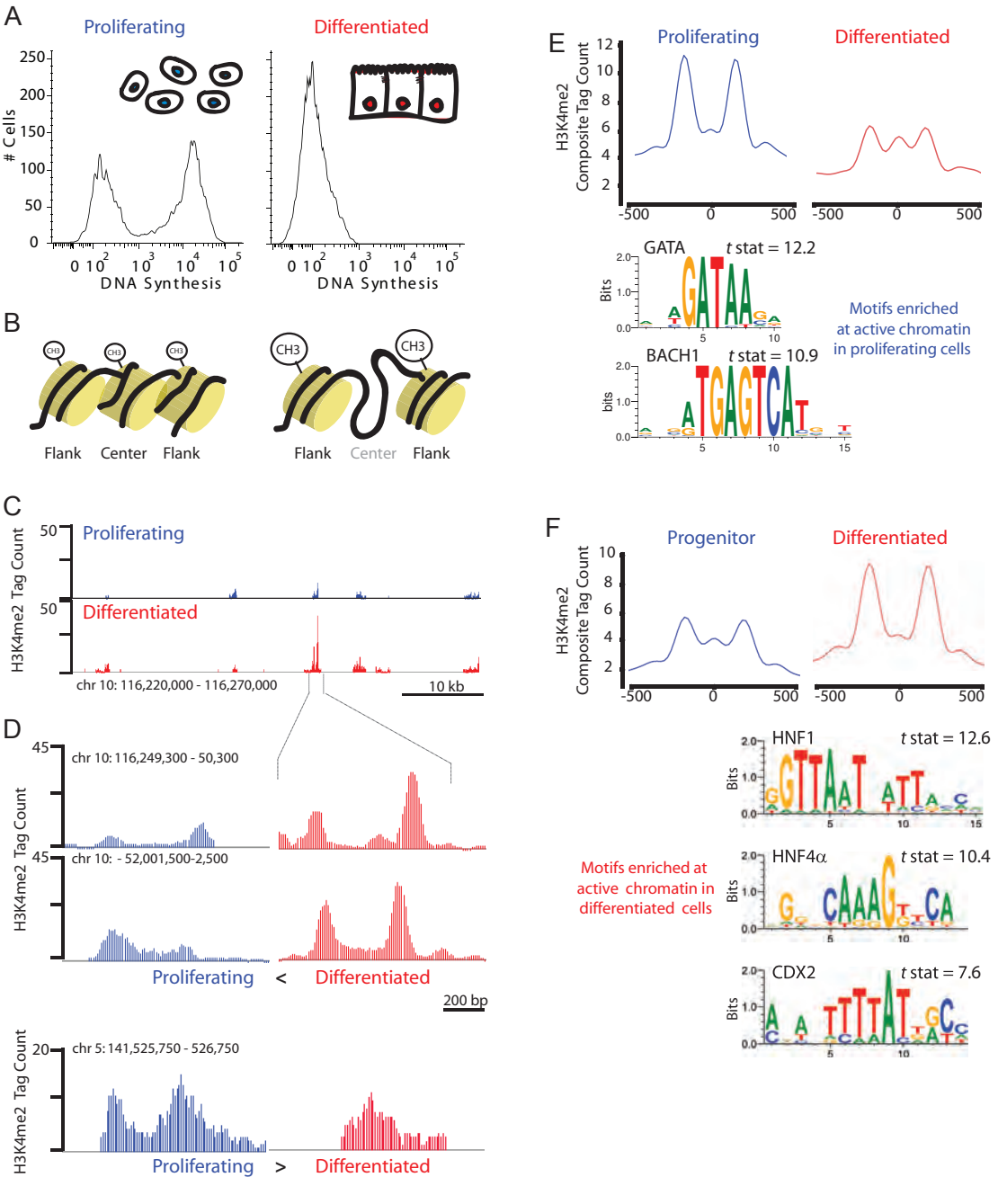

A
Proliferating
Differentiated
Cells
DNA Synthesis
B
CH3
Flank
Center
Flank
C
H3K4me2 Tag Count
Proliferating
Differentiated
chr 10: 116,220,000 - 116,270,000
10 kb
D
chr 10: 116,249,300 - 50,300
chr 10: - 52,001,500-2,500
Proliferating < Differentiated
200 bp
chr 5: 141,525,750 - 526,750
Proliferating > Differentiated
E
H3K4me2 Composite Tag Count
Proliferating
Differentiated
GATA
t stat = 12.2
BACH1
t stat = 10.9
Bits
Motifs enriched at active chromatin in proliferating cells
F
Progenitor
Differentiated
HNF1
t stat = 12.6
HNF4α
t stat = 10.4
CDX2
t stat = 7.6
Motifs enriched at active chromatin in differentiated cells

Figure 1

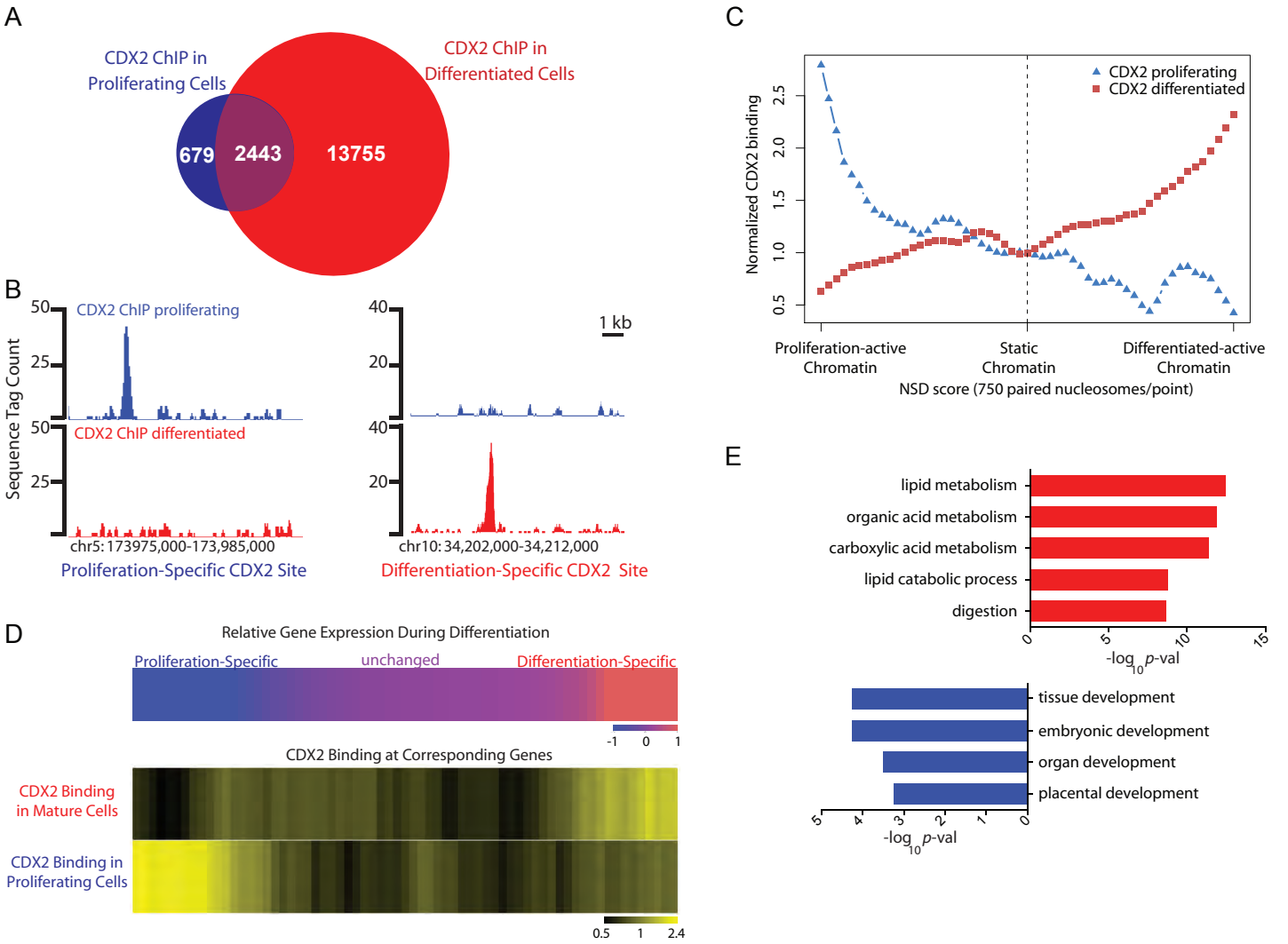

A
CDX2 ChIP in Proliferating Cells
CDX2 ChIP in Differentiated Cells
679
2443
13755
B
CDX2 ChIP proliferating
CDX2 ChIP differentiated
Sequence Tag Count
1 kb
chr5: 173975,000-173,985,000
chr10: 34,202,000-34,212,000
Proliferation-Specific CDX2 Site
Differentiation-Specific CDX2 Site
C
CDX2 proliferating
CDX2 differentiated
Normalized CDX2 binding
Proliferation-active Chromatin
Static Chromatin
Differentiated-active Chromatin
NSD score (750 paired nucleosomes/point)
D
Relative Gene Expression During Differentiation
Proliferation-Specific
unchanged
Differentiation-Specific
CDX2 Binding at Corresponding Genes
CDX2 Binding in Mature Cells
CDX2 Binding in Proliferating Cells
E
lipid metabolism
organic acid metabolism
carboxylic acid metabolism
lipid catabolic process
digestion
-log10 p-val
tissue development
embryonic development
organ development
placental development
-log10 p-val

Figure 2

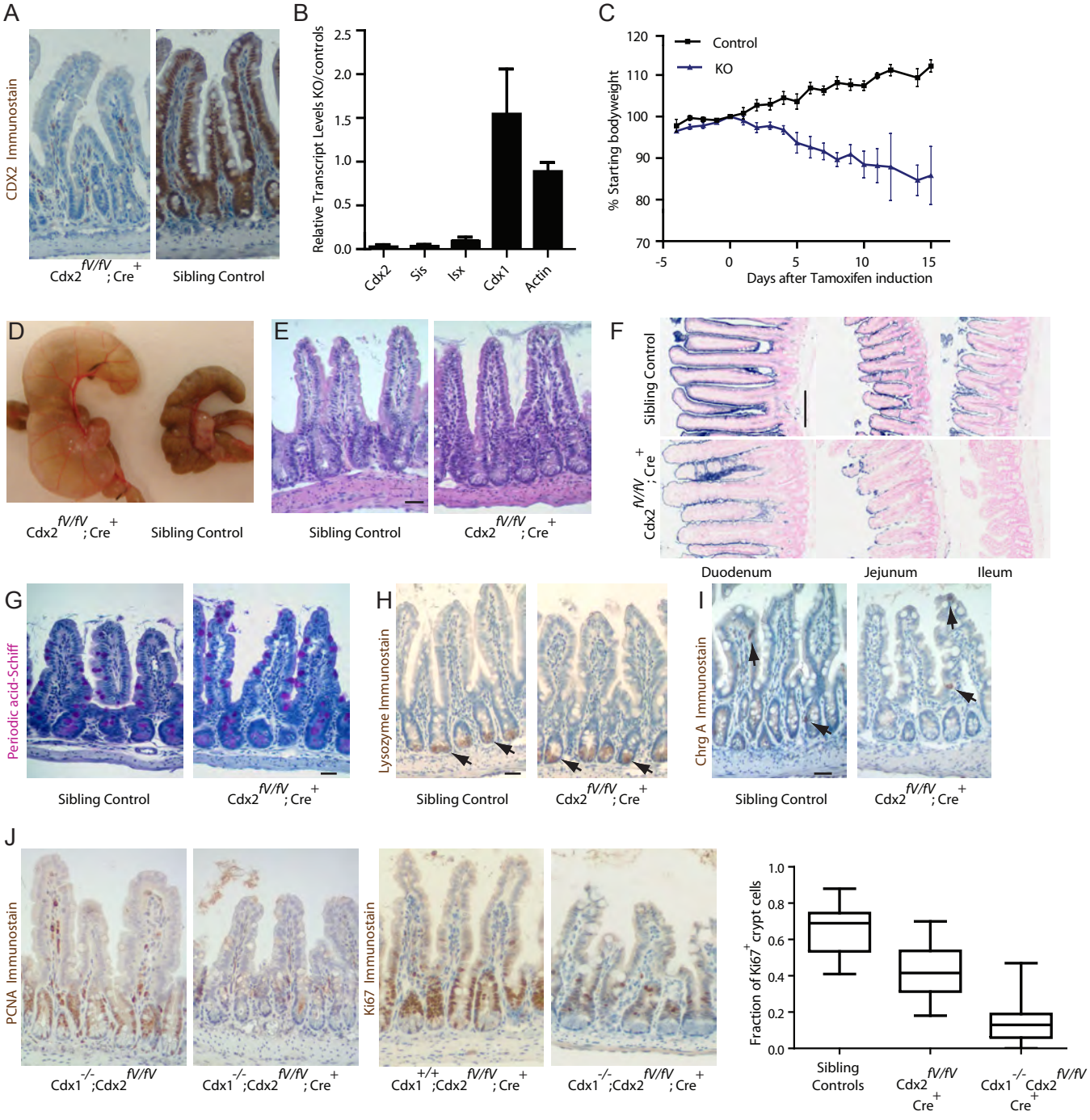

A
CDX2 Immunostain
Cdx2 fV/fV ; Cre +
Sibling Control
B
Relative Transcript Levels KO/controls
2.5
2.0
1.5
1.0
0.5
0.0
Cdx2
Sis
Isx
Cdx1
Actin
C
Control
KO
% Starting bodyweight
120
110
100
90
80
70
-5
0
5
10
15
Days after Tamoxifen induction
D
Cdx2 fV/fV ; Cre +
Sibling Control
E
Sibling Control
Cdx2 fV/fV ; Cre +
F
Sibling Control
Cdx2 fV/fV ; Cre +
Duodenum
Jejunum
Ileum
G
Periodic acid-Schiff
Sibling Control
Cdx2 fV/fV ; Cre +
H
Lysozyme Immunostain
Sibling Control
Cdx2 fV/fV ; Cre +
I
Chrg A Immunostain
Sibling Control
Cdx2 fV/fV ; Cre +
J
PCNA Immunostain
Cdx1 -/- ;Cdx2 fV/fV
Cdx1 -/- ;Cdx2 fV/fV ; Cre +
Ki67 Immunostain
Cdx1 +/+ ;Cdx2 fV/fV ; Cre +
Cdx1 -/- ;Cdx2 fV/fV ; Cre +
Fraction of Ki67+ crypt cells
1.0
0.8
0.6
0.4
0.2
0.0
Sibling Controls
Cdx2 fV/fV Cre +
Cdx1 -/- Cdx2 fV/fV Cre +

Figure 3

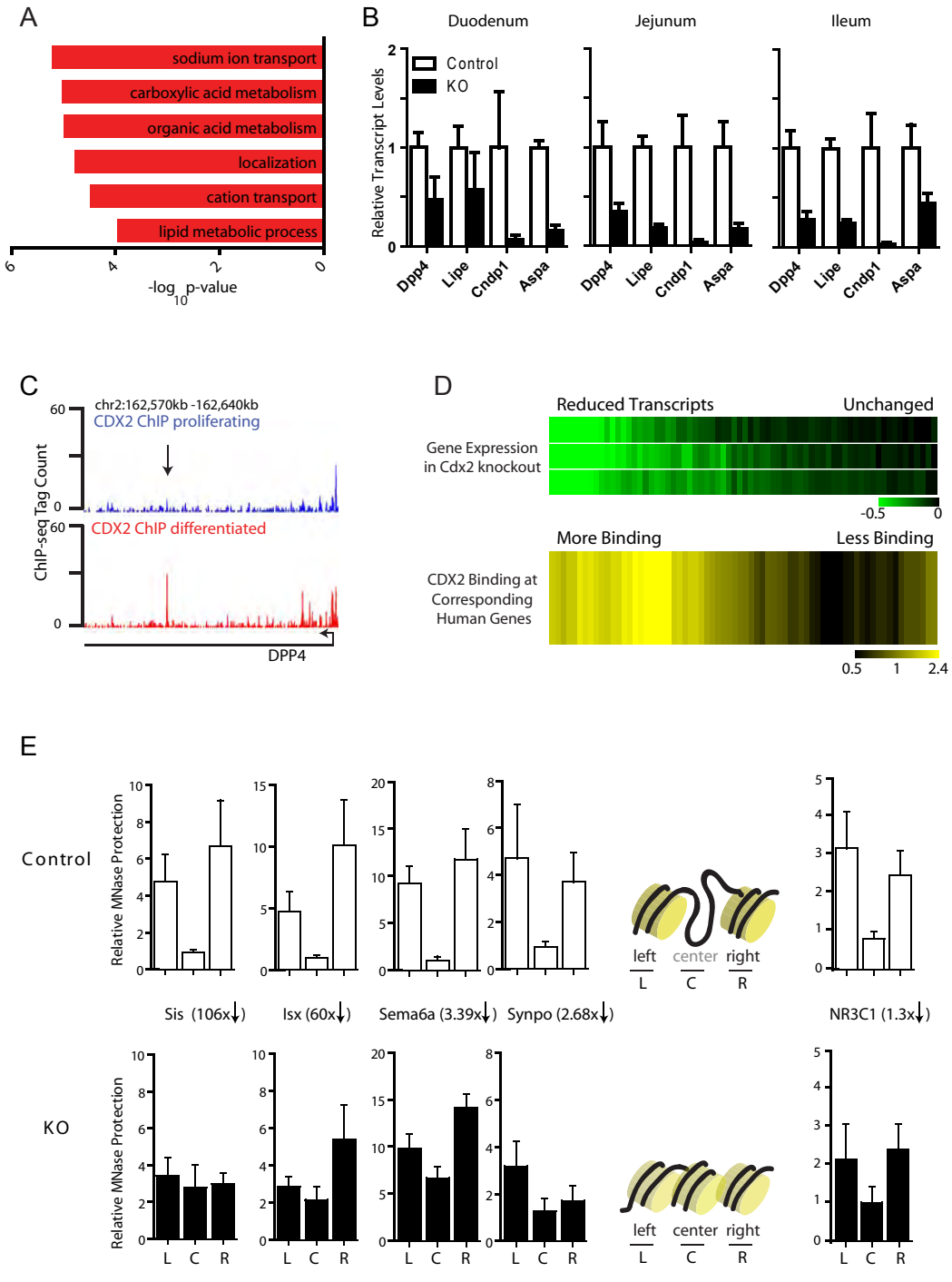

A
sodium ion transport
carboxylic acid metabolism
organic acid metabolism
localization
cation transport
lipid metabolic process
6
4
2
0
-log10 p-value
B
Duodenum
Jejunum
Ileum
Control
KO
Relative Transcript Levels
2
1
0
Dpp4
Lipe
Cndp1
Aspa
C
chr2:162,570kb -162,640kb
CDX2 ChIP proliferating
CDX2 ChIP differentiated
ChIP-seq Tag Count
60
0
DPP4
D
Reduced Transcripts
Unchanged
Gene Expression in Cdx2 knockout
-0.5
0
More Binding
Less Binding
CDX2 Binding at Corresponding Human Genes
0.5
1
2.4
E
Control
KO
Relative MNase Protection
Sis (106x↓)
Isx (60x↓)
Sema6a (3.39x↓)
Synpo (2.68x↓)
NR3C1 (1.3x↓)
left
center
right
L
C
R

Figure 4

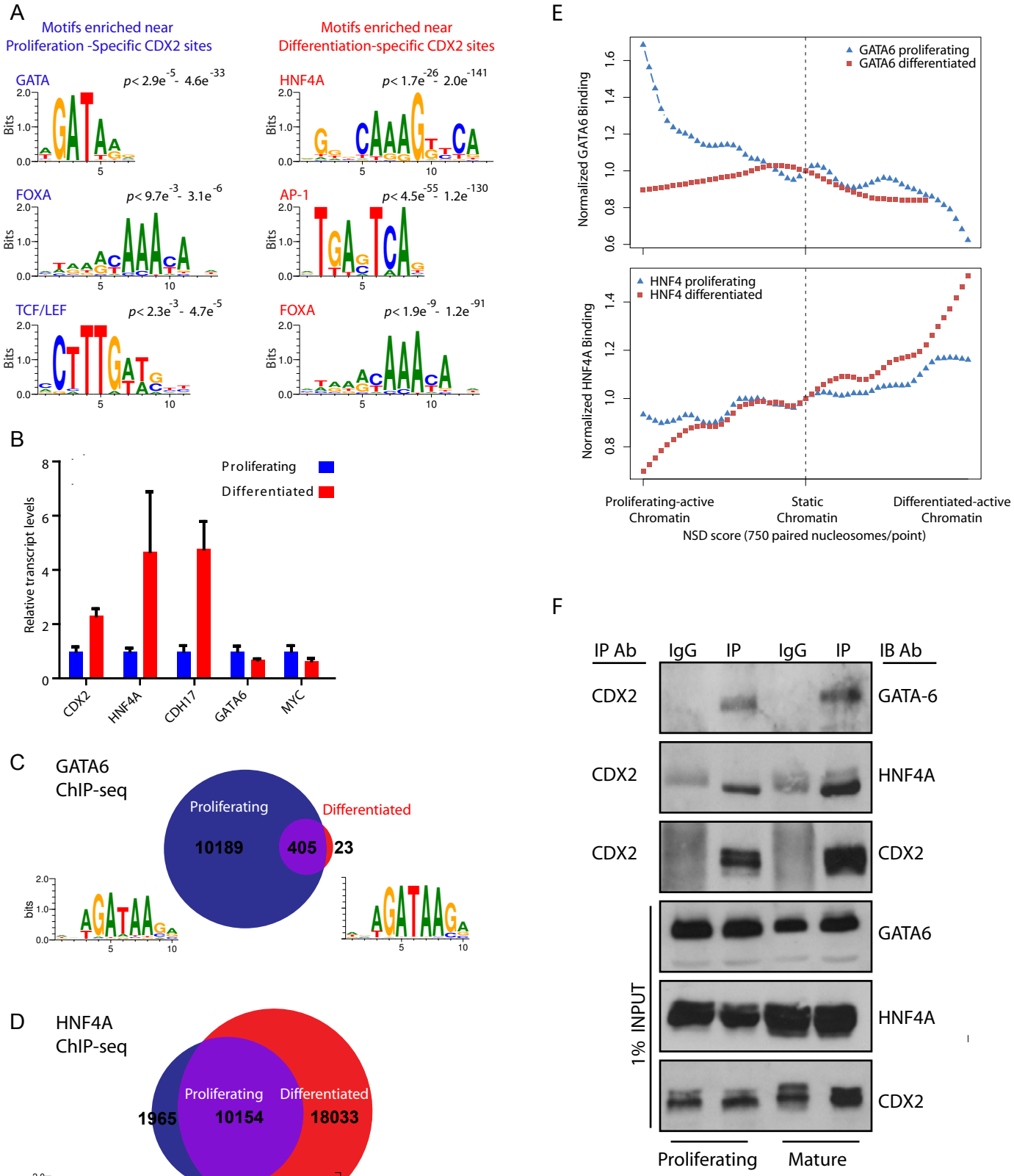

A
Motifs enriched near
Proliferation -Specific CDX2 sites
Motifs enriched near
Differentiation-specific CDX2 sites
GATA
p< 2.9e-5 - 4.6e-33
HNF4A
p< 1.7e-26 - 2.0e-141
FOXA
p< 9.7e-3 - 3.1e-6
AP-1
p< 4.5e-55 - 1.2e-130
TCF/LEF
p< 2.3e-3 - 4.7e-5
FOXA
p< 1.9e-9 - 1.2e-91
Bits
B
Proliferating
Differentiated
Relative transcript levels
CDX2
HNF4A
CDH17
GATA6
MYC
C
GATA6
ChIP-seq
Proliferating
Differentiated
10189
405
23
bits
D
HNF4A
ChIP-seq
Proliferating
Differentiated
1965
10154
18033
bits
E
GATA6 proliferating
GATA6 differentiated
Normalized GATA6 Binding
HNF4 proliferating
HNF4 differentiated
Normalized HNF4A Binding
Proliferating-active
Chromatin
Static
Chromatin
Differentiated-active
Chromatin
NSD score (750 paired nucleosomes/point)
F
IP Ab
IgG
IP
IgG
IP
IB Ab
CDX2
GATA-6
CDX2
HNF4A
CDX2
CDX2
1% INPUT
GATA6
HNF4A
CDX2
Proliferating
Mature

Figure 5

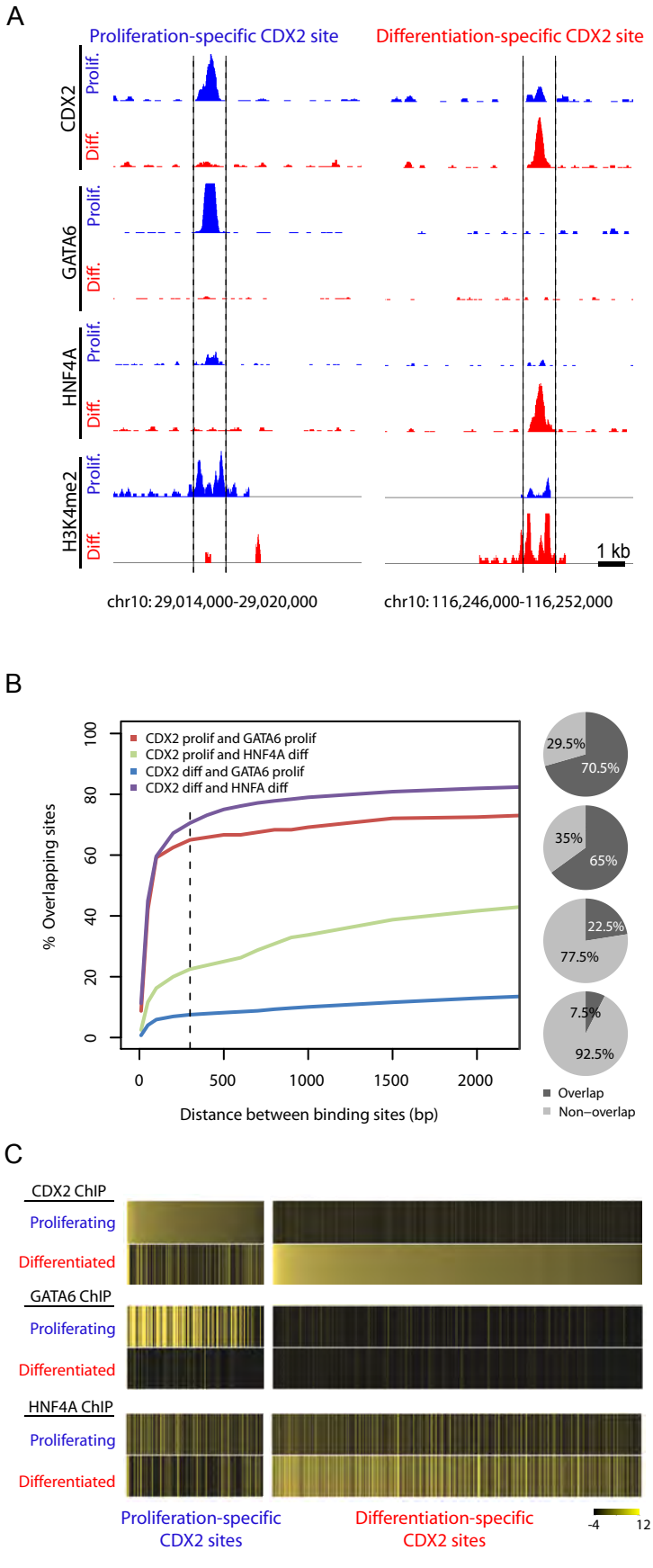
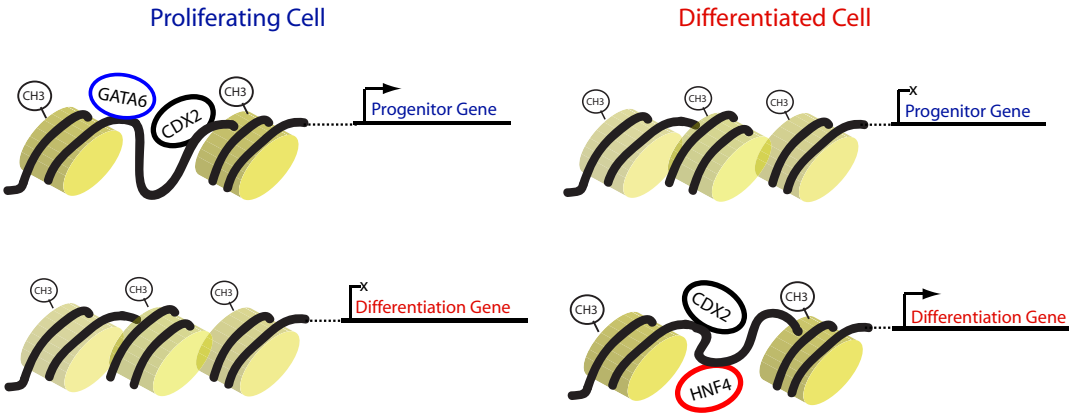

A
Proliferation-specific CDX2 site
Differentiation-specific CDX2 site
CDX2
GATA6
HNF4A
H3K4me2
Prolif.
Diff.
1 kb
chr10: 29,014,000-29,020,000
chr10: 116,246,000-116,252,000
B
CDX2 prolif and GATA6 prolif
CDX2 prolif and HNF4A diff
CDX2 diff and GATA6 prolif
CDX2 diff and HNFA diff
% Overlapping sites
Distance between binding sites (bp)
29.5%
70.5%
35%
65%
22.5%
77.5%
7.5%
92.5%
Overlap
Non−overlap
C
CDX2 ChIP
GATA6 ChIP
HNF4A ChIP
Proliferating
Differentiated
Proliferation-specific CDX2 sites
Differentiation-specific CDX2 sites
-4
12

Figure 6


Proliferating Cell
Differentiated Cell
CH3
GATA6
CDX2
Progenitor Gene
Differentiation Gene
HNF4

Inventory of Supplemental Information:

Figure S1. Illustrates similarity between H3K4Me2 and H3K27Ac data related to Figure 1.

Figure S2. Describes the sequencing depth of each ChIP-seq experiment, indicating further that sequencing depths were at or near saturation for detection of binding sites (peaks). This relates to Figures 1, 2, and 5.

Table S1. Lists the full results from the epigenomic screen implemented in Figure 1.

Figure S3. Provides additional examples of the data presented in Figure 2.

Figure S4. Data demonstrating quality of the CDX2 ChIP-seq results summarized in Figures 2-3.

Table S2. Provides the full results of GO analysis from which data in Figures 2-3 are derived.

Table S3. Lists transcripts reduced most significantly in CDX2 conditional-null intestines.

Figure S5. Illustrates the strategy for generating CDX2 conditional knockout mice and the efficiency of Cdx2 depletion. Supplements the results presented in Figures 3-4.

Figure S6. Quality control for GATA6 and HNF4A ChIP-seq results presented in Figures 5-6.

Figure S7. Summary examples of integrated ChIP-seq data from Figures 1, 2, 5, and 6.

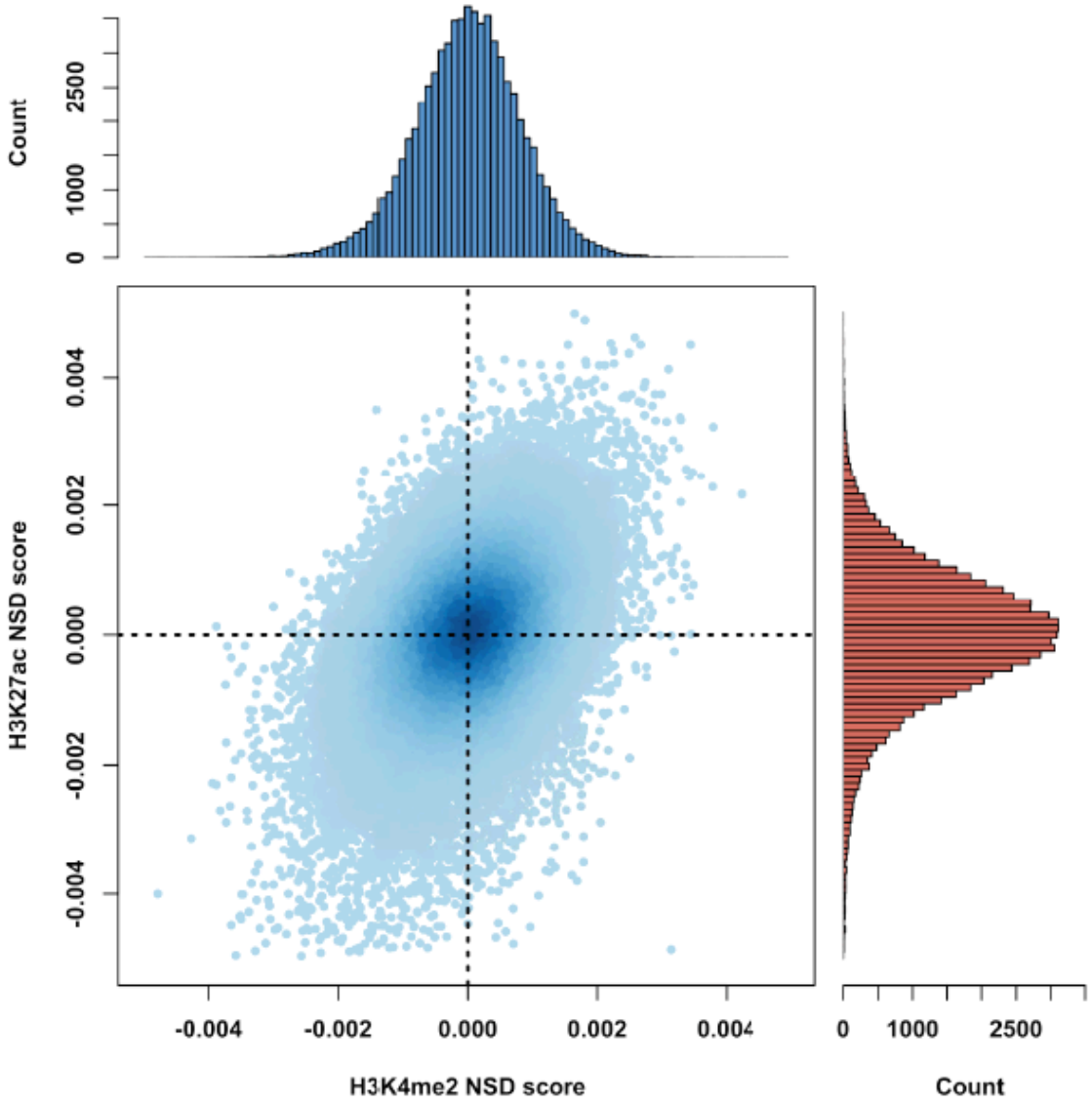
Included below are:

Supplemental Figures 1-7

Supplemental Tables 1-3

Extended Experimental Procedures

Supplemental References

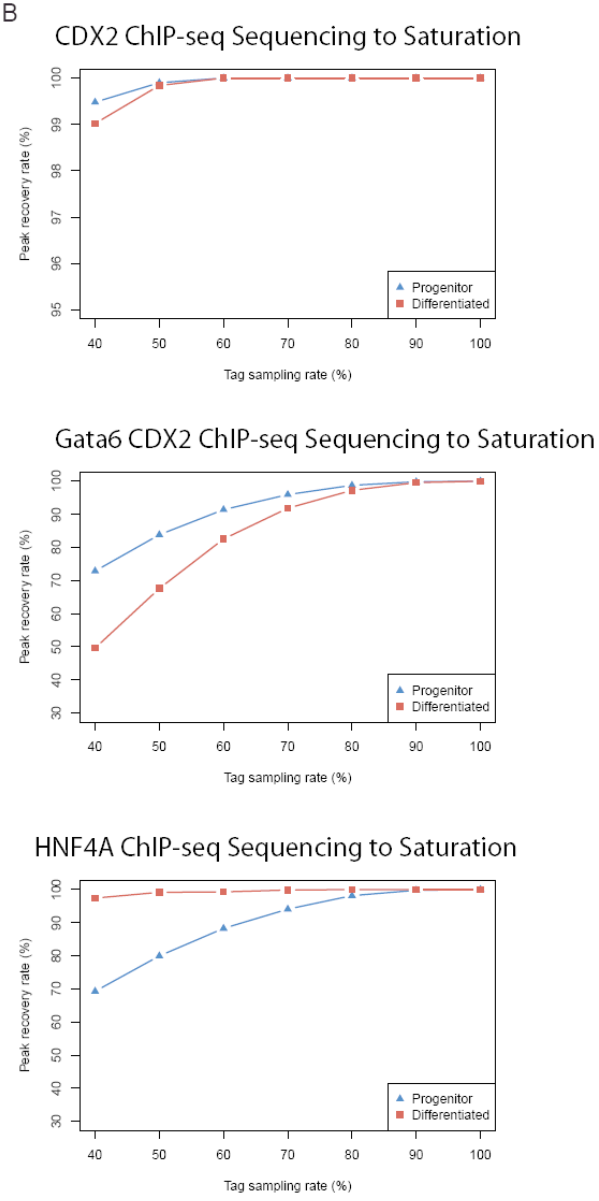


Supplemental Figure 1. Comparison of H3K4Me2 and H3K27Ac NSD scores reveals a strong correlation of active-enhancer patterns for the two H3 modifications during Caco-2 cell differentiation. Differentially active chromatin regions were identified from H3K4Me2 or H3K27Ac ChIP-seq, as described in the text. Regions showing differential activity (as defined and plotted by NSD score) in proliferating or differentiated cells with one mark tend to show similar differential activity with the other (scatter plot). Histograms also show the distribution of NSD scores (H3K4Me2 on top and H3K27Ac on the right). As results from both experiments were similar, only H3K4Me2 data are presented in this report, for simplicity.

A

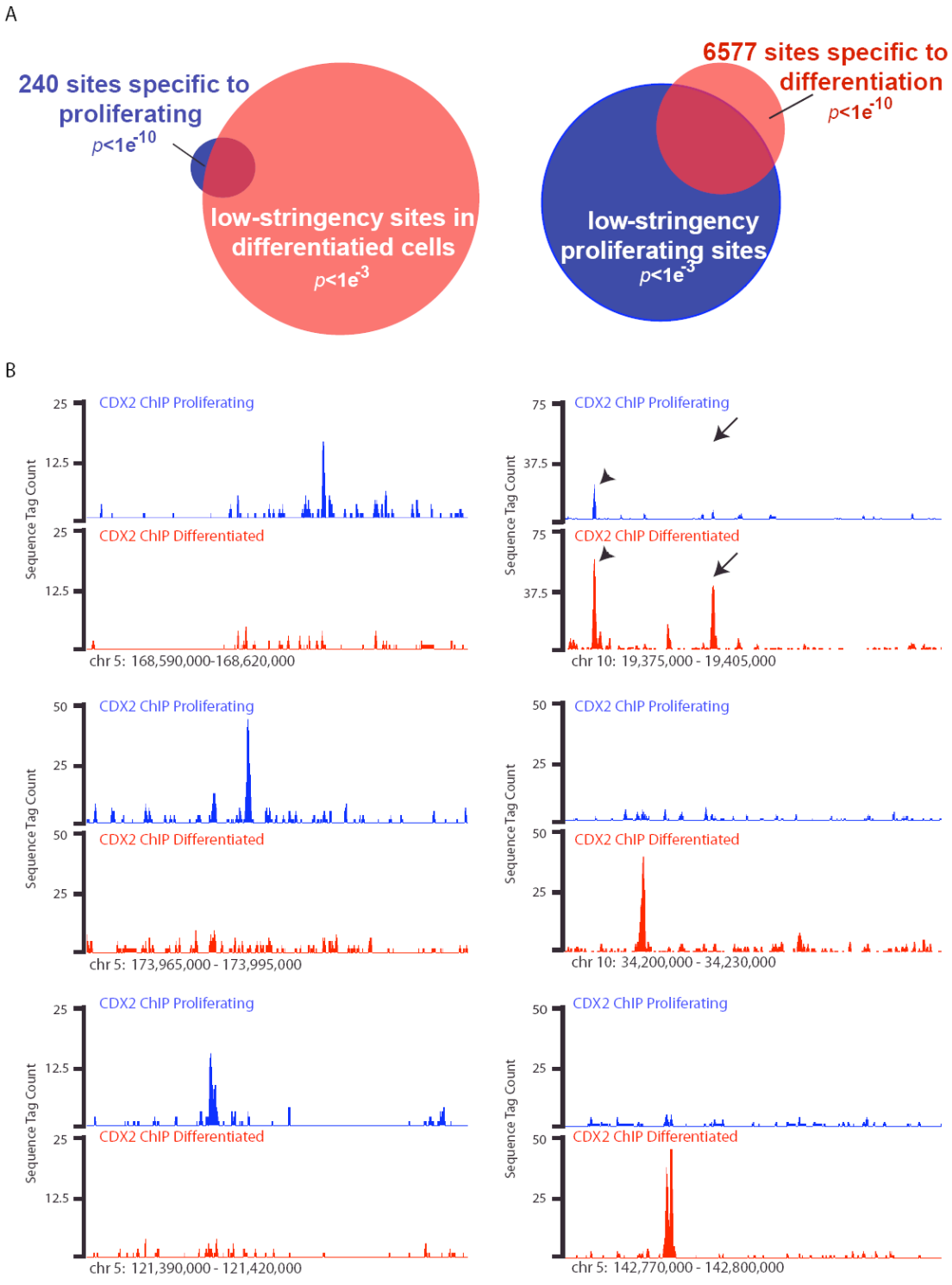
Antibody	Cell State	Total tag number	Uniquely mapped tags	MACS peaks	positioned nucleosomes
CDX2	proliferating	16518515	10962632	3122	NA
CDX2	differentiated	22897414	14995565	16198	NA
GATA6	proliferating	21340553	8855924	10594	NA
GATA6	differentiated	23579311	9178995	428	NA
HNF4A	proliferating	23333744	11278924	12119	NA
HNF4A	differentiated	23346845	13132116	28187	NA
H3K4me2	proliferating	21043206	14481117	86491	219266
H3K4me2	differentiated	21639827	13035690	96931	208762
H3K27ac	proliferating	21608020	14103038	37174	182107
H3K27ac	differentiated	20871655	14628901	46094	193496

B



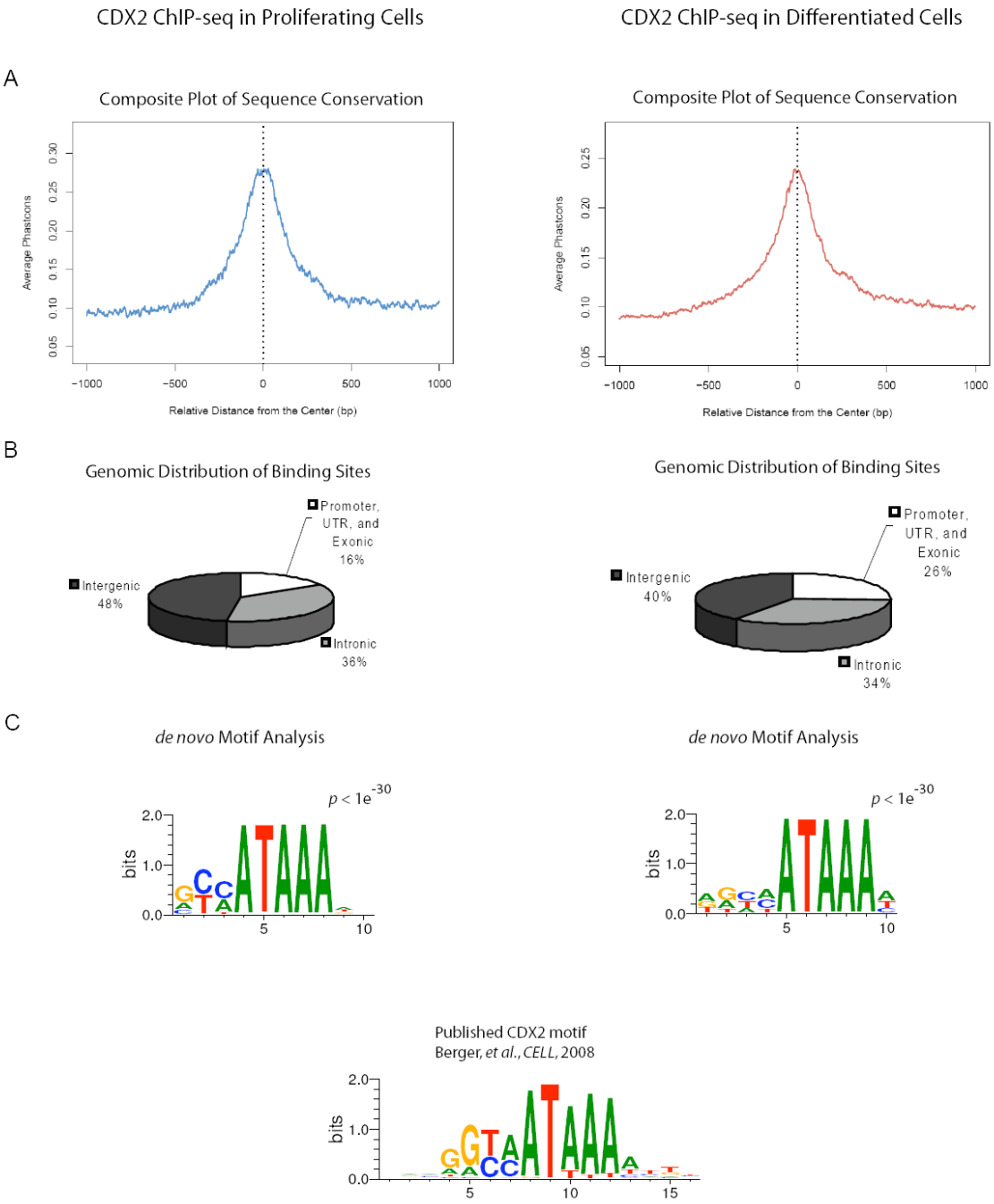
Supplemental Figure 2

Supplemental Figure 2. Summary of ChIP-seq results and Peak Calling. A, Summary table of ChIP targets, cellular differentiation state, sequencing depth, number of uniquely mapped sequences, and resulting peaks called by MACS(Zhang et al., 2008a) or positioned nucleosomes identified by NPS(Zhang et al., 2008b). **B,** Graphs of the percent of total peaks identified (y-axis) for any subset of total sequencing tags (x-axis), indicating that sequencing depth for each ChIP-seq experiment was adequate to saturate the number of binding sites observed at a normalized fold-change of >20 in excess of the input.

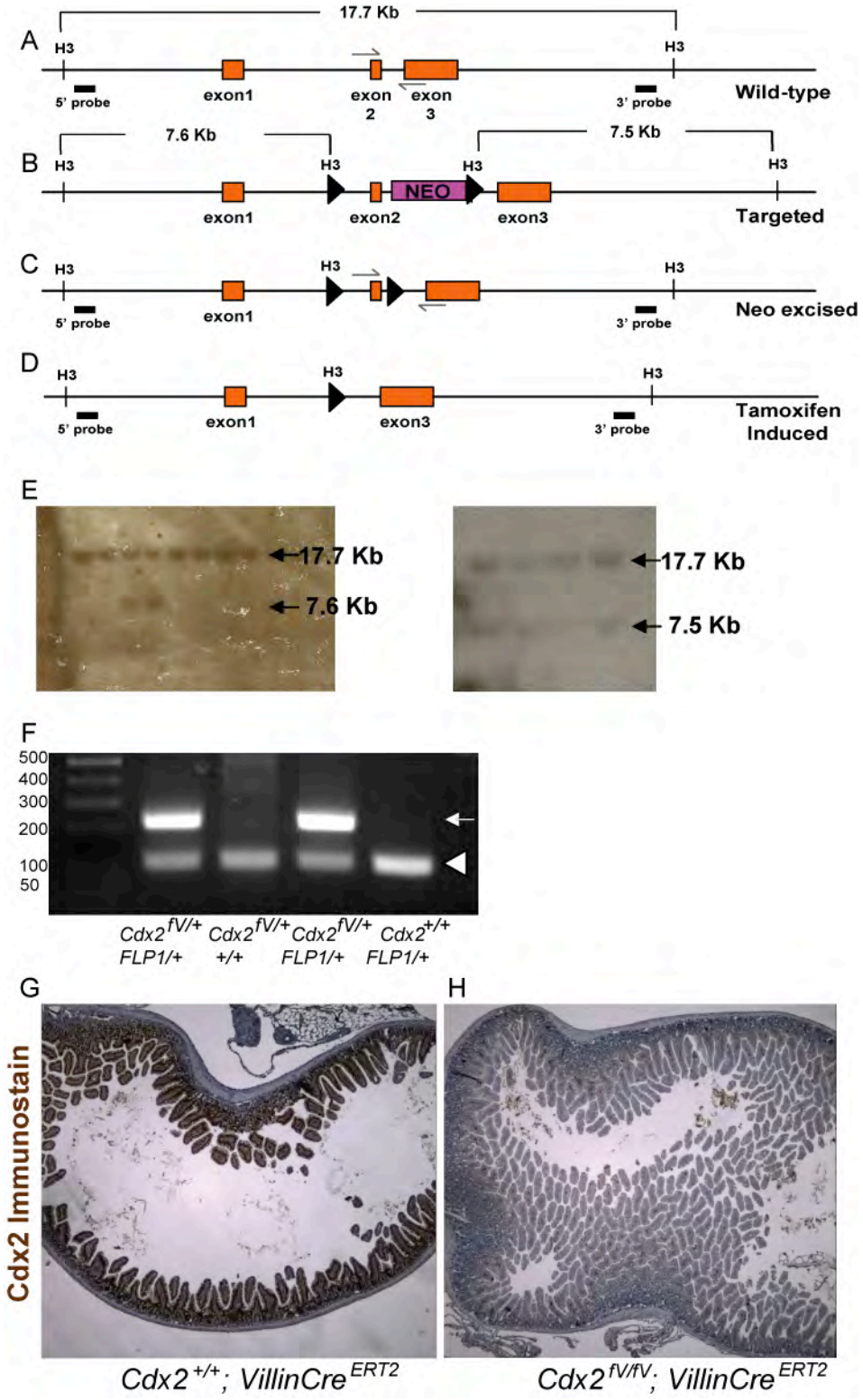


Supplemental Figure 3. Delineation of condition-specific CDX2 occupancy. Although comparison of CDX2 occupancy, determined by ChIP-seq in proliferating and differentiated cells ($p<10^{-10}$ for both states) revealed sites with distinct regulatory properties, such regions were not always exclusively occupied in one condition. Areas showing occupancy at high confidence

in one condition ($p<10^{-10}$) sometimes showed occupancy with lower confidence ($p<10^{-3}$) in the other. To rigorously delineate sites that are uniquely bound in each state, we subtracted low-stringency sites in one condition ($p<10^{-3}$) from high-stringency sites in the other (**A**). This approach yielded selective sets of 240 regions uniquely occupied in proliferating and 6577 exclusive sites in differentiated cells. **B**, Examples of these regions demonstrate strict condition-specific CDX2 occupancy in proliferating (left) or in differentiated cells (right). The sample intergenic region near the *ARL5B* locus (top right panel) showed differential affinity in proliferating and differentiated cells at one site (arrows) but failed to satisfy the strict requirements for condition-specific binding at a second site (arrowheads).

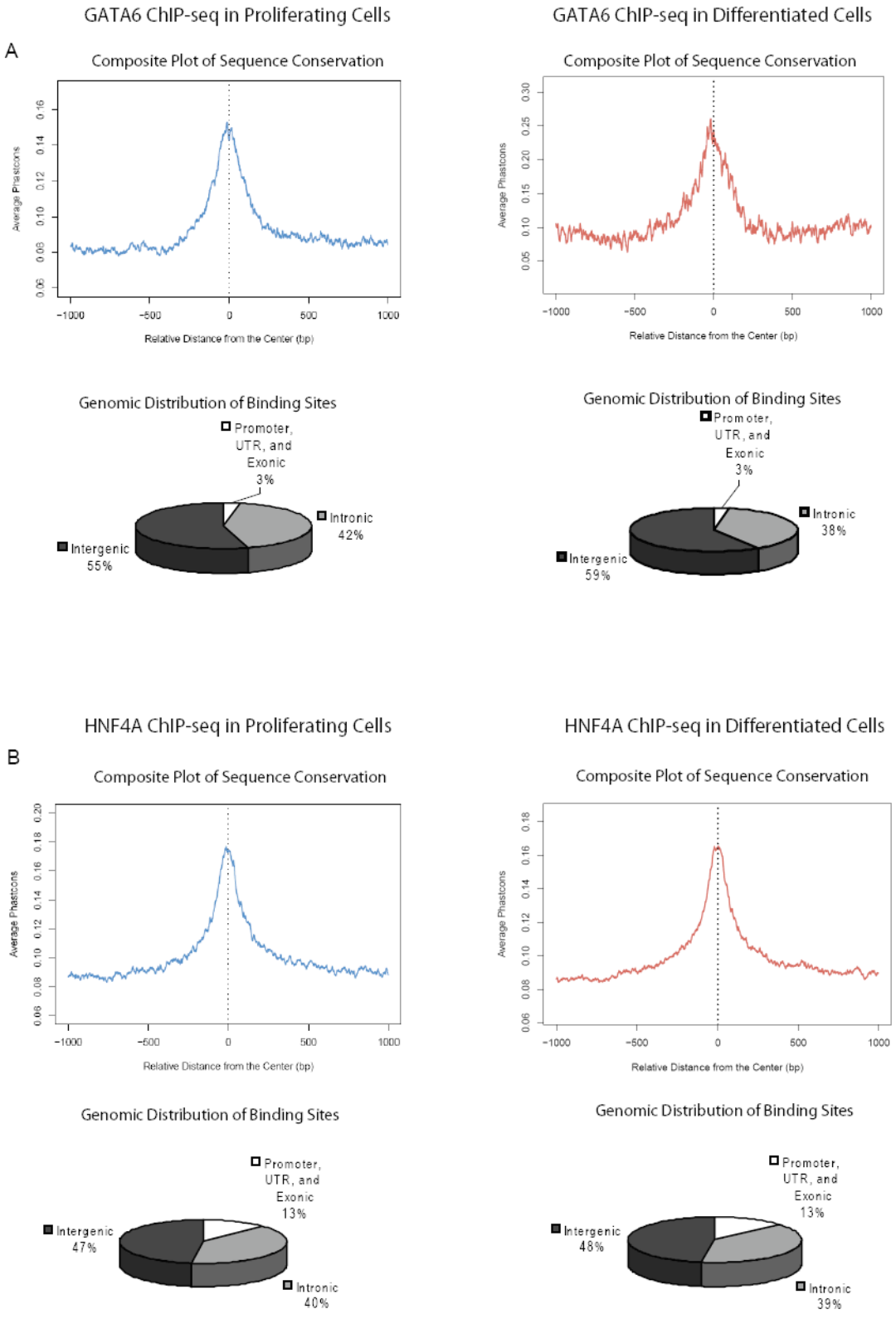


Supplemental Figure 4. CDX2 ChIP-seq regions are conserved, enriched for the consensus motif, and located predominantly at putative distal regulatory regions. **A**, Composite plot of sequence conservation across all ChIP regions reveals increased evolutionary conservation at the centers, implying function as regulatory elements. Average Phastcons scores were generated for placental mammals from the UCSC Phastcons Track. **B**, The distribution of ChIP sites indicates that CDX2 principally binds far from classical promoters, i.e., >5 kb from TSSs. **C**, Following ChIP-seq, *de novo* motif discovery using SeqPos tools identified the most centered, enriched motif to be highly similar to one that *in vitro* binding data previously attributed to CDX2.



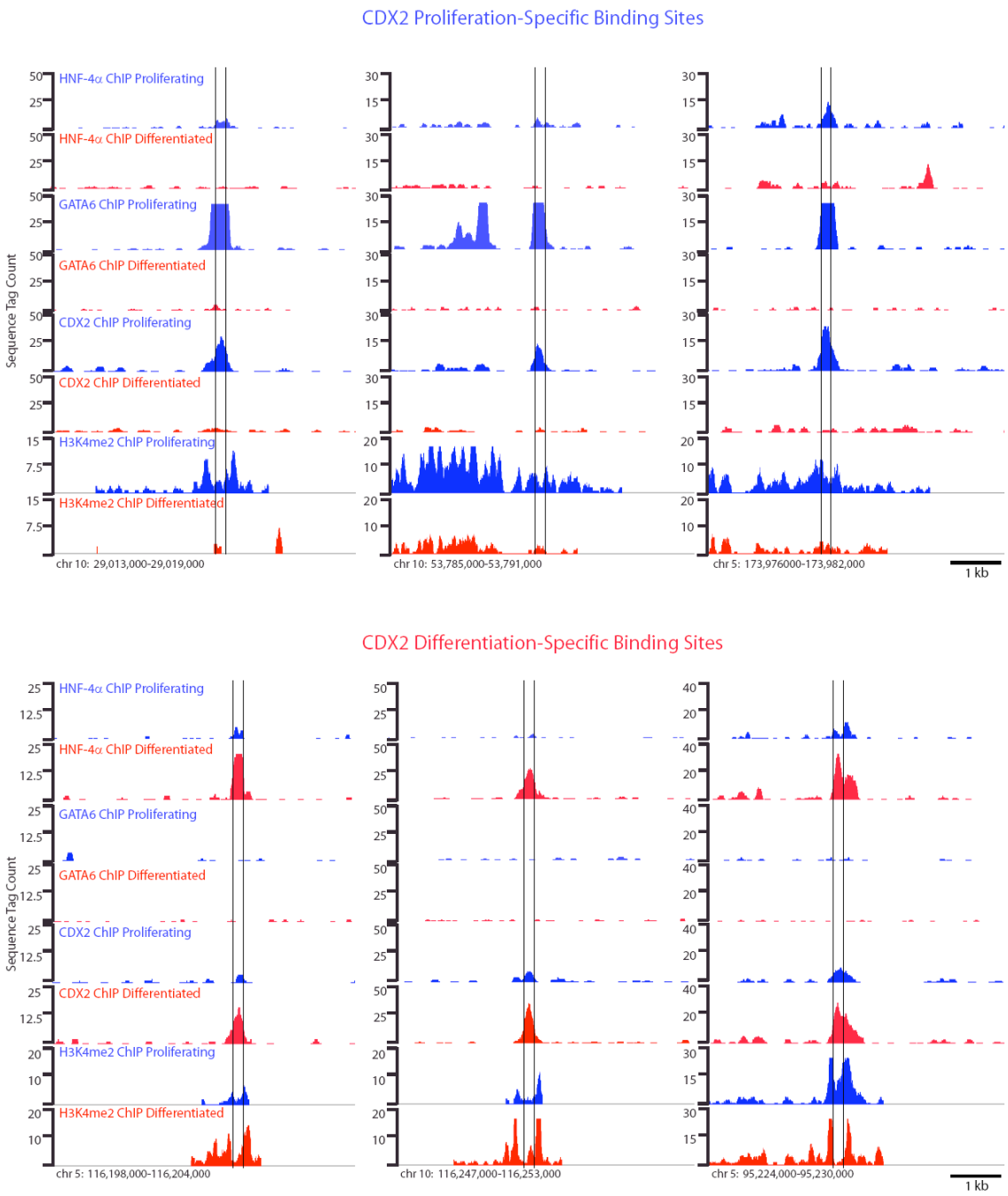
Supplemental Figure 5. Generation of inducible, intestine-specific *Cdx2* knockout mice. Our conditional *Cdx2*^{fV} allele carries *LoxP* elements flanking exon 2, which encodes the DNA-binding domain, and Frt sites flanking the Neomycin-resistance (*Neo*^{R}) cassette (**B**). Mice were

crossed first with Flp-deletor mice(Farley et al., 2000) to excise Neo^R (**C**); the resulting progeny were crossed with $Villin\text{-}Cre^{ER(T2)}$ transgenic mice and bred to $Cdx2^{fV}$ homozygosity. $Cdx2^{fV/fV};Villin\text{-}Cre^{ER(T2)}$ mice were treated with 1 mg intraperitoneal tamoxifen daily for 5 days to induce *Cre*-mediated excision of exon 2 (**D**). **E**, (left) Southern blot of ES cell clones, showing two properly targeted clones. The faster migrating band (7.6kb) is the result of HindIII (H3) sites. The result shown is with an external 5' probe (left); targeting was verified with an external 3' probe detecting the targeted H3 fragment (7.5kb) in addition to the wild-type band (17.7kb) (right). **F**, PCR genotyping confirms excision of the Neo^R cassette. As depicted by half-arrows in **A** and **C**, primers flanking Neo^R were designed to amplify either the wild-type sequence (arrowhead) or the targeted modification resulting from Neo^R excision (arrow). The product expected from targeted regions that retain Neo^R is too large to be amplified under these conditions. **G-H**, Cdx2 immunostaining detects robust expression in control mouse ileal epithelium (brown nuclear signal) but none after tamoxifen treatment of $Cdx2^{fV/fV};Villin\text{-}Cre^{ER(T2)}$ mice, indicating efficient depletion of Cdx2.



Supplemental Figure 6. GATA6 and HNF4A ChIP-seq regions are conserved and located mainly at putative distal regulatory regions. Composite plots of sequence conservation across

all ChIP regions for GATA6 (**A**) and HNF4A (**B**) reveal increased evolutionary conservation at the centers, implying regulatory functions. Average Phastcons scores were generated for placental mammals from the UCSC Phastcons Track. The distribution of ChIP sites indicates that binding of both GATA6 (**A**) andHNF4A (**B**) predominates at distances >5 kb from TSSs.



Supplemental Figure 7. ChIP-seq tag densities at individual regions demonstrate condition-specific CDX2 occupancy and differential association with GATA6 or HNF4A binding and H3K4Me2-marked nucleosomes. Three regions selectively bound by CDX2 in proliferating (Top) and differentiated cells (Bottom) are shown. In each case, 6 kb regions with the indicated genome coordinates are shown, with ChIP-seq tag counts on the y-axis. Eight ChIP-seq tracings are shown as labeled: H3K4Me2, CDX2, GATA6, and HNF4A, each in proliferating (blue) and differentiated (red) states. GATA6 almost exclusively occupies proliferation-specific sites in the

proliferating state, whereas HNF4A binds predominantly at differentiation-specific sites in differentiated cells. Proliferation-specific sites lie near genes expressed selectively in proliferating cells (*BAMBI*, *DKK1*, *MSX2*); differentiation-specific sites lie near genes that increase expression as cells differentiate (*SEMA6A*, *TMEM23*, *ABLIM1*).

Differentiation-Specific Active Chromatin Structure Motifs

Motif ID	Motif Factor	tstat
M00790	HNF1	12.61
M00132	HNF1	11.66
M01011	HNF1	11.57
M01031	HNF4	10.44
M00206	HNF1	9.93
M00411	HNF4alpha1	8.49
M00134	HNF4	8.47
M00638	HNF4alpha	8.05
M00729	Cdx-2	7.59
M00489	Nkx6-2	7.47
M00149	SBF-1	7.35
M00510	Lhx3	7.3
M00266	Croc	7.23
M00292	Freac-4	6.81
M00763	PPAR direct repeat 1	6.67
M00158	COUP-TF, HNF-4	6.58
M00268	XFD-2	6.35
M00762	PPAR, HNF-4, COUP, RAR	6.17
M01152	RXR (PXR:RXR)	6.09
M00724	HNF3alpha	6.02
M00294	HFH8 (FOXF1A)	5.91
M00099	S8	5.86
M00710	Zen	5.8
M00138	OCT_1	5.78
M01094	Abd-B	5.73
M00455	dri	5.67
M01084	Antp	5.66
M00094	BR-C Z4	5.64
M00764	HNF4 direct repeat 1	5.64
M00672	TEF	5.63
M00290	Freac-2	5.57
M01012	HNF3	5.39
M00820	HAHB-4	5.31
M00424	Nkx6-1	5.18
M00713	TBP	5.13
M00639	HNF6	5.08
M00991	Cdx	5.06

Proliferation-specific Active Chromatin Structure Motifs

Motif ID	Motif Factor	tstat
M00348	GATA-2	-12.21
M00027	AbaA	-11.71
M00495	Bach1	-10.93
M00349	GATA-2	-10.77
M00490	Bach2	-9.96
M00347	GATA-1	-9.86
M00038	GCN4	-9.39
M00517	AP-1	-9.04
M00199	AP-1	-8.94
M00926	AP-1	-8.78
M00203	GATA-X	-8.7
M00925	AP-1	-8.51
M00350	GATA-3	-7.94
M00128	GATA-1	-7.77
M00174	AP-1	-7.71
M00924	AP-1	-7.69
M00088	Ik-3	-7.13
M00346	GATA-1	-6.89
M00082	Evi-1	-6.83
M00278	Lmo2 complex	-6.7
M00079	Evi-1	-6.54
M00080	Evi-1	-6.5
M00652	Nrf-1	-6.44
M00695	ETF	-6.33
M00204	GCN4	-6.25
M00053	c-Rel	-6.14
M00037	NF-E2	-6.09
M00459	STAT5B (homodimer)	-5.76
M00052	NF-kappaB (p65)	-5.64
M00916	CREB	-5.58
M01057	ERF2	-5.56
M00188	AP-1	-5.39
M00208	NF-kappaB	-5.29
M00981	CREB, ATF	-5.28
M00917	CREB	-5.22
M00036	v-Jun	-5.18
M00513	ATF3	-5.07
M00196	Sp1	-5.07
M00362	dTCF	-5.05
M00457	STAT5A (homodimer)	-5.03

Supplemental Table 1. Complete results of motif discovery at the centers of H3K4Me2 regions differentially active in proliferating and differentiated Caco-2 cells. NSD analysis(He et al., 2010) followed by motif detection identified motifs enriched among differentiation-specific and proliferation-specific active chromatin regions.

Supplemental Table 2. Complete results of Gene Ontology term analysis for Biological Processes, the source for data presented in Figures 2 and 3. See Excel File.

Supplemental Table 3. The 100 transcripts most significantly reduced upon Cdx2 depletion *in vivo* include many terminally differentiated intestinal genes. Shown are the transcripts most significantly reduced in 3 replicate mRNA profiles, with fold-changes indicated on a $\log_2$ scale. See Excel File.

EXTENDED EXPERIMENTAL PROCEDURES

Identification of mononucleosomes. Mononucleosomes carrying H3K4Me2 or H3K27Ac marks were identified using Nucleosome Positioning from Sequencing (NPS) with default parameters(Zhang et al., 2008b). NPS extends every sequence tag toward its 3' end by the nucleosome length of 145 nucleotides and takes the middle 75 nucleotides to obtain the optimal nucleosome position signal. After deriving this signal from the raw ChIP-Seq tags, NPS removes genome background by wavelet denoising and identifies the genomic locations of positioned nucleosomes using Laplacian of Gaussian (LoG) edge detection. Using a p-value cutoff of 10^{-4}, NPS detected about 200,000 mono-nucleosomes with each type of modified histone in proliferating and mature cells (Figure S2).

Chromatin Immunoprecipitation (ChIP) and co-immunoprecipitation for transcription factors, ChIP-seq, and data analysis. Caco-2 cells were cultured in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum (FBS) and 2mM L-glutamine. Proliferating cells were passaged every 2 days to maintain subconfluency; to induce maturation, cells were grown to confluence and the medium replaced every other day for an additional 26 days(Halbleib et al., 2007; Saaf et al., 2007). Mouse villi were isolated as described(Guo et al., 2009) and passed through a 70uM filter (BD Falcon 352350); the material retained by the filter was harvested. For each immunoprecipitate (IP), ~$15E^6$ cells were cross-linked with 1% formaldehyde for 10 min at 37^0C, washed in cold phosphate-buffered saline (PBS), resuspended in lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.1, and complete protease inhibitors (Roche)), and sonicated to obtain chromatin fragments between 200 bp and 1200 bp. Sonicated chromatin was resuspended in IP buffer (1% triton, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl, pH 8.1) and incubated overnight at 4^oC with magnetic beads (Dynal) conjugated to one of the following antibodies: CDX2 (Bethyl BL3194), GATA6 c-20 (Santa Cruz 7244), GATA6 h-92 (Santa Cruz 9055), HNF4α c-19 (Santa Cruz 6547) or HNF4α h-171 (Santa Cruz 8987). The IP was washed 6 times with RIPA buffer (50 mM HEPES, pH 7.6, 1 mM EDTA, 0.7% Na deoxycholate, 1% NP-40, 0.5 M LiCl) and the DNA recovered by reversing the cross-links in 1% SDS, 0.1 M $NaHCO_3$ for 8 h at 65^0C. DNA was purified and quantified by Picogreen (Manufacturer). 700 pg of DNA was used per qPCR reaction (Applied Biosystems).

For co-immunoprecipitation, nuclear extracts were prepared from sub-confluent and post-confluent Caco-2 by lysing cells first with a Cell Lysis Buffer (20 mM Tris pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl) followed by Nuclear Lysis Buffer (20mM Tris pH 7.9, 1.5 mM $MgCl_2$, 0.42M NaCl) in the presence of protease and phosphatase inhibitors. Lysates were pre-cleared with protein A/G agarose (Santa Cruz sc-2003) and precipitated with CDX2 Ab (BD Pharmigen 560171) for 4h. Immune complexes were collected with protein A/G agarose for 2 h at 4°C, washed 3 times in PBS, and eluted with SDS sample buffer. Control precipitations with equal amounts of mouse IgG (Santa Cruz, sc-2025) were carried on parallel lysates. Samples were resolved by SDS-PAGE, transferred to nitrocellulose membranes, and probed with CDX2 (Santa Cruz, sc-134468), HNF4A (Santa Cruz sc-6556) and GATA6 (Santa Cruz sc-7244) Ab followed by horseradish peroxidase-coupled secondary Ab and detection by chemiluminescence. Lysates were immunoblotted with the same Ab.

DNA motif searches within transcription factor binding sites. 600-bp sequences surrounding the summit of every ChIP region (*i.e.*, the local maximum of ChIP signals in the binding site) were retrieved and their subsequences searched exhaustively for over-representation of known motifs compared to the genome background. Both motif signal strength and the regional bias of location toward the center of binding sites were considered in assessing enrichment scores based on z-statistics (Ref. (He et al., 2010) and Meyer CA, unpublished). In addition to known motifs in the TRANSFAC database, we attempted to detect *de novo* motifs using the MDscan motif-finding module(Liu et al., 2004) within the Cistrome Analysis Pipeline (http://cistrome.dfci.harvard.edu/ap/).

Expression analyses. To detect dysregulated gene expression resulting from Cdx2 loss, we isolated RNA from the jejunal epithelium of three 4-8-week old control and 3 $Cdx2^{f/f}$;$Villin\text{-}Cre^{ERT2}$ mice 3 days or 4 days after completion of a 5-day course of tamoxifen treatment and followed the manufacturer's instructions to process for hybridization to Mouse Genome 430A 2.0 microarrays (Affymetrix). Array data were analyzed directly with dChIP(Li and Wong, 2001), or pre-processed and normalized using RMA and quantile normalization in R/Bioconductor(Bolstad et al., 2003; Irizarry et al., 2003). Limma was applied to the pre-processed gene expression data(Smyth, 2004) to identify genes that are differentially expressed with respect to controls, along with $\log_2$ fold-changes and p-values.

For immunostaining, tissues were fixed overnight in 4% paraformaldehyde at 4°C before embedding in paraffin and preparing 5-μm thick sections. Slides were first treated in a pressure cooker for 2 min at 15 psi in 10 mM sodium citrate, pH 6.0. Endogenous peroxidase activity was blocked with 3% H_2O_2 followed by blocking in 5% FBS for 1 h at 25°C. Antibodies were applied overnight at 4°C. Biotin-conjugated secondary IgG (Jackson ImmunoResearch) and avidin-biotin complex solution (Vector Laboratories) were each applied for 1 h at 25°C. Slides were developed in 0.05% 3,3'-diaminobenzidine and 0.1% H_2O_2, and counterstained with hematoxylin. Cell lineages were counted as a fraction of positive cells in crypts (Ki67) or villi (PAS and ChromograninA).

Gene association and GO analyses. We used Fisher's exact test to examine how genes that are differentially expressed in proliferating and mature Caco-2 cells are related with condition-specific CDX2 binding. Given a set of differentially expressed genes (*e.g.,* increased in mature Caco-2 cells relative to progenitors) and a certain cut-off for distance to the nearest binding site (*e.g.,* 100 kb), the Fisher's exact test estimates the significance of potential effect of CDX2 binding on gene expression changes based on the hypergeometric distribution. Sets of genes with CDX2 binding sites that are unique to proliferating or differentiated Caco-2 cells and whose transcripts are differentially expressed between the two cell states (GEO accession number GSE1614) were examined for enrichment of Gene Ontology (GO) terms using MetaCore (GeneGo, St. Joseph, MI). Significant GO terms for gene expression data from *Cdx2*$^{fV/fV}$*;Villin-Cre*$^{ER(T2)}$ mice were similarly obtained using DAVID Gene Functional Analysis tools (http://david.abcc.ncifcrf.gov/), where differentially expressed genes were detected using limma (R/Bioconductor) with a 5% false-discovery rate cut-off.

Oligonucleotides used to determine relative transcript levels

Mouse

Cndp1-RTb	CCAGTCCCAAGAACTGTGGT
Cndp1-RTb	GTGTCGGGGAACAAATGAAG
Abca1-RT	TTATCATTGGCCCTCATTCC
Abca1-RTb	CAGCAGGACTGTCACAGCTTTA
Dpp4-Rt	TCAAGTGCGGAATAGCTGTG
Dpp4-RTb	TATCATCTGCCGTTCCATGA
Aspa-RT	CCTCACGGTGTCCTTAGAGC
Aspa-RTb	GGTGCAATGGTTTCCAGTCT
Lipe-RT	CTCCACATGCCCCTCTACAC
Lipe-RTr	GCGGCAGATCTTCTACCACT
mCdx1RTf	GTTTACTTTGCGCTCCTTGG
mCdx1RTr	CCGAACCAAGGACAAGTACC
mCDX2-RTf	TCACCATCAGGAGGAAAAGTG
mCDX2-RTr	GCAAGGAGGTCACAGGACTC
mPCNA rt pcr	CAGTGGAGTGGCTTTTGTGA
mPCNA rt pcr	CCACATTGGAGATGCTGTTG
mGapdh F	ACCACAGTCCATGCCATCAC
mGapdh R	TCCACCACCCTGTTGCTGTA
mActin F	TGTTACCAACTGGGACGACA
mActin R	CCATCACAATGCCTGTGGTA

Human

CDH17f	CTGACCACCCTTCTGGTGAT
CDH17r	GGATGAGATTGTTGTTGCTG
CDX2f	GCGGAACCTGTGCGAGTG
CDX2r	GACTGTAGTGAAACTCCTTCTC
GATA6f	CACAGGCAGGTTGGTTTACA
GATA6r	TACAAAACGGCTCCAAAAGG
HNF4Af	GAGCTGCAGAGTCGATGACAA
HNF4Ar	TACTGGCGGTGTCGTTGATGTA

Oligonucleotides used in MNase protection assays

mSema6aL1	TGACAGCTCTGTGAACAAAACA
mSema6aL2	CCCATTGGACCTGTTCTCAT
mSema6aC1	TGAGAAGGTAGCCTGGCATT
mSema6aC2	TGATTCGACTCGCAGGTATG
mSema6aR1	TGCCTGACATTTTGATTTTCA
mSema6aR2	CCACCATCCTTTGCCATAAT
mISX-L1	TTGTGTGAGCAAACTGTCCA
mISX-L2	TCACTTGGGTCAAGTAAAGAAATC
mISX-C1	TGAGGAAAGCACCAGGATCT
mISX-C2	CGAGAGCAAAGTGGAGAGTG
mISX-R1	TGAGCCTCTGTAGATCACTGC
mISX-R2	GATGGGGAAACAGACGAAGA

mSis-L1	TGATGAAACAGGTTGTGAGTCC
mSis-L2	TTCACACAGACAATGATCAAACA
mSis-C1	AGCATCCGGAAGTCATCAAC
mSis-C2	ATTGGAGGAAGGCCAAAGAC
mSis-R1	AGCATCCTTCCCATGTTCTG
mSis-R2	GGGAAATCAAAGTGGGAACTC
mSynpo-L1	GGACAGAGACGAGGCTGGTA
mSynpo-L2	CTCAGCCATCAGGCGATAG
mSynpo-C1	GGACGTGCTTAGAGGATAGGG
mSynpo-C2	CATTCACGTGCCTCTCTTCC
mSynpo-R1	GGTTCATGGTGAGCTTGGAG
mSynpo-R2	TGAGGAAGAATTAGGCAGATGG
mNr3c1-L1	CGGATGGCAGCTTTTTACAT
mNr3c1-L2	CGGATTCATTCGTTCCTCAT
mNr3c1-C1	CTGCGTGCAGAGCAGCTA
mNr3c1-C2	AGGAATTACTCGGGCAGCTT
mNr3c1-R1	CCCATGGCTAACTCCTGAAA
mNr3c1-R2	GACCAGGAGAGAAGGGAAGC

REFERENCES FOR SUPPLEMENTAL MATERIALS

Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185-193.

Farley, F. W., Soriano, P., Steffen, L. S., and Dymecki, S. M. (2000). Widespread recombinase expression using FLPeR (flipper) mice. Genesis *28*, 106-110.

Guo, J., Longshore, S., Nair, R., and Warner, B. W. (2009). Retinoblastoma protein (pRb), but not p107 or p130, is required for maintenance of enterocyte quiescence and differentiation in small intestine. J Biol Chem *284*, 134-140.

Halbleib, J. M., Saaf, A. M., Brown, P. O., and Nelson, W. J. (2007). Transcriptional modulation of genes encoding structural characteristics of differentiating enterocytes during development of a polarized epithelium in vitro. Mol Biol Cell *18*, 4261-4278.

He, H. H., Meyer, C. A., Shin, H., Bailey, S., Wei, G., Wang, Q., Zhany, Y., Xu, K., Ni, M., Lupien, M*., et al.* (2010). Positioned nucleosomes flanking a labile nucleosome characterize transcriptional enhancers. Nat Genet *42*, 343-347.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostat *4*, 249-264.

Li, C., and Wong, W. H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci USA *98*, 31-36.

Liu, Y., Wei, L., Batzoglou, S., Brutlag, D. L., Liu, J. S., and Liu, X. S. (2004). A suite of web-based programs to search for transcriptional regulatory motifs. Nucleic Acids Res *32*, W204-207.

Saaf, A. M., Halbleib, J. M., Chen, X., Yuen, S. T., Leung, S. Y., Nelson, W. J., and Brown, P. O. (2007). Parallels between global transcriptional programs of polarizing Caco-2 intestinal epithelial cells in vitro and gene expression programs in normal colon and colon cancer. Mol Biol Cell *18*, 4245-4260.

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology *3*, Article3.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008a). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zhang, Y., Shin, H., Song, J. S., Lei, Y., and Liu, X. S. (2008b). Identifying positioned nucleosomes with epigenetic marks in human from ChIP-Seq. BMC Genomics *9*, 537.