TCF4 and CDX2, major transcription factors for intestinal function, converge on the same *cis*-regulatory regions

Michael P. Verzi[a,b], Pantelis Hatzis[c], Rita Sulahian[a,b], Juliet Philips[a], Jurian Schuijers[c], Hyunjin Shin[d], Ellen Freed[a], John P. Lynch[e], Duyen T. Dang[f], Myles Brown[a,b], Hans Clevers[c], X. Shirley Liu[d], and Ramesh A. Shivdasani[a,b,1]

[a]Department of Medical Oncology and [d]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02115; [b]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115; [c]Netherlands Institute of Developmental Biology and Hubrecht Institute, 3508 AD, Utrecht, The Netherlands; [e]Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA 19104; and [f]Department of Medicine, University of Michigan School of Medicine, Ann Arbor, MI 48109

Edited by Eric N. Olson, University of Texas Southwestern, Dallas, TX, and approved July 14, 2010 (received for review March 23, 2010)

Surprisingly few pathways signal between cells, raising questions about mechanisms for tissue-specific responses. In particular, Wnt ligands signal in many mammalian tissues, including the intestinal epithelium, where constitutive signaling causes cancer. Genome-wide analysis of DNA *cis*-regulatory regions bound by the intestine-restricted transcription factor CDX2 in colonic cells uncovered highly significant overrepresentation of sequences that bind TCF4, a transcriptional effector of intestinal Wnt signaling. Chromatin immunoprecipitation confirmed TCF4 occupancy at most such sites and co-occupancy of CDX2 and TCF4 across short distances. A region spanning the single nucleotide polymorphism rs6983267, which lies within a MYC enhancer and confers colorectal cancer risk in humans, represented one of many co-occupied sites. Co-occupancy correlated with intestine-specific gene expression and CDX2 loss reduced TCF4 binding. These results implicate CDX2 in directing TCF4 binding in intestinal cells. Co-occupancy of regulatory regions by signal-effector and tissue-restricted transcription factors may represent a general mechanism for ubiquitous signaling pathways to achieve tissue-specific outcomes.

context specificity of signaling | genome-wide chromatin immunoprecipitation | signaling | Wnt signaling in intestine | tissue-specific gene regulation

Diverse cellular interactions in animals are served by remarkably few signaling pathways: transforming growth factor-β, Wnt, and fibroblast growth-factor ligands, for example, signal in most tissues. Although different combinations of ligands and receptors may enable some of the diversity, ligands are often interchangeable and their receptors redundant with one another. Furthermore, signals feed into common intracellular pathways that impinge on homologous transcription factors with similar or identical DNA sequence preference. The wide variety of transcriptional outcomes that follow activation of the same signaling pathway in different tissues thus remains poorly explained. Our investigation of genome occupancy of the homeodomain transcription factor CDX2 led to previously unexplored insights into mechanisms for signal specificity.

CDX2 expression in adult mammals is restricted to intestinal epithelium (1). Forced CDX2 expression drives maturation of cultured gut epithelial cells (2) and intestinal differentiation in transgenic mouse stomachs (3, 4) or cultured human esophageal cells (5). Conversely, Cdx2 loss converts mouse intestinal epithelium into an esophageal type (6); CDX2 thus behaves as a dominant regulator of intestinal fate. However, its direct activity has been demonstrated only at the proximal promoters of a few intestine-expressed genes (7–10) and its modes of interaction with the intestinal cell genome are unknown. We mapped CDX2 associations with chromatin in the human colon cell line Caco-2 using ChIP-chip: chromatin immunoprecipitation (IP) followed by hybridization of DNA to tiled whole-genome arrays. Analysis of CDX2-bound *cis*-regulatory modules revealed highly significant overrepresentation of binding sites for transcriptional effectors of the Wnt signaling pathway, a prime determinant of intestinal cell proliferation and differentiation (11). Experimental and computational approaches subsequently confirmed CDX2 and TCF4 co-occupancy at hundreds of genomic sites and revealed their significance in tissue-specific gene expression. Such partnering between signaling-effector and cell-restricted transcription factors at *cis*-regulatory regions likely accounts in part for tissue-specific effects of ubiquitous signaling pathways.

Results and Discussion

Genome Binding Sites for CDX2. Caco-2 cells constitute a reliable model for intestinal epithelial function and differentiation (12, 13) and express CDX2 far in excess of its homolog CDX1 (14). Model-based analysis of tiling arrays (MAT) (15) from three ChIP-chip replicates revealed ~1,100 regions of CDX2 interaction at a false discovery rate (FDR) of zero (Fig. 1*A*), 1,400 sites at FDR < 0.01, and 2,500 sites at FDR < 0.1. In independent ChIP isolates with CDX2 antibody (Ab) compared with IgG, quantitative PCR confirmed 5- to >100-fold enrichment at all 24 sites selected from high-confidence ChIP-chip data (Fig. 1*B*); a second CDX2 Ab confirmed association with the same regions (Fig. S1*A*). The known target locus *CDH17* illustrates that ChIP-chip verified published results on CDX2 binding to the promoter (9) and uncovered an additional distant interaction (Fig. S1*C*). CDX2 binding occurred mainly in intergenic regions and introns (Fig. S1*B*), and although nearby genes were not appreciably enriched in particular protein classes (Fig. S1*D*), they showed enrichment for Gene Ontology (GO) terms for developmental and transcriptional processes (*P* values < $1e^{-7}$) (Table S2). Newly identified CDX2 binding sites generally gave stronger ChIP signals than the few previously reported sites (examples in Fig. 1*B*), which tend to lie in promoters. Aggregate analysis of CDX2-binding regions indicated sequence conservation across multiple vertebrate species, with highest average conservation at the center of the binding site (Fig. 1*C*); strong evolutionary conservation is one reliable indicator of function.

Author contributions: M.P.V. and R.A.S. designed research; M.P.V., P.H., R.S., J.P., J.S., and E.F. performed research; J.P.L. and D.T.D. contributed new reagents/analytic tools; M.P.V., J.P., H.S., M.B., H.C., X.S.L., and R.A.S. analyzed data; and M.P.V. and R.A.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: All microarray data have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE22572).

[1]To whom correspondence should be addressed. E-mail: ramesh_shivdasani@dfci.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1003822107/-/DCSupplemental.

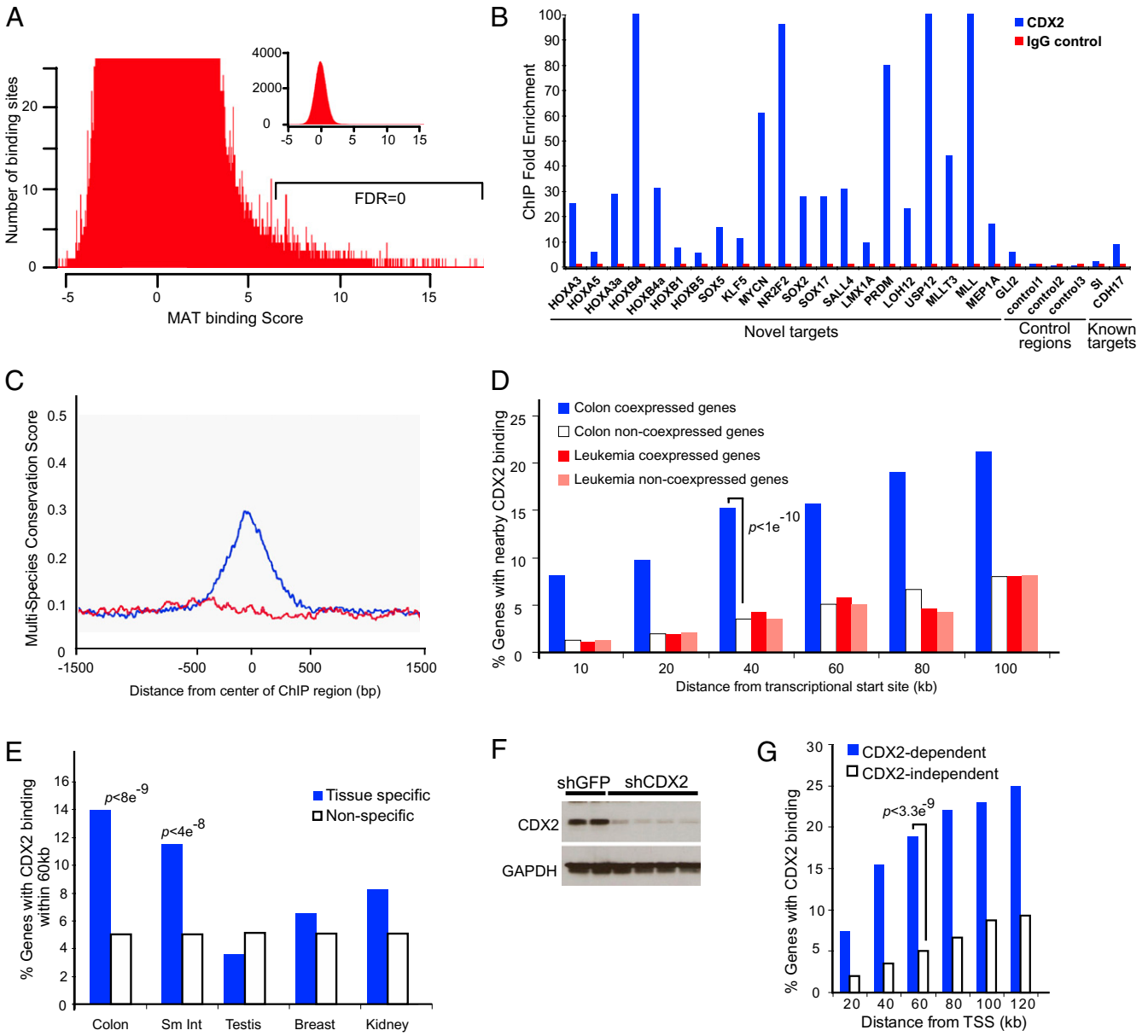

Fig. 1. CDX2 binding in the Caco-2 genome. (*A*) Histogram showing CDX2 ChIP-chip binding sites identified by MAT (15) at different MAT scores. Sites outside the normal distribution were assigned confidence scores to impute FDRs. (*B*) Representative validation of unique sites in independent CDX2 IPs, followed by quantitative PCR (qPCR) analysis of selected genomic regions. Mean fold-enrichment in ChIP is expressed relative to input or IgG IPs and control regions. Sites are named according to the nearest locus; genome coordinates are listed in Table S1. (*C*) Composite plot of multispecies conservation at CDX2-occupied regions, showing peak conservation at the centers (blue); random regions (red) show background conservation. (*D*) Oncomine was used to identify genes coexpressed with CDX2 in cancer samples (16). Genes coexpressed with CDX2 in colon tumors are more likely to have nearby CDX2 binding than genes not coexpressed with CDX2; gene coexpression in leukemia showed no correlation with intestinal CDX2 binding. (*E*) RNA expression data from 55 mouse tissues (18) were correlated with CDX2 occupancy within 60-kb windows. Human orthologs of genes highly expressed in mouse small and large intestine are much more likely to have nearby CDX2 occupancy in Caco-2 cells than genes not expressed in mouse intestine. (*F*) Immunoblot demonstration of CDX2 depletion in Caco-2 cells following lentiviral delivery of shCDX2. (*G*) Genes dysregulated upon CDX2 depletion are more likely to have nearby CDX2 binding sites than unaffected genes.

To judge the relevance of CDX2 occupancy, first we used Oncomine analysis (16) of transcripts that are coexpressed with CDX2 mRNA in multiple colon cancers, including 184 coexpressed genes determined from data from the International Genomics Consortium (GEO accession GSE2109). This group carried significantly higher probability of nearby CDX2 binding in Caco-2 cells than genes that are not coexpressed with CDX2 in colon tumors. The relationship held true for CDX2-binding regions at least as far as 100 kb from transcriptional start sites (TSSs, *P*-value range $1.7e^{-9}$–$1.1e^{-10}$) (Fig. 1*D*) and for other colon cancer expression datasets (Fig. S2*A*). CDX2 occupancy also correlated highly with genes that respond to forced CDX2 expression in cultured esophageal cells (5) (Fig. S2*C*). Although CDX2 is aberrantly expressed in human acute myeloid leukemia (17), genes coexpressed with CDX2 in leukemia samples did not show frequent nearby CDX2 binding in Caco-2 cells (Fig. 1*D* and Fig. S2*B*).

Second, we interrogated a catalog of gene expression in 55 adult and embryonic mouse tissues (18). CDX2-occupied sites appear at a higher frequency near the loci for 286 transcripts most highly expressed in the small or large intestine, compared with

inactive loci in those organs (Fig. 1*E*). These associations were not evident near genes expressed in tissues where CDX2 is absent and, in the intestine, persisted at nearest-gene distances extending >100 kb from CDX2-binding sites (Fig. S3). Third, we profiled transcripts in Caco-2 cells depleted of CDX2 by lentiviral shRNA. Total CDX2 removal caused cell death and we could not achieve more than ~70% reduction in the protein level in viable cells (Fig. 1*F*). As a result, we detected residual CDX2 occupancy at many sites, albeit with reduced signal strength. Nevertheless, the levels of 148 transcripts were significantly altered ($P < 0.05$) and the loci encoding these transcripts showed a higher frequency of nearby CDX2 occupancy compared with loci unaffected by CDX2 depletion (*P*-value range $2e^{-4}$–$6.9e^{-10}$ for CDX2 occupancy between 20 and 100 kb from the TSS) (Fig. 1*G*). Whereas luciferase reporter assays gave inconclusive results for different binding regions, probably because the assay fails to replicate native chromatin and gene regulation, the strong associations with expressed genes indicate that ChIP-chip accurately mapped CDX2-occupied regions with probable *cis*-regulatory function. Furthermore, CDX proteins in several species bind the sequence a/cATAAAa/t (19, 20) and de novo motif discovery identified a highly similar string as the most-centered motif among CDX2 ChIP-chip regions in Caco-2 cells (Fig. 2*A*).

Identification of Potential CDX2 Cofactors. To determine whether CDX2 might commonly regulate intestinal genes in conjunction with other proteins, we searched for consensus motifs for other sequence-specific proteins within the central 100 bp of CDX2-occupied regions, where conservation is highest (Fig. 1*C*). CEAS (*cis*-regulatory element annotation system) (21) and SeqPos (22), two algorithms designed to seek motifs present in the TRANSFAC database (http://www.gene-regulation.com), uncovered high enrichment for sequences corresponding to GATA, FOXA1, HNF1, TCF/LEF, and HNF4A transcription factors (Fig. 2*B* and Table S3). As expected, this analysis also identified the core CDX2 motif at high frequency ($P < 1.3e^{-93}$). In contrast, 100-bp stretches from the 3′ edges of the same CDX2-ChIP regions lacked significant enrichment of any motifs, including the short t/aGATAa/g sequence recognized by the GATA family (Fig. 2*B*). These results are striking because GATA proteins, HNF1, FOXA1, and HNF4A are among the few factors known to regulate genes in tissues derived from the embryonic endoderm. GATA proteins, for example, control expression of certain mouse intestinal markers, interact physically with CDX2, and affect CDX2-dependent reporter constructs (23–25). Elucidation of CDX2-occupied regions hence offers one explanation for recurrent observation of intestinal gene control by a small group of tissue-restricted factors. By contrast, CDX2 or other intestine-restricted proteins have not previously been associated with TCF/LEF proteins, the transcriptional effectors of canonical Wnt signaling. A distribution analysis of their respective motifs across CDX2-occupied regions revealed significant enrichment of both near the centers of binding regions, with a short distance between their peaks; as a control, estrogen-responsive motifs are evenly distributed across CDX2-occupied areas and hardly enriched over the genomic background (Fig. 2*C*). As examples, both CDX2-binding regions in the *CDH17* locus show close proximity of consensus CDX2 and TCF/LEF motifs (Fig. S1*C*). Together, these results suggested that CDX2 and TCF proteins may commonly interact with the same *cis*-regulatory regions in colonic cells.

Confirmation of CDX2 and TCF4 Co-Occupancy. The principal effector of Wnt/β-catenin signaling in the intestine is TCF4 (26). To determine whether TCF4 binds DNA near CDX2 sites, as the motif distributions implied, we first compared CDX2 occupancy in Caco-2 with 6,900 TCF4-binding regions (FDR = 0) identified in LS174t colon cancer cells (27). Of 1,100 CDX2-binding regions in Caco-2, 340 (FDR = 0) overlapped with the TCF4-occupied sites, whereas

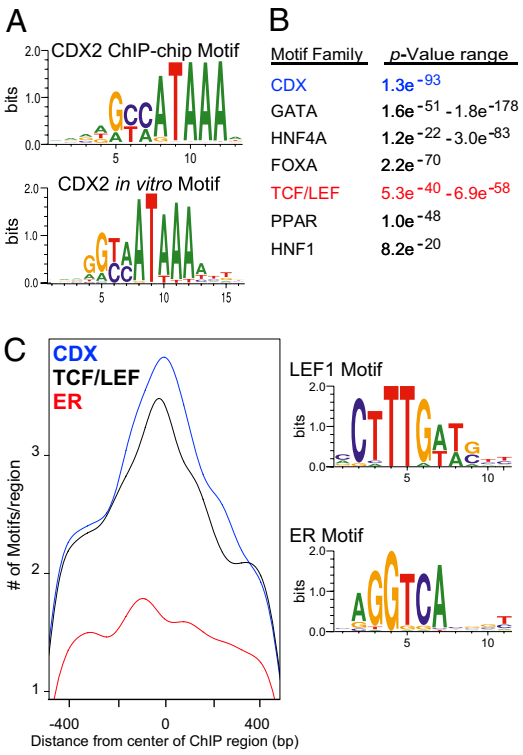

Fig. 2. Computational prediction that intestinal transcription factors co-occupy CDX2 binding regions. (*A*) Position-weighted matrices demonstrating that the most centered sequence motif identified de novo ($P < 1e^{-30}$, *Upper*) within CDX2-occupied regions closely resembles the consensus motif attributed to CDX2 (*Lower*) (19). (*B*) TRANSFAC motifs identified using CEAS (21), with the corresponding *P* value for enrichment in CDX2-bound regions over the genomic background. Each of these motifs is associated with transcription factors known to regulate intestinal genes. (*C*) Density plots of motif occurrences at CDX2 ChIP regions shows disproportionate enrichment of CDX2 and TCF/LEF motifs at the centers of CDX2-occupied regions, suggesting functionality; the estrogen-response element (ERE) is not enriched or centered. Motif matrices used in the analysis are shown on the right.

only 11 regions were found in common with 6,900 random intergenic regions (Fig. 3*A*). Although TCF4-occupied regions comprise about 0.2% of the genome and CDX2-interacting regions about 0.075%, nearly one in every three CDX2 regions also binds TCF4, with >91% of the overlapping regions sharing >50% common sequence; the probability of this occurring by chance is almost 0 ($P < 4.94e^{-324}$, one-side binomial test).

To confirm regional co-occupancy in the same cells, we conducted ChIP-chip analysis for TCF4 in Caco-2 cells and detected 476 binding sites ($P < e^{-5}$) on tiled microarrays covering all non-repetitive sequences from chromosomes 8, 11, and 12 (Fig. 3*B*). On these three chromosomes, 169 of the 318 CDX2-bound regions (53.1%) also showed TCF4 occupancy ($P < 4.94e^{-324}$, one-side binomial test). Similar to CDX2 sites, TCF4-bound regions were mostly located far from TSSs (Fig. S4*A*), evolutionarily conserved (Fig. S4*B*), and enriched in consensus motifs for other intestinal factors (Fig. S4*C*). We also performed independent ChIP-chip analysis in LS174t cells and detected similar concordance in CDX2 and TCF4 occupancy, confirming binding of the two factors close together ($P < 1.76e^{-227}$, one-side binomial test) (Fig. 3*B*). This extraordinary and unexpected degree of overlap in ChIP signals suggests that CDX2 and TCF4 may act on many of the same *cis*-regulatory regions.

Of particular interest for human disease, a SNP located on chromosome 8q24 confers increased risk for colorectal cancer, and three groups place this SNP within a TCF4-bound enhancer that

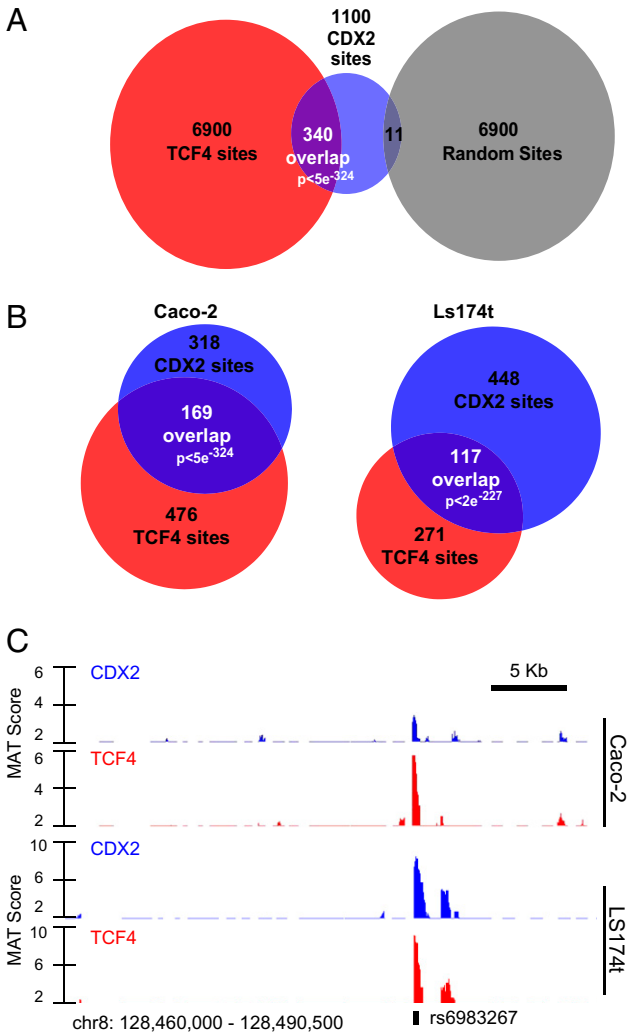

Fig. 3. The *cis*-regulatory region co-occupancy by CDX2 and the Wnt-effector transcription factor TCF4. (*A*) Extensive overlap between 6,900 TCF4-binding regions previously identified (FDR = 0) in Ls174t colon cancer cells (27) and CDX2-occupied regions in Caco-2. Nearly 1 in 3 CDX2-binding sites (FDR = 0) falls within a TCF4-occupied region, compared with <1% that overlap with 6,900 random 1-kb regions. (*B*) TCF4 ChIP-chip on chromosomes 8, 11, and 12 in Caco-2 (*Left*) and Ls174t (*Right*) cells revealed TCF4 occupancy of 26% (Ls174T) to 53% (Caco-2) of CDX2-binding regions on these chromosomes. (*C*) Raw traces of CDX2 (blue) and TCF4 (red) ChIP-chip data in Caco-2 (*Upper*) and LS174t (*Lower*) cells, demonstrating precise co-occupancy at an upstream *MYC* enhancer that encompasses the colon cancer risk SNP rs6983267.

controls MYC transcription (28–30). In both Caco-2 and LS174t cells, binding of CDX2 and TCF4 coincided precisely over the implicated region (Fig. 3*C*), with the next nearest occupancy in either direction occurring >350 kb away. To determine if the two factors occupy target loci simultaneously, we precipitated LS174t chromatin first with TCF4 Ab and, after extensive washing, again with CDX2 Ab; in reciprocal experiments, we precipitated first with CDX2 Ab and then with TCF4 Ab. In both cases, sequential precipitation retained significantly more target DNA than mouse or goat IgG controls, which did not enrich for target DNA in either primary or sequential precipitation (Fig. 4*A* and Fig. S4 *D* and *E*). These results indicate that both factors occupy the same chromatin and provide further evidence for co-occupancy of regulatory regions in intestinal epithelial cells.

Functional Significance of CDX2 and TCF4 Co-Occupancy. We next asked if CDX2 binding is required for TCF4 to bind nearby. To overcome incomplete CDX2 depletion by shRNAs (Fig. 1*F*), we used the colon cell line LoVo as an alternative model. Endogenous CDX2 expression in LoVo cells has been ablated by targeted recombination (31), providing an opportunity to examine effects in paired cell lines with and without CDX2 expression (Fig. 4*B*). ChIP assays in parental and derivative LoVo lines confirmed CDX2 and TCF4 co-occupancy at several representative sites we had originally identified as co-occupied in Caco-2 cells (Fig. 4*C*). Absence of CDX2 abrogated CDX2 occupancy, as expected, and also disrupted TCF4 binding at several but not all such sites. By contrast, absence of CDX2 did not affect TCF4 binding at representative sites where ChIP-chip had revealed TCF4 but not CDX2 occupancy (Fig. 4*C*). These results corroborate *cis*-regulatory region co-occupancy and suggest that CDX2 is required for TCF4 binding at certain sites.

Canonical Wnt signaling is central to intestinal epithelial homeostasis (11) but also functions in other organs; it is unclear how the pathway elicits distinct transcriptional responses in different tissues or developmental stages. In light of CDX2 and TCF4 co-occupancy at many sites in gut epithelial cells, intestine-restricted CDX2 expression might account in part for a tissue-specific Wnt response. To test this idea, we extended the analysis that had revealed strong association between CDX2 occupancy and intestine-specific gene expression (Fig. 1*E*). Compared with CDX2-binding sites, those occupied by TCF4 but not CDX2 in Caco-2 cells showed measurable but lesser association with intestine-enriched genes (Fig. 4*D*). Regions that bound both CDX2 and TCF4 showed the most robust association with intestine-specific transcripts ($P < 5.5e^{-7}$, logistic regression model), much greater than TCF4 alone ($P < 0.077$). Significant association between cobinding and tissue-specific gene expression was confined to the small and large intestine; gene expression in three tissues lacking CDX2 showed no correlation. Co-occupancy is thus a significant predictor of intestinal gene expression.

Conclusions. Inactivating *APC* or activating *CTNNB1* mutations are found in nearly all colon cancers (32, 33) and the role of constitutive Wnt signaling in intestinal tumorigenesis is well established (11). In contrast, the role of CDX2 in colon cancer is unclear. Heterozygosity at the *Cdx2* locus reduces small intestine adenomas almost 10-fold in $APC^{+/\Delta 716}$ mice, suggesting cooperation between Cdx2 and Wnt signaling; however, colonic polyp numbers are increased in the compound mutant strain (34). *CDX2* is usually expressed but rarely mutated in the human disease, and manipulation of CDX2 levels in cancer cell lines increases or decreases malignant behavior (35), without a clear pattern. Although many CDX2 properties likely occur independent of TCF4, our results are compatible with such experimental variability, as the combined activity of the two proteins at given *cis*-regulatory regions may be cooperative or antagonistic.

In summary, this study reveals that CDX2 binds many of the same *cis*-regulatory regions as TCF4 in intestinal epithelial cells, that CDX2 is required for TCF4 binding in some regions, and that TCF4 and CDX2 co-occupancy predicts intestinal gene expression better than occupancy by either factor alone. Caudal and TCF/LEF proteins may also cooperate during development; combined deficiency of TCF4 and TCF1 in mice causes severe caudal truncation and anteriorizes the digestive tract (36), defects similar to those reported in *Cdx2* mutants (37). We propose that *cis*-regulatory region co-occupancy by a Wnt effector protein and a tissue-restricted transcription factor represents a means for tissue-specific response to Wnt signaling in the intestine, as shown in this study, and probably also in other tissues.

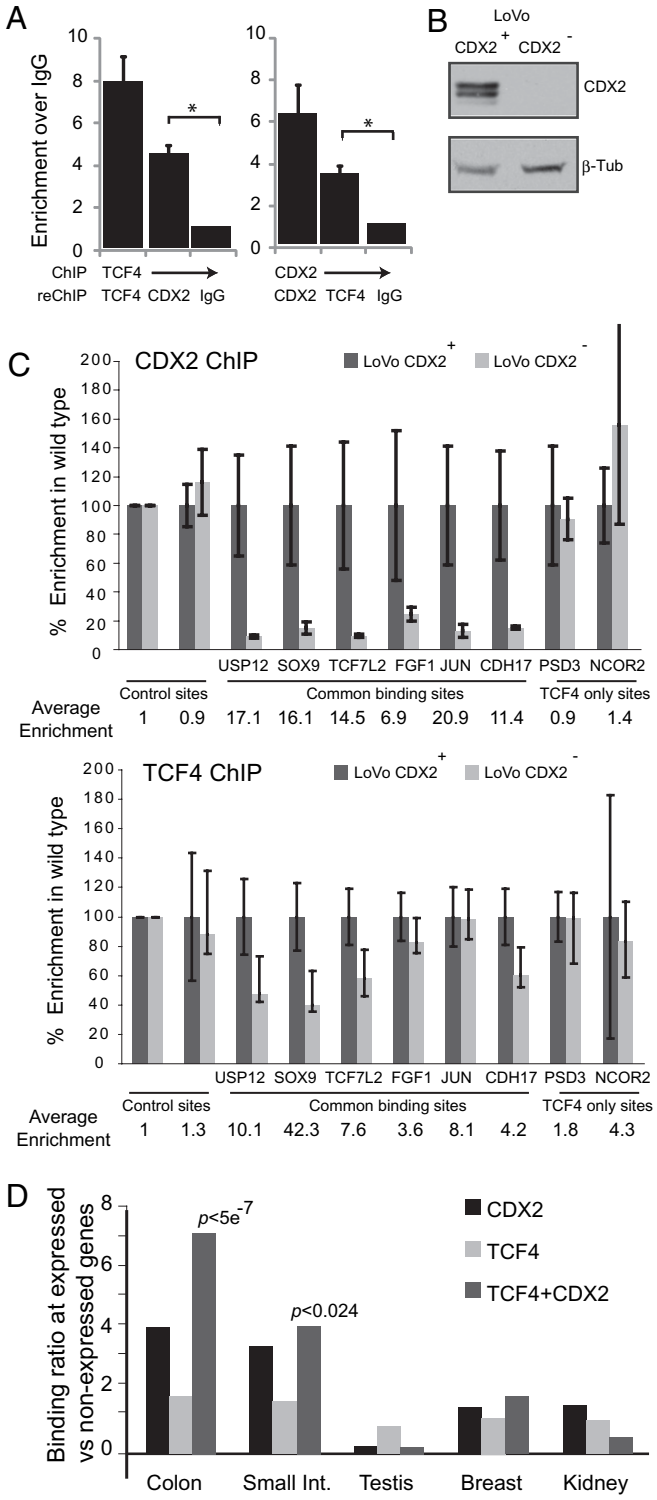

Fig. 4. Demonstration and functional significance of *cis*-regulatory region co-occupancy by CDX2 and TCF4. (*A*) Sequential ChIP with antibody against TCF4 followed by antibody against CDX2 and vice versa reveals the presence of both transcription factors on the same chromatin binding region near *EPHB2*. The chromatin eluted from the first ChIP was immunoprecipitated again (reChIP) with CDX2, TCF4, or mouse or goat IgG controls. Enrichment is calculated based upon qPCR relative to the IgG control, which is set to 1. Enrichments are indicated with SEM and are the results of three independent experiments. The asterisk represents $P < 0.001$ (t test). Similar results at a region near *ADRA2C* are shown in Fig. S4*D*. (*B*) Immunoblot evidence of CDX2 expression in parental and modified LoVo cells. (*C*) ChIP analysis of representative individual and co-occupied sites in paired CDX2$^+$ and CDX2$^-$ LoVo lines, with fold-enrichment relative to input DNA determined by qPCR. ChIP enrichment is shown relative to the parental, CDX2-expressing LoVo line. The average ChIP enrichment in the parental line for each region is listed below the graph. CDX2 binding showed dependence on CDX2. TCF4 ChIP revealed that binding at certain co-occupied sites also depends on the presence of CDX2, whereas binding at sites without demonstrable CDX2 occupancy does not. Sites are named according to the nearest locus and genome coordinates are listed in Table S1. (*D*) Comparison of the ratios of tissue-specific versus nonspecific loci bound by each transcription factor. TCF4- and independent CDX2-binding sites show significant but relatively modest correlation with transcripts expressed in the mouse intestine, whereas co-occupancy of CDX2 and TCF4 is much better associated with intestinal gene expression. Breast, kidney, and testis transcripts showed no association with TCF4 or CDX2 binding near expressed genes.

Materials and Methods

Chromatin Immunoprecipitation. Caco-2 and LS174t cells were cultured in DMEM and LoVo cells in F-12K medium, each supplemented with 10% FBS. Next, 15×10^6 cells were cross-linked with 1% formaldehyde for 10 min at 37 °C, washed in cold PBS, resuspended in lysis buffer [50 mM Tris-HCl, pH 8.1, 10 mM EDTA, 1% SDS, complete protease inhibitors (Roche)], and sonicated to obtain chromatin fragment lengths between 200 and 1,200 bp. This chromatin was diluted in IP buffer (20 mM Tris-HCl, pH 8.1, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100) and incubated overnight at 4 °C with magnetic beads (Dynal) conjugated to either CDX2 (mix of Bethyl BL3194 and Biogenex CDX2-88), TCF4 (Santa Cruz Biotechnology, sc-8613), or control rabbit IgG (sc-2027) antibodies. Immunoprecipitates were washed six times with RIPA buffer (50 mM Hepes, pH 7.6, 0.5 M LiCl, 1 mM EDTA, 0.7% Na deoxycholate, 1% Nonidet P-40) and DNA recovered over 8 h by incubation in 1% SDS, 0.1 M $NaHCO_3$ at 65 °C. DNA was purified and quantified by Picogreen (Invitrogen). A total of 700 pg of DNA was used per qPCR reaction (Applied Biosystems) or 6 to 10 ng were used for ChIP-chip.

For ChIP-chip, DNA immunoprecipitated with specific Ab or IgG control was prepared for array hybridization, as described previously (22). Briefly, DNA was treated with Proteinase K and RNaseA, ends were repaired and blunted with End-It (Epicenter), and adapter oligos were ligated onto the library for ligation-mediated PCR. The amplified library was fragmented with DNase, labeled according to the manufacturer's instructions, and hybridized to GeneChip Human Tiling 2.0R array sets (Affymetrix), or Human tiling array 2.0R F only. Each experiment consisted of at least three independent ChIP assays pooled. Binding sites were detected by MAT, as described previously (15). Data are publicly available at GEO (accession no. GSE22572).

Small Hairpin RNA Depletion, Immunoblotting, and RNA Preparation. Lentiviral delivery of Mission TRCN13684 shRNA (Sigma) was used to deplete CDX2 in Caco-2 cells and compared with cells infected with shGFP control virus (SH005, Sigma). To generate virus, 95% confluent 10-cm plates of 293 cells were cultured in DMEM containing 10% FBS and transfected with a mixture of the hairpin-containing plasmid pLKO.1-puro and the packaging vectors pMISSIONgagpol and pMISSIONvsvg (all from Sigma), using Lipo293 (Signagen). Virus-containing supernatants were collected in 4 mL of DMEM supplemented with 30% FBS, filtered, supplanted with 8 μg/mL polybrene, and used to infect 30% confluent 6-cm plates of Caco-2 cells. After 2 d of infection (6 h each day), cells were selected for hairpin expression in the presence of Puromycin (2 μg/mL) and cultured in 1 μg/mL Puromycin until harvest for RNA analysis and immunoblotting with CDX2 (Bethyl 3194) or GAPDH (Abcam ab9485-200) Ab. For transcriptional profiling of Caco-2 cells treated with shCDX2 or shGFP, RNA was extracted using TRIzol (Invitrogen), labeled, and hybridized to Affymetrix Human Genome U133A 2.0 Arrays. Data were deposited for public accession at GEO (accession no. GSE22572) and analyzed using dChip (38).

Computational and Statistical Analyses. To identify overrepresented sequence motifs de novo, we applied the Motif Decoder component of the SeqPos algorithm. SeqPos weights the occurrence of binding motifs according to their locations within a binding region and gives higher scores to motifs near the centers, based on the assumption that insignificant DNA motifs will be uniformly distributed whereas functional motifs will lie closer to the center (22). SeqPos and CEAS (21) were also used to identify enrichment of TRANSFAC motifs, and CEAS to score the distribution of binding sites and multispecies conservation plots using PhastCons conservation scores from University of California Santa Cruz Genome Bioinformatics, based on alignment of human, chimpanzee, mouse, rat, dog, chicken, fugu, and zebrafish genomes. Density plots of motif enrichment across ChIP-chip

GENETICS

regions and association of trends in gene expression with transcription factor binding sites were determined as reported previously (22). Position-weighted motif logos were generated by enoLOGOS (39). For association between gene expression and CDX2 occupancy, we assigned each binding site to the gene with the nearest annotated transcriptional start site within the indicated distance.

To calculate the significance of whole-genome binding-site overlaps, the null probability of CDX2 binding to the genome was estimated as the ratio of the binding sites with respect to the whole genome. The fraction of TCF binding sites residing in CDX2-binding regions and the significance of this overlap was estimated using the one-side binomial test. Fisher's exact test was used to calculate the significance of association between gene expression datasets and nearby CDX2 or TCF4 binding sites. Student's *t* test was used to calculate the significance of binding determined in ChIP-reChIP assays between the specific Ab and no-Ab controls. To assess the effect of cobinding of CDX2 and TCF4 on tissue-specific gene expression, we applied a logistic regression model to each group of binding sites and their associated genes. The log-odds ratio between expressed and unexpressed genes in each tissue was considered as the response variable; cobinding of CDX2 and TCF4 and solitary binding of TCF4 were the explanatory variables. The *P* values of the explanatory variables were estimated using their Pearson residuals.

ACKNOWLEDGMENTS. We thank Clifford Meyer for computational support and sharing algorithms. This work was supported in part by the Caring for Carcinoid Foundation, National Institutes of Health Grant RC2CA14822 (to R.A.S.), and the Dana-Farber/Harvard Cancer Center SPORE Program in Gastrointestinal Cancers (P50CA127003). M.P.V. was supported by National Institutes of Health Training Grant T32DK07477 and a fellowship from the Crohn's and Colitis Foundation of America. P.H. was supported by a Human Frontier Science Program Long-Term Fellowship.

1. James R, Kazenwadel J (1991) Homeobox gene expression in the intestinal epithelium of adult mice. *J Biol Chem* 266:3246–3251.
2. Suh E, Traber PG (1996) An intestine-specific homeobox gene regulates proliferation and differentiation. *Mol Cell Biol* 16:619–625.
3. Mutoh H, et al. (2002) Conversion of gastric mucosa to intestinal metaplasia in Cdx2-expressing transgenic mice. *Biochem Biophys Res Commun* 294:470–479.
4. Silberg DG, et al. (2002) Cdx2 ectopic expression induces gastric intestinal metaplasia in transgenic mice. *Gastroenterology* 122:689–696.
5. Liu T, et al. (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.
6. Gao N, White P, Kaestner KH (2009) Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. *Dev Cell* 16:588–599.
7. Dang DT, Mahatan CS, Dang LH, Agboola IA, Yang VW (2001) Expression of the gut-enriched Krüppel-like factor (Krüppel-like factor 4) gene in the human colon cancer cell line RKO is dependent on CDX2. *Oncogene* 20:4884–4890.
8. Fang R, Santiago NA, Olds LC, Sibley E (2000) The homeodomain protein Cdx2 regulates lactase gene promoter activity during enterocyte differentiation. *Gastroenterology* 118:115–127.
9. Hinoi T, et al. (2002) CDX2 regulates liver intestine-cadherin expression in normal and malignant colon epithelium and intestinal metaplasia. *Gastroenterology* 123: 1565–1577.
10. Park J, Schulz S, Waldman SA (2000) Intestine-specific activity of the human guanylyl cyclase C promoter is regulated by Cdx2. *Gastroenterology* 119:89–96.
11. Clevers H (2006) Wnt/beta-catenin signaling in development and disease. *Cell* 127: 469–480.
12. Grasset E, Bernabeu J, Pinto M (1985) Epithelial properties of human colonic carcinoma cell line Caco-2: Effect of secretagogues. *Am J Physiol* 248:C410–C418.
13. Halbleib JM, Sääf AM, Brown PO, Nelson WJ (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.
14. Lu X, et al. (2008) Differential regulation of CDX1 and CDX2 gene expression by deficiency in methyl group donors. *Biochimie* 90:697–704.
15. Johnson WE, et al. (2006) Model-based analysis of tiling-arrays for ChIP-chip. *Proc Natl Acad Sci USA* 103:12457–12462.
16. Rhodes DR, et al. (2007) Oncomine 3.0: Genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* 9:166–180.
17. Scholl C, et al. (2007) The homeobox gene CDX2 is aberrantly expressed in most cases of acute myeloid leukemia and promotes leukemogenesis. *J Clin Invest* 117: 1037–1048.
18. Zhang W, et al. (2004) The functional landscape of mouse gene expression. *J Biol* 3:20.
19. Berger MF, et al. (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* 133:1266–1276.
20. Noyes MB, et al. (2008) A systematic characterization of factors that regulate *Drosophila* segmentation via a bacterial one-hybrid system. *Nucleic Acids Res* 36: 2547–2560.
21. Ji X, Li W, Song J, Wei L, Liu XS (2006) CEAS: *cis*-regulatory element annotation system. *Nucleic Acids Res* 34 (Web Server issue):W551–W554.
22. Lupien M, et al. (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* 132:958–970.
23. Bosse T, et al. (2006) Gata4 is essential for the maintenance of jejunal-ileal identities in the adult mouse small intestine. *Mol Cell Biol* 26:9060–9070.
24. Boudreau F, et al. (2002) Hepatocyte nuclear factor-1 alpha, GATA-4, and caudal related homeodomain protein Cdx2 interact functionally to modulate intestinal gene transcription. Implication for the developmental regulation of the sucrase-isomaltase gene. *J Biol Chem* 277:31909–31917.
25. Jonckheere N, et al. (2007) The human mucin MUC4 is transcriptionally regulated by caudal-related homeobox, hepatocyte nuclear factors, forkhead box A, and GATA endodermal transcription factors in epithelial cancer cells. *J Biol Chem* 282: 22638–22650.
26. Korinek V, et al. (1998) Two members of the Tcf family implicated in Wnt/beta-catenin signaling during embryogenesis in the mouse. *Mol Cell Biol* 18:1248–1256.
27. Hatzis P, et al. (2008) Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells. *Mol Cell Biol* 28:2732–2744.
28. Pomerantz MM, et al. (2009) The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet* 41:882–884.
29. Tuupanen S, et al. (2009) The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet* 41: 885–890.
30. Wright JB, Brown SJ, Cole MD (2010) Upregulation of c-MYC in *cis* through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. *Mol Cell Biol* 30:1411–1420.
31. Dang LH, et al. (2006) CDX2 has tumorigenic potential in the human colon cancer cell lines LOVO and SW48. *Oncogene* 25:2264–2272.
32. Korinek V, et al. (1997) Constitutive transcriptional activation by a β-catenin-Tcf complex in APC-/- colon carcinoma. *Science* 275:1784–1787.
33. Sparks AB, Morin PJ, Vogelstein B, Kinzler KW (1998) Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. *Cancer Res* 58:1130–1134.
34. Aoki K, Tamai Y, Horiike S, Oshima M, Taketo MM (2003) Colonic polyposis caused by mTOR-mediated chromosomal instability in Apc+/Delta716 Cdx2+/– compound mutant mice. *Nat Genet* 35:323–330.
35. Guo RJ, Suh ER, Lynch JP (2004) The role of Cdx proteins in intestinal development and cancer. *Cancer Biol Ther* 3:593–601.
36. Gregorieff A, Grosschedl R, Clevers H (2004) Hindgut defects and transformation of the gastro-intestinal tract in Tcf4(–/–)/Tcf1(–/–) embryos. *EMBO J* 23:1825–1833.
37. Chawengsaksophak K, de Graaff W, Rossant J, Deschamps J, Beck F (2004) Cdx2 is essential for axial elongation in mouse development. *Proc Natl Acad Sci USA* 101: 7641–7645.
38. Li C, Wong WH (2001) Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proc Natl Acad Sci USA* 98:31–36.
39. Workman CT, et al. (2005) enoLOGOS: A versatile web tool for energy normalized sequence logos. *Nucleic Acids Res* 33 (Web Server issue):W389–W392.