

CDX2 is an amplified lineage-survival oncogene in colorectal cancer

Keyan Salari^{a,b}, Mary E. Spulak^a, Justin Cuff^a, Andrew D. Forster^a, Craig P. Giacomini^{a,c}, Stephanie Huang^a, Melissa E. Ko^c, Albert Y. Lin^d, Matt van de Rijn^a, and Jonathan R. Pollack^{a,c,1}

Departments of ^aPathology and ^bGenetics, and ^cCancer Biology Program, Stanford University School of Medicine, Stanford, CA 94305; and ^dDepartment of Medicine, Division of Oncology, Santa Clara Valley Medical Center, San Jose, CA 95128

Edited by Michael Wigler, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, and approved October 9, 2012 (received for review April 13, 2012)

The mutational activation of oncogenes drives cancer development and progression. Classic oncogenes, such as *MYC* and *RAS*, are active across many different cancer types. In contrast, “lineage-survival” oncogenes represent a distinct and emerging class typically comprising transcriptional regulators of a specific cell lineage that, when deregulated, support the proliferation and survival of cancers derived from that lineage. Here, in a large collection of colorectal cancer cell lines and tumors, we identify recurrent amplification of chromosome 13, an alteration highly restricted to colorectal-derived cancers. A minimal region of amplification on 13q12.2 pinpoints caudal type homeobox transcription factor 2 (*CDX2*), a regulator of normal intestinal lineage development and differentiation, as a target of the amplification. In contrast to its described role as a colorectal tumor suppressor, *CDX2* when amplified is required for the proliferation and survival of colorectal cancer cells. Further, transcriptional profiling, binding-site analysis, and functional studies link *CDX2* to Wnt/ β -catenin signaling, itself a key oncogenic pathway in colorectal cancer. These data characterize *CDX2* as a lineage-survival oncogene deregulated in colorectal cancer. Our findings challenge a prevailing view that *CDX2* is a tumor suppressor in colorectal cancer and uncover an additional piece in the multistep model of colorectal tumorigenesis.

lineage dependency | lineage-addiction oncogene | 13q amplification | CDK8 | integrative genomics

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and the third leading cause of cancer death in both men and women, with ~141,000 new cases and ~49,000 deaths per year in the United States (1). The classical model of colorectal tumorigenesis proposed by Vogelstein et al. (2) more than 20 y ago describes the most common mutations found in CRC. In cancers with the more common chromosomal instability (CIN), operationally defined by the presence of multiple chromosomal gains, losses, and translocations that alter expression of cancer genes (3, 4), the sequence of events from early adenoma formation to carcinoma includes loss of adenomatous polyposis coli (*APC*), leading to transcriptional activation of Wnt target genes, activating mutations in the small GTPase *KRAS*, resulting in constitutive Ras-signaling, mutations in *TP53*, and loss of the long arm of chromosome 18 (i.e., *DCC*, *SMAD4*), which is the most common cytogenetic abnormality in CRC and is associated with an unfavorable outcome (5). In contrast, tumorigenesis in CRCs with microsatellite instability (MSI), defined by defects in DNA mismatch repair, follows a pathway distinct from but analogous to that typified by CIN tumors. The initial step involves alterations in Wnt signaling (6), followed by activating mutations in *BRAF* (instead of *KRAS*) (7). Mismatch repair deficiency further disrupts cancer genes harboring coding-sequence microsatellite tracts, such as TGF- β receptor 2 (*TGFBR2*), insulin-like growth factor 2 receptor (*IGF2R*), and the apoptosis regulator *BAX*.

More recently, large-scale exome sequencing efforts have added a layer of complexity to the genomic landscape of colorectal cancers (8, 9), identifying additional mutations and highlighting the considerable genetic heterogeneity among patient tumors. Additionally, several studies have surveyed DNA copy number

alterations (CNAs) in CRC, establishing that CIN tumors harbor numerous broad gains and losses, often affecting whole chromosome arms; common alterations include gain of chromosome arms 7p (*EGFR*), 8q (*MYC*), 11p, 13q, and 20q and loss of 1p, 4q, 5q (*APC*), 8p, 17p (*TP53*), and 18q (*DCC*, *SMAD4*) (10–13). Because of the large number of genes spanned by most of these alterations, pinpointing new driver cancer genes has proven challenging. A promising strategy to hone in on such drivers is to compare genomic data across multiple tumor types, where lineage-specific alterations might point to cancer genes playing a critical role in the development of cancers arising from a specific cell lineage (14).

Here, we carried out an integrative genomic analysis of colorectal cancer, surveying a collection of 29 colorectal cancer cell lines and publicly available data on 226 primary colorectal tumors for somatic DNA CNAs. By comparison with the genomic profiles of thousands of samples from diverse tumor types, we unveiled a region of amplification highly specific to colorectal-derived tumors. We identified caudal type homeobox transcription factor 2 (*CDX2*), a master regulator of intestinal cell survival and differentiation, as a target gene of this amplification. We performed functional studies and genome-wide location and transcriptome analysis to uncover a context-dependent role for *CDX2* in Wnt/ β -catenin signaling and CRC tumorigenesis. These data show that for a subset of colorectal-derived tumors, cell survival and proliferation are dependent on the abnormal amplification and overexpression of *CDX2*. Further, they characterize *CDX2* as a lineage-survival oncogene in colorectal cancer and provide additional insight into colorectal pathobiology.

Results

Recurrent 13q12 Amplification in Colorectal Cancer Targets *CDX2*. We profiled somatic DNA CNAs in 29 colorectal cancer cell lines and reanalyzed data from 226 primary colorectal tumors (15, 16), using Genomic Identification of Significant Targets in Cancer (GISTIC) (17) to identify recurrent amplifications or deletions and peak regions likely to harbor driver cancer genes. Consistent with previous studies (10–13), we identified numerous somatic CNAs including gain of chromosome arms 1q, 7p (*EGFR*), 8q (*MYC*), 13q, and 20q and loss of 1p, 4q, 5q (*APC*), 8p, 17p (*TP53*), and 18q (*DCC*, *SMAD4*) (Fig. 1A and *SI Appendix*, Fig. S1).

Author contributions: K.S., M.E.S., and J.R.P. designed research; K.S., M.E.S., J.C., A.D.F., C.P.G., S.H., and M.E.K. performed research; A.Y.L. and M.v.d.R. contributed new reagents/analytic tools; K.S., M.E.S., J.C., and J.R.P. analyzed data; and K.S., M.E.S., and J.R.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The microarray and short-read sequencing data reported in this paper have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) and Short Read Archive (GEO) www.ncbi.nlm.nih.gov/geo (accession no. GSE30475).

¹To whom correspondence should be addressed. E-mail: pollack1@stanford.edu.

See Author Summary on page 18651 (volume 109, number 46).

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1206004109/-DCSupplemental.

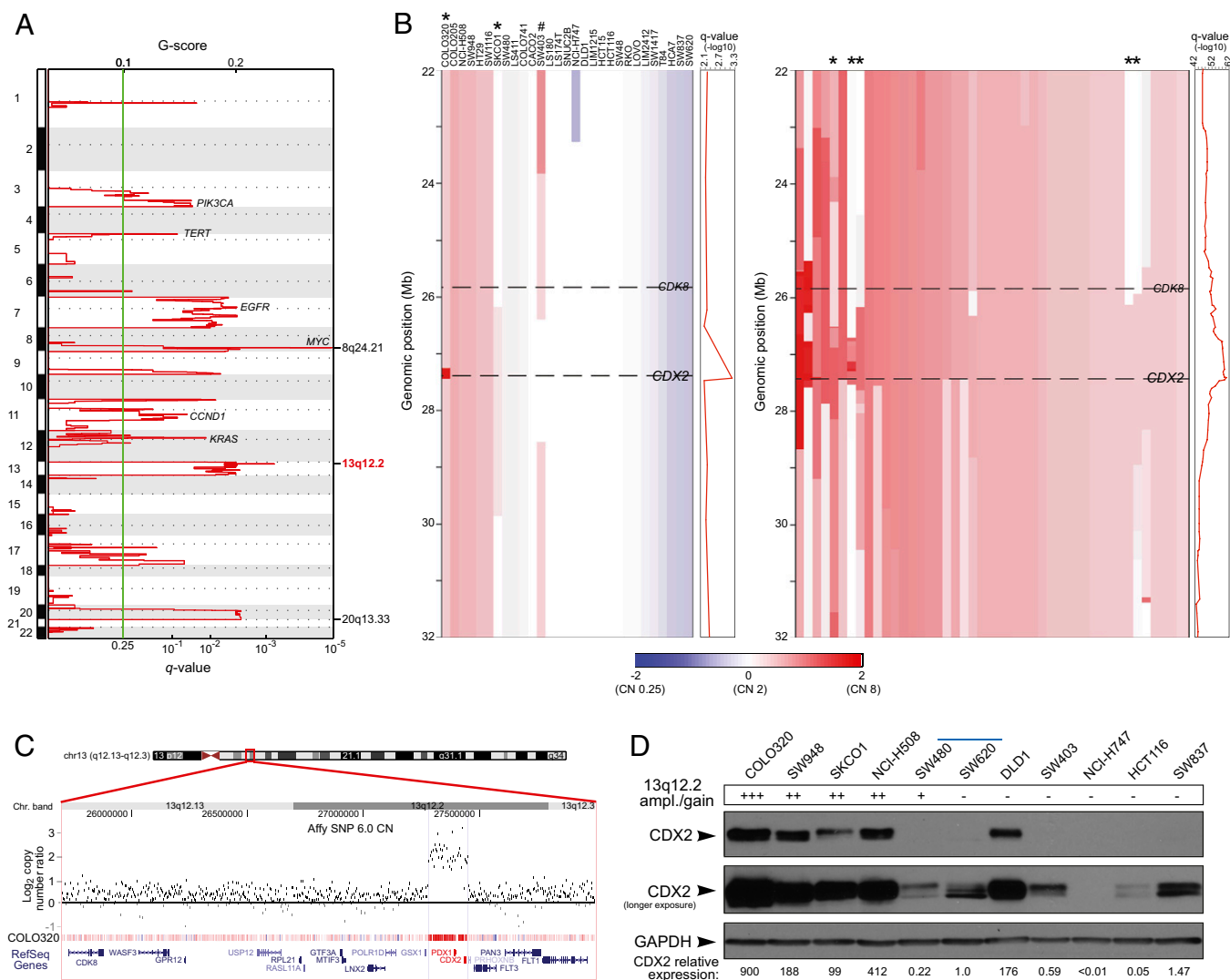


Fig. 1. Recurrent amplification and overexpression of *CDX2* in colorectal cancer cell lines and primary tumors. (A) Peaks of recurrent amplification among 29 colorectal cancer cell lines. The G-score is shown on the upper x-axis, and the q-value is shown on the lower x-axis for recurrent amplifications across the genome (y-axis) identified by GISTIC analysis of aCGH data. The positions of the most significant recurrent amplifications are labeled on the right. The green line indicates a false discovery rate of 0.25. (B) Heatmaps of the 10-Mb region of chromosome segment 13q12.2 displaying the minimal region of amplification in 29 colorectal cancer cell lines (Left) and 226 primary colorectal tumors (Right; only the amplified subset of samples is shown). Each sample is represented by a column in which areas of genomic amplification and deletion are depicted in red and blue, respectively. Alongside are graphs of the corresponding GISTIC q-value, defining the amplicon peak. The positions of *CDX2* and *CDK8* are indicated by dashed lines. Asterisks indicate samples excluding *CDK8* from the minimal region of amplification; the hash symbol (#) marks the one sample including *CDK8* that excludes *CDX2*. (C) High-resolution mapping of the 13q12.2 amplicon in COLO320 using an Affymetrix SNP 6.0 array. Log₂ copy number ratios plotted on the y-axis indicate the boundaries of the focal region of amplification spanning two RefSeq genes, *PDX1* and *CDX2*. The position of candidate colorectal oncogene *CDK8* outside the region of amplification is indicated. (D) Western blot analysis of CDX2 protein expression in 11 colorectal cancer cell lines with and without 13q12.2 amplification, including one line with single-copy loss of 13q (SW837). Shown are short and long exposures using an anti-CDX2 antibody; GAPDH serves as a loading control. Relative CDX2 expression levels are indicated, normalized to GAPDH and referenced to SW620 (set to 1). Note: SW480 and SW620 are matched primary and metastasis-derived lines from the same patient; the absence of 13q gain in SW620 and the relatively low CDX2 expression in both lines suggest that 13q gain is not a driving event for this tumor.

After amplification of 8q24 (encompassing the colorectal oncogene *MYC*), the most significant recurrent alteration was amplification on the long arm of chromosome 13, observed in 11 of 29 (38%) colorectal cancer cell lines and 121 of 226 (54%) primary tumors. To validate this finding, we performed FISH on a tissue microarray (TMA) of 76 evaluable primary colorectal tumors (using a commercially validated FISH probe mapping to 13q14) and confirmed increased copy number of 13q in 32 of 76 (42%) cases (SI Appendix, Fig. S24 and Table S1). Although most amplification events were broad (i.e., more than half the length of the chromosome arm), a number of focal amplifications helped define

three GISTIC-identified peaks on 13q (Table 1). The most significant peak maps to chromosome segment 13q12.2 (Fig. 1B), and reprofiling the sample harboring the most focal region of amplification (CRC cell line COLO320) using the high-resolution Affymetrix SNP 6.0 array revealed a 168-kb high-level amplification spanning just two genes, pancreatic-duodenal homeobox transcription factor 1 (*PDX1*) and *CDX2* (Fig. 1C); among primary tumors, the region of highest amplification includes only *CDX2*. This amplification resides adjacent to a recently described focus of copy number gain in colorectal cancers at 13q12.13 containing candidate oncogene *CDK8* (15). Although *CDX2* and *CDK8* often

Table 1. Candidate targets of chromosome 13q amplification in colorectal cancer

Cytoband	q-value*	Peak boundaries (Mb)	Genes in peak	Candidate target gene(s)
13q12.2	3.70×10^{-61}	27.41–27.47	2	<i>CDX2</i>
13q22.1	1.28×10^{-54}	72.50–72.59	1	<i>KLF5</i>
13q34	3.03×10^{-54}	108.88–110.08	4	<i>IRS2, RAB20</i>

See also [SI Appendix, Fig. S3](#) and [Table S2](#).

*GISTIC-defined significant peaks of recurrent amplification.

were coamplified, in our analysis of 255 colorectal cancer samples (including the ones used to nominate *CDK8*) (15, 16), we found six primary tumors and two cell lines that exclude *CDK8* from the peak region of 13q12.2 amplification; in contrast, only one cell line harbors an amplification excluding *CDX2* (Fig. 1B and [SI Appendix, Fig. S3](#)). This result suggests that *CDX2* is more often targeted for amplification, and thus *CDK8* may not be the primary target of 13q12.2 amplification in colorectal cancers. Additional 13q peaks of amplification include transcriptional activator *KLF5* at 13q22.1 and insulin-signaling gene *IRS2* at 13q34 (Table 1 and [SI Appendix, Fig. S1 and Table S2](#)); a colorectal cancer susceptibility variant resides in the latter (18).

PDX1 and *CDX2*, along with *GSX1* (outside the amplification), form the ParaHox gene cluster at 13q12.2. *PDX1* is a tissue-specific transcriptional activator in the early developing and adult pancreas, and defects of *PDX1* are a cause of pancreatic agenesis (19, 20). *CDX2*, a master transcriptional regulator of intestinal cell fate and survival, has been used as a marker to indicate colorectal differentiation in adenocarcinomas of unknown origin. Given its lineage-specific function, we examined a compendium of genomic profiles of 3,131 samples from 26 distinct histological types (21) and found that *CDX2* is amplified significantly only in colorectal-derived tumors ($q < 10^{-6}$) ([SI Appendix, Table S3](#)). Further, among 1,911 tumor specimens across 16 distinct tumor types (22), *CDX2* expression also is significantly associated with colorectal-derived tumors ($P < 10^{-153}$) ([SI Appendix, Fig. S4](#)).

Interestingly, *CDX2* has been characterized as a colorectal tumor suppressor in mouse models (23–25). However, *CDX2* sequence mutations are exceedingly rare—only 3 of 224 previously sequenced colorectal cancer cell lines and tumors harbor a mutation (0.9% mutant allele frequency), and all of those mutations occur in repeat sites of cancers with microsatellite instability (9, 26–29)—and although it harbors tumor-suppressor gene *RBI*, loss of chromosome 13 is rarely observed in colorectal cancers ([SI Appendix, Figs. S1 and S3](#)). Among the 29 cell lines surveyed in this study, one MSI⁺ cell line (RKO) harbors a previously identified heterozygous *CDX2* mutation; we uncovered one additional heterozygous frameshift mutation (c.1090delG) in a repeat sequence in MSI⁺ cell line LIM2412. The frameshift is predicted to replace the eight terminal amino acids of *CDX2* with one alternate amino acid (followed by a stop codon encoded by the shifted reading frame); the functional significance (if any) of this mutation is unknown. Additionally, we analyzed the CpG island spanning the *CDX2* promoter and transcription start site in 11 representative cell lines and found the majority of CpG sites in all cell lines to be unmethylated ([SI Appendix, Fig. S5](#)). These observations suggest that the infrequent *CDX2* mutations are likely secondary to genomic instability and argue against an important role for *CDX2* as a colorectal tumor-suppressor gene.

Colorectal Cancer Cells with 13q12 Amplification Exhibit *CDX2* Dependency. To determine the impact of 13q amplification on *CDX2* expression, we performed immunohistochemistry on the same colorectal cancer specimens analyzed by FISH and found nuclear *CDX2* staining in 51 of 76 (67%) cases, with a correlation between 13q copy number and *CDX2* protein expression ($\rho = 0.37$, $P = 0.001$; Spearman correlation) ([SI Appendix, Fig. S2 A](#)

and [B](#) and [Table S1](#)). We next analyzed the panel of 11 colorectal cancer cell lines and found that cell lines with 13q12.2 gain or amplification (validated by quantitative PCR ([SI Appendix, Fig. S2C](#)) generally display elevated *CDX2* transcript ($P = 0.04$, Mann–Whitney u test; [SI Appendix, Fig. S2D](#)) and protein levels ($P = 0.07$; Fig. 1D) as compared with cell lines without the amplification or as compared with normal colonic epithelium (i.e., isolated colonic mucosa uninvolved by cancer) (transcript levels $P = 0.09$, protein levels $P = 0.02$; [SI Appendix, Fig. S2 D and E](#)). A notable exception was the disomic cell line DLD1, which showed *CDX2* expression levels comparable to those in amplified cell lines. In contrast, we found no significant difference in *PDX1* transcript or protein levels in colorectal cancer cell lines with and without 13q12.2 amplification or as compared with normal colonic epithelium ([SI Appendix, Fig. S6 A and B](#)), indicating that *PDX1* is likely a passenger of 13q12.2 amplification.

Within the cohort of 76 colorectal cancer cases evaluated for 13q gain (by FISH) and *CDX2* expression (by immunohistochemistry, IHC), 13q gain was associated with advanced pathologic stage ($P = 0.039$; Fisher's exact test) but not with tumor grade or overall survival ([SI Appendix, Table S4](#)). *CDX2* expression showed a similar although nonsignificant trend ([SI Appendix, Table S4](#)). Within the collection of colorectal cancer cell lines, *CDX2* gain/overexpression did not correlate significantly with other commonly occurring colorectal cancer gene mutations, including *TP53*, *KRAS*, *BRAF*, *APC*, *CTNNB1*, *PIK3CA*, and *SMAD4* ([SI Appendix, Table S5](#)).

We next evaluated the dependence on *CDX2* expression of colorectal cancer cell lines bearing 13q12.2 amplification. RNAi by five independent shRNA and four independent siRNA in COLO320 cells revealed a profound sensitivity to reduced *CDX2* expression (Fig. 2A and [SI Appendix, Fig. S7 A and B](#)). To characterize this dependency in more detail, we suppressed *CDX2* expression in six cell lines, three with amplification of 13q12.2 (COLO320, SW948, and SKCO1) and three without (DLD1, NCI-H747, and SW837), using *CDX2*- or control GFP-targeting shRNAs (sh*CDX2*-1, sh*CDX2*-3, and shGFP) (Fig. 2B). Knockdown of *CDX2* inhibited proliferation in all three cell lines with 13q12.2 amplification but not in any cell line without the amplification (Fig. 2C). Notably, proliferation was not reduced in DLD1 cells expressing high levels of *CDX2*, suggesting an amplification-associated, context-specific dependency on *CDX2* not explainable by expression levels alone. To dissect the observed effect further, we measured the immediate effect of *CDX2* suppression on cell-cycle progression and apoptosis and found that in COLO320 cells both of two *CDX2*-directed siRNAs inhibited cell-cycle progression and induced apoptosis ([SI Appendix, Fig. S7 C and D](#)). These observations suggest that colorectal cancers with 13q12.2 amplification depend on deregulated *CDX2* expression.

We also tested tumor cell dependency on *CDK8* expression in the same panel of colorectal cancer cell lines. By *CDK8*-directed siRNA knockdown, suppression of *CDK8* expression in our hands had no significant effect on cell viability in cell lines with or without 13q12.2 amplifications spanning *CDK8* ([SI Appendix, Fig. S8](#)). Together with the observation of multiple 13q12.2 amplifications excluding *CDK8*, and only one amplification spanning *CDK8* that excludes *CDX2* (in cell line SW403), these data strongly suggest that *CDX2* (and not *CDK8*) is the important target gene of 13q12.2 amplification.

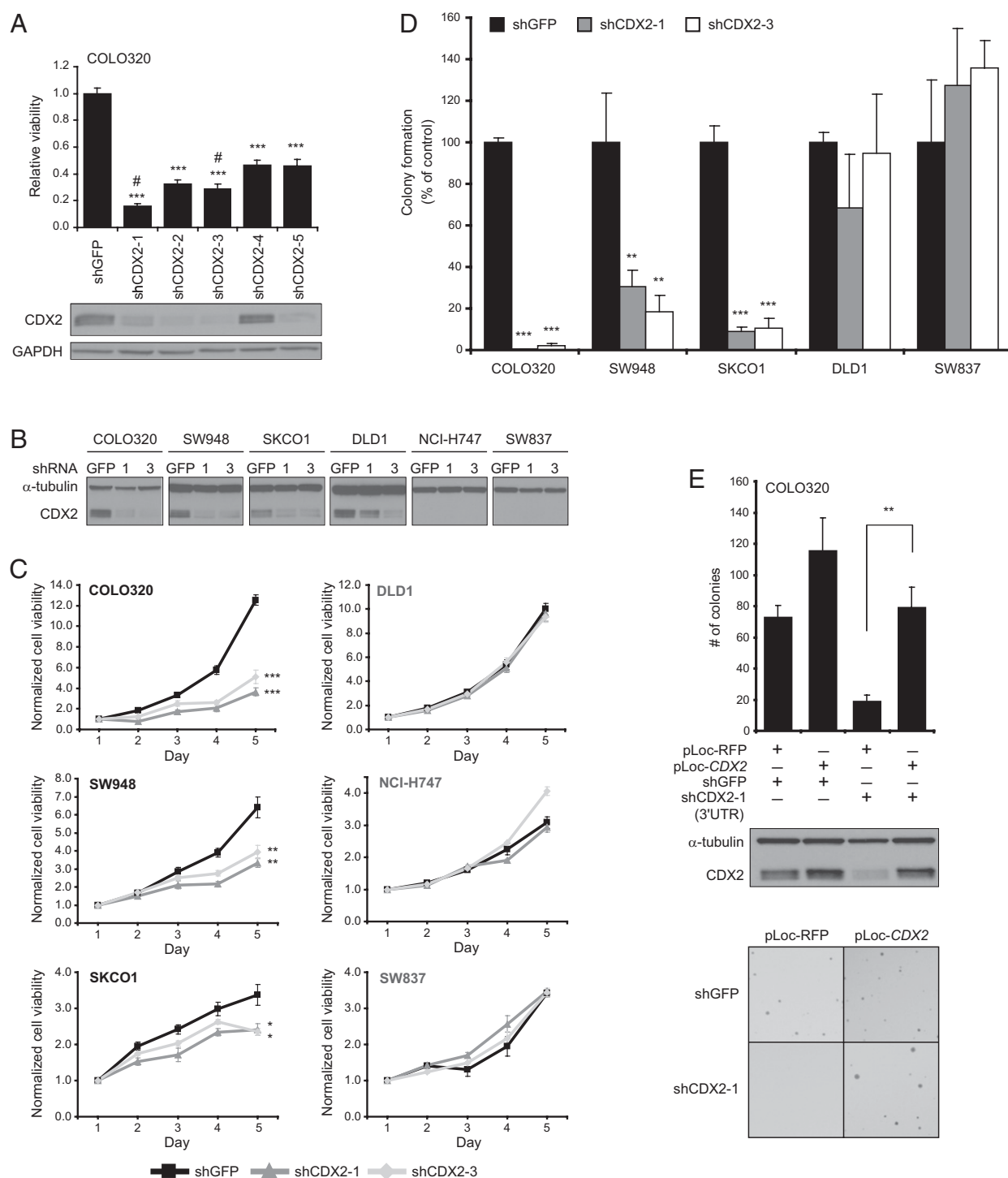


Fig. 2. RNAi of CDX2 inhibits proliferation and anchorage-independent growth of CDX2-amplified cell lines. (A) Cell viability upon stable knockdown of CDX2 expression by five independent shRNA constructs relative to shGFP in COLO320 cells. Mean cell viabilities (± 1 SD), assayed 5 d after plating cells, are reported. Hash symbols (#) indicate the two CDX2-directed shRNA constructs used in subsequent experiments. Western blot of CDX2 and GAPDH are shown. $***P < 0.001$ (Student *t* test, shCDX2 compared with shGFP). Note that shCDX2-4 is not associated with substantial CDX2 knockdown, and therefore reduced viability likely reflects an off-target effect; that shRNA is not used in subsequent studies. (B) Western blot analysis of six colorectal cancer cell lines stably expressing shRNA constructs targeting CDX2 or GFP (shCDX2-1, shCDX2-3, and shGFP). α -Tubulin serves as a loading control. COLO320, SW948, and SKCO1 bear 13q12.2 amplification; DLD1, NCI-H747, and SW837 do not. (C) Effects of CDX2 knockdown on cell viability over 5 d. Data are normalized to day 1 values to correct for differences in cell plating. Mean cell viabilities (± 1 SD) of representative experiments done twice in triplicate are reported. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. (D) Colony formation in soft agar of colorectal cancer cell lines with and without 13q12.2 gain/amplification reported as percentage relative to shGFP (± 1 SD). $**P < 0.01$, $***P < 0.001$. (E) Soft agar colony formation of COLO320 cells coinfecting to express stably shCDX2-1 or shGFP with CDX2 or control RFP vector. Western blot of CDX2 and α -tubulin and representative images of soft agar colonies are shown. Colony formation is reported as percentage relative to cells expressing shGFP and RFP vector (± 1 SD). $**P < 0.01$.

Similarly, we determined that knockdown of *PDX1* does not affect cell viability in COLO320 cells, in which *PDX1* and neighboring *CDX2* are highly coamplified (*SI Appendix, Fig. S6 C and D*).

We next examined the effect of *CDX2* amplification on anchorage-independent growth. *CDX2*-directed shRNAs markedly diminished soft agar colony formation in *CDX2*-amplified COLO320, SW948, and SKCO1 cells but had no effect in DLD1 and SW837 control cells lacking the amplification (Fig. 2D). NCI-H747 was not tested because in our hands this cell line failed to form colonies in soft agar. Because of the noted effect on proliferation, we sought to determine whether the observed reduction in colony formation upon *CDX2* suppression represented a distinct phenotype or simply reflected an attenuated capacity for cell proliferation. By comparing colony formation in soft agar with that on tissue-culture plastic, we noted that knockdown of *CDX2* represses anchorage-independent growth beyond its inhibitory effect on proliferation (*SI Appendix, Fig. S7E*). In other phenotypes assayed, suppression of *CDX2* did not affect cell migration or invasion consistently, irrespective of 13q12.2 amplification status (*SI Appendix, Fig. S9*). Finally, in contrast to the necessity of *CDX2* for anchorage-independent growth, heterologous expression of *CDX2* was not sufficient to transform immortalized rodent fibroblasts (NIH 3T3 cells) (*SI Appendix, Fig. S7H*).

To confirm that the phenotypes observed using *CDX2*-directed shRNAs were caused specifically by *CDX2* suppression, we attempted to rescue this phenotype by ectopic expression of *CDX2*. We coinfect COLO320 cells with sh*CDX2*-1, which targets the 3' UTR of *CDX2* transcripts, and a *CDX2* cDNA lacking the endogenous 3' UTR. Forced expression of *CDX2* successfully restored *CDX2* protein levels and rescued both cell proliferation and colony formation to wild-type levels (Fig. 2E and *SI Appendix, Fig. S7 F and G*). These data demonstrate that the effects of our shRNA knockdowns are not caused by off-target RNAi and are attributable specifically to *CDX2* suppression.

When Amplified, *CDX2* Promotes Wnt/ β -Catenin Signaling. We next sought to identify genes that might be under the regulatory control of *CDX2* in the context of its aberrant amplification and oncogenic phenotype. We performed gene-expression profiling of COLO320 cells after siRNA-mediated *CDX2* knockdown to identify genes that correlate with decreased *CDX2* transcript levels (*SI Appendix, Table S6*). Using gene set enrichment analysis (30), we found that the 185 genes most correlated with *CDX2* expression levels (*SI Appendix, Supplemental Materials and Methods*) were involved in Wnt/ β -catenin signaling ($n = 9$; $P = 10^{-3}$), HOX gene signaling ($n = 5$; $P = 2.6 \times 10^{-4}$), phosphatidylinositol signaling ($n = 5$; $P = 2.1 \times 10^{-3}$), and insulin signaling ($n = 5$; $P = 1.0 \times 10^{-2}$) (Fig. 3A). To determine the relevance of the cell line *CDX2*-knockdown signature to actual colorectal tumor samples, we investigated the signature genes in publicly available microarray expression datasets. Indeed, the genes correlating with *CDX2* expression levels in COLO320 showed enrichment (significant or strong trend) among the top genes correlated with *CDX2* expression in tumors (*SI Appendix, Table S7*).

Inspection of the enriched Wnt-signaling gene set (Fig. 3B) revealed genes associated with canonical Wnt/ β -catenin signaling (*FZD8*, *CSNK2A1*, *FBXW11*) as well as with the noncanonical planar cell polarity (*RAC2*) and Wnt/calcium (*PLCB3*, *CAMK2D*, *PRICKLE2*) pathways. Also included, *WNT5A* encodes a promiscuous noncanonical Wnt ligand that can either inhibit or stimulate β -catenin activity depending on the constellation of Wnt receptors (31). Notably, the top-ranked gene, *FZD8*, encodes a Wnt receptor recently found to be overexpressed in lung cancer where it promotes cell proliferation and Wnt/ β -catenin signaling (32). The second-ranked gene, *CSNK2A1*, has been found to be essential for Wnt/ β -catenin signaling (33). Thus, the microarray data provide a strong rationale for further investigating a link between *CDX2* and canonical Wnt signaling.

To complement our expression analysis and to identify direct transcriptional targets, we also performed ChIP followed by mas-

sively parallel sequencing (ChIP-seq) of *CDX2* in COLO320 cells, uncovering 14,631 high-confidence binding peaks (>30-fold enrichment over input) across the genome. As one indicator that we identified occupancy of functional elements, we observed strong sequence conservation across multiple vertebrate species peaking at the center of our *CDX2*-binding regions (Fig. 3C). Further, de novo motif scanning recovered an AT-rich motif highly similar to a *CDX2*-binding motif previously defined in vitro by protein-binding microarrays (Fig. 3D) (34). Finally, we identified a strong binding peak in the proximal promoter region of the intestine-specific sucrase-isomaltase gene, a well-characterized *CDX2* transcriptional target (Fig. 3E) (35–37).

Integrating ChIP-seq data with our *CDX2* gene-expression signature revealed that 56 of the 185 differentially expressed genes had nearby *CDX2* binding peaks ($P = 0.006$; hypergeometric test), including three of the Wnt-signaling genes (*CSNK2A1*, *WNT5A*, and *CAMK2D*) and one known *CDX2* target gene, liver-intestine cadherin 17 (*CDH17*) (Fig. 3E and *SI Appendix, Table S8*) (38). Additionally, we identified *CDX2*-binding peaks within and near Wnt/ β -catenin effector genes *LEF1* and *TCF4* (Fig. 3E). Taken together, these observations further implicate *CDX2* as a positive regulator of Wnt/ β -catenin signaling.

Buttressing this hypothesis, we analyzed a 100-bp window around our *CDX2*-binding peaks, where sequence conservation was highest, for enrichment of other transcription factor motifs in the TRANSFAC database (39); strikingly, the T-cell factor/lymphoid enhancer factor (TCF/LEF) consensus binding motif was significantly enriched ($P < 10^{-41}$; Fig. 3F). This finding suggests that *CDX2* might cooperate with TCF/LEF proteins to promote β -catenin transcriptional activity in COLO320, concordant with a recent report of *CDX2* and TCF4 co-occupancy in Caco2 colorectal cancer cells (40). As expected, the CDX consensus motif also was significantly enriched ($P < 10^{-166}$), as were motifs of other transcription factors known to interact with *CDX2* to induce intestine-specific gene expression, such as GATA and HNF1 (Fig. 3F and *SI Appendix, Table S10*) (41, 42).

To test directly whether *CDX2* promotes β -catenin transcriptional activity, we transfected β -catenin/TCF-luciferase and mutant control-luciferase reporter constructs (“TOPflash” and “FOPflash,” respectively) into 13q12.2-amplified and nonamplified cells stably expressing *CDX2*-directed or control shRNAs. Remarkably, *CDX2* suppression significantly reduced (42–66% decrease) the TOPflash/FOPflash ratio in COLO320 and SKCO1 cells (Fig. 4A), suggesting that in the context of 13q12.2 amplification *CDX2* promotes β -catenin/TCF-dependent transcriptional activity, a finding consistent with our expression array and ChIP-seq results. In contrast, *CDX2* knockdown increased the TOPflash/FOPflash ratio in DLD1 and SW837 cells (although the basal ratio also was ~40–400-fold lower than in COLO320 cells) (Fig. 4B), suggesting that in a nonamplified context *CDX2* actually inhibits β -catenin activity [concordant with recently reported findings for DLD1 (43)]. Consistent with our data linking *CDX2* amplification/overexpression to β -catenin activity, we noted a significant correlation between levels of *CDX2* and nuclear (active) β -catenin by IHC staining of TMA specimens ($\rho = 0.48$, $P < 0.001$; Spearman correlation) (Fig. 4C and *SI Appendix, Fig. S10 and Table S1*).

Discussion

We report here that *CDX2* is an amplified oncogene in colorectal cancer. When amplified and overexpressed, colorectal cancers demonstrate a marked dependency on *CDX2* levels for continued growth and survival. We further uncover a connection between *CDX2* and Wnt/ β -catenin signaling, possibly providing a unifying mechanism for the *CDX2* contribution to colorectal tumorigenesis.

Collectively, *CDX2* is gained/amplified in 52% of samples, but in the majority of cases a broad chromosomal event results in coamplification of *CDX2* and *CDK8*. Although focal amplification generally is rare among known oncogenes and thus is not to be

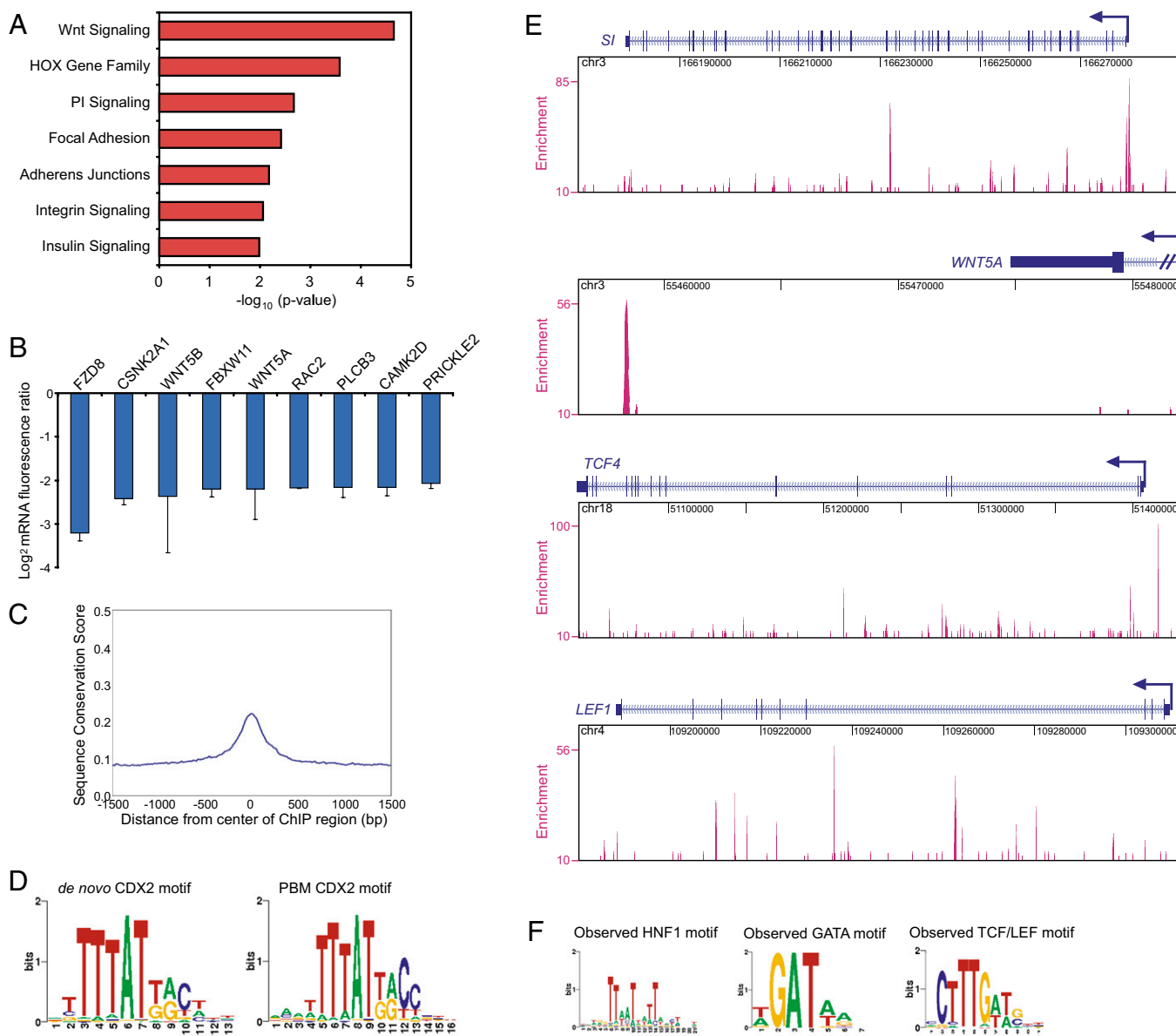
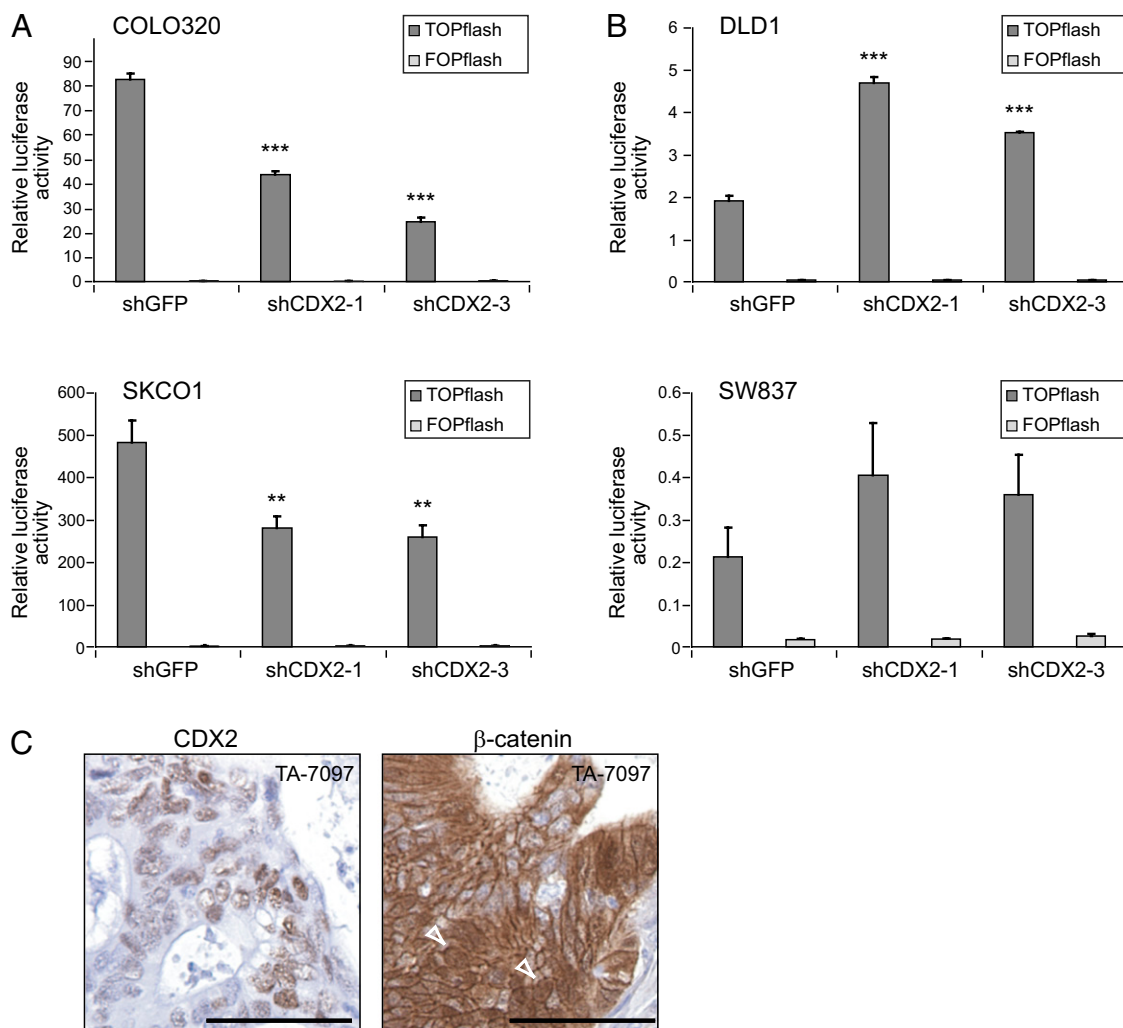


Fig. 3. CDX2 promotes expression of Wnt/ β -catenin pathway genes. (A) Pathways significantly enriched among differentially expressed genes in COLO320 cells after transfection with either of two CDX2-targeting siRNA constructs (siCDX2-1 or siCDX2-3) compared with an NTC siRNA construct. The statistical significance of enriched gene sets is indicated on the x-axis as the $-\log_{10}$ of the hypergeometric P value. (B) Gene expression of Wnt-signaling gene set members down-regulated upon transient CDX2 knockdown. Log₂ mRNA fluorescence ratios between cells transfected with CDX2-targeting or NTC siRNA constructs are shown. Data are reported as the average of microarray experiments assessing transcript abundance by the two independent siCDX2 constructs. (C) Sequence conservation of CDX2 ChIP-seq binding regions. Score indicating conservation across multiple vertebrate species is shown on the y axis for all CDX2 ChIP-seq binding regions analyzed in aggregate. Scores are reported relative to the center of CDX2 ChIP regions. (D) Sequence logos of CDX2-binding motifs. (Left) The motif identified *de novo* from CDX2 ChIP-seq binding peaks in COLO320 cells. (Right) A CDX2 motif previously defined *in vitro* by protein-binding microarrays (PBM). (E) CDX2 ChIP-seq binding peaks within and near the known intestine-specific CDX2 target gene sucrose-isomaltase and newly identified targets *WNT5A*, *TCF4*, and *LEF1*. Normalized enrichment of CDX2-immunoprecipitated chromatin over input DNA is shown on the y-axis, and genomic coordinates are labeled on upper x-axis. Gene structure is indicated above each plot. (F) Sequence logos of significantly enriched transcription factor-binding motifs within 100 bp of CDX2 ChIP-seq binding peaks in COLO320 cells.

expected (21, 44), we were able to identify six primary tumors and two cell lines in which the peak region of amplification coincides with *CDX2* and excludes *CDK8*. Conversely, only one cell line (SW403) and no primary tumor harbors an amplification that spans *CDK8* but does not include *CDX2*. This genetic evidence, illustrated by the GISTIC amplification peak residing squarely over *CDX2*, points to *CDX2* as the more likely target gene. Further, our side-by-side comparison of tumor cell dependency on *CDK8* expression in the same panel of cell lines used to explore *CDX2* demonstrates that suppression of *CDK8* expression has no effect

on cell viability in cell lines with or without amplification. In comparison with the study by Firestein et al. (15), because the cell lines used were overlapping but ultimately different (DLD1 and SW837 were common to both studies, and *CDK8* knockdown led to reduced viability in their hands but not in ours), we cannot overstate our own conclusions. We also can only speculate as to why Firestein et al. did not identify *CDX2*; this disparity may reflect the cell lines used (much of their initial work was done with non-amplified lines DLD1 and HCT116) and/or their apparent reliance on a single shRNA hairpin targeting *CDX2*. Nevertheless, we



of *AR* amplification in prostate cancer (54), *MITF* amplification in melanoma (55), and *TITF1* or *SOX2* amplification in adenocarcinoma or squamous cell carcinoma of the lung (44, 56–58), respectively. The highly lineage-restricted pattern of *CDX2* amplification and expression we found across thousands of samples from diverse tumor types supports this hypothesis and suggests that *CDX2* invokes lineage-dependent mechanisms to orchestrate an oncogenic phenotype. These data characterize *CDX2* as a lineage-survival oncogene in colorectal cancer.

Despite its amplification and dependency for cell proliferation and survival (hallmarks of oncogenes), we found that *CDX2* alone does not transform immortalized rodent fibroblasts. However, this finding is consistent with a contextual role in tumorigenesis, the sine qua non of lineage-survival oncogenes. Indeed, to transform immortalized cells, *MITF* requires *BRAF* (55), *SOX2* requires *FOXEl* or *FGFR2* (44), and *TITF1* requires *NKX2-8* or *PAX9* (56). An important focus of future studies will be to elucidate the genetic context in which *CDX2* transforms cells.

Finally, we implicate Wnt/ β -catenin signaling as a possible mechanism by which *CDX2* contributes to colorectal tumorigenesis. Multiple lines of evidence support a connection between *CDX2* and the canonical Wnt/ β -catenin pathway: (i) *CDX2* knockdown reduces expression of select canonical Wnt pathway genes; (ii) *CDX2* binds genomic sites nearby select canonical Wnt pathway genes; (iii) *CDX2* binds adjacent to TCF/LEF motifs genome-wide; (iv) *CDX2* knockdown (in amplified cell lines) inhibits β -catenin/TCF reporter activity; and (v) *CDX2* expression in colorectal tumors significantly correlates with nuclear β -catenin levels. Nonetheless, the detailed mechanism(s) by which *CDX2* affects β -catenin activity and their relative importance remain to be clarified. Furthermore, because it is likely that Wnt/ β -catenin is not the sole mediator of *CDX2* oncogenic activity, future studies should explore other possible mediators, including noncanonical Wnt pathways, *HOX* genes, phosphatidylinositol signaling, and others.

Wnt/ β -catenin signaling plays an integral role in the development of the colonic epithelium and represents a key pathway altered during colorectal tumorigenesis (59). More than 90% of patients are purported to have alterations that affect the Wnt/ β -catenin pathway (60). Among our cell line collection, nearly all *MSI*[−] lines harbor the *APC* mutation, and nearly all *MSI*⁺ cell lines harbor mutation in *APC* or β -catenin. Nonetheless, the high frequency of *APC*/ β -catenin mutations does not preclude the importance of *CDX2* amplification and involvement in the Wnt pathway. In fact, despite the near ubiquity of *APC* mutations, less than half of colorectal cancers show positive nuclear β -catenin staining by IHC [42% in our study and 40% in a previous report (61)]. This result implies that more than one upstream mutation may be necessary to activate nuclear β -catenin signaling sufficiently. Our data are consistent with *CDX2* amplification possibly functioning in this role and warrant further investigation of the relationship between *CDX2* and Wnt/ β -catenin signaling. Additional work to characterize the genes or pathway dependencies invoked by *CDX2*-driven tumors, perhaps through synthetic lethal screening, may point to new selective vulnerabilities in colorectal cancers harboring 13q12.2 amplifications and should be of potential clinical value.

Materials and Methods

See *SI Appendix, Supplemental Materials and Methods* for detailed discussion of experimental procedures.

Colorectal Cancer Cell Lines. Colorectal cancer cell lines were obtained from the American Type Culture Collection or the European Collection of Cell Cultures, unless otherwise specified (*SI Appendix, Supplemental Materials and Methods*). All cell lines were grown in RPMI-1640 with 10% (vol/vol) FBS.

Array-Based Comparative Genomic Hybridization. Array-based comparative genomic hybridization (aCGH) was done using cDNA microarrays representing 22,488 mapped human genes, according to previously published protocols (62, 63), with map positions for arrayed cDNAs assigned using the National Center

for Biotechnology Information genome assembly (Build 36.1). Copy number data were segmented using Circular Binary Segmentation (CBS) (64), and statistically significant recurrent alterations were identified using GISTIC (17). COLO320 also was profiled using an Affymetrix SNP 6.0 array. For primary colorectal tumors, DNA copy number data were obtained from 123 tumors profiled on Affymetrix 250K Mapping SNP arrays (15) and from 98 tumors profiled on Agilent 244K oligonucleotide arrays (16). SNP array data were processed using GenePattern 2.0 (65), and copy number estimates were obtained using the SNP analysis module. SNP array and Agilent array copy number data were segmented using CBS and were analyzed using GISTIC. All genomic coordinates were mapped to the hg18 reference genome build. DNA amplifications and deletions were defined by CBS segments with mean log₂ ratios greater than 0.1 and less than −0.1, respectively.

Gene-Expression Profiling. To derive a *CDX2* gene-expression signature, COLO320 cells were transfected with either of two *CDX2*-targeting siRNAs (siCDX2-1 or siCDX2-3) or a nontargeting control (NTC) siRNA. Cells were harvested 24 h later (coinciding with the time of maximal *CDX2* protein knockdown), and RNA from siCDX2 and NTC was cohybridized to Agilent Whole Human Genome 44K gene-expression arrays. Expression data were normalized, and the intersection of the top 500 genes down-regulated upon *CDX2* knockdown by each of the two *CDX2*-targeting siRNA constructs, compared with NTC, comprised our *CDX2* gene-expression signature (185 genes). Enrichment for functionally related genes was evaluated by overlapping this signature with curated pathway gene sets from the Molecular Signatures Database (30), and enrichment was calculated using the hypergeometric distribution.

ChIP-seq. ChIP was performed using two independent antibodies against human *CDX2*. ChIP and input DNA were sequenced using a Genome Analyzer IIX (Illumina). Raw sequence image data were analyzed using the Illumina analysis pipeline and mapped to the unmasked human genome (NCBI36, hg18) using ELAND (Illumina). Mapped sequences were analyzed for enriched binding peaks using QuEST (66). The SeqPos program was used for de novo motif scanning of *CDX2*-binding regions and to identify enrichment of transcription factor-binding motifs. To integrate ChIP-seq-binding peaks with gene-expression data, we assigned each ChIP-seq peak to the nearest annotated gene.

FISH and IHC. A TMA was constructed using archived formalin-fixed, paraffin-embedded colon tissue specimens obtained with Institutional Review Board approval from the Santa Clara Valley Medical Center (San Jose, CA). The TMA contained duplicate 0.6-mm cores representing 76 evaluable colorectal adenocarcinomas. A set of overlapping, SpectrumOrange-labeled clones mapping to 13q14 (LSI13; Vysis) was cohybridized with a SpectrumGreen-labeled control chromosome 10-centromeric probe (CEP10; Vysis) (chromosome 10 exhibited few CNAs in the aCGH data). Slides were imaged using a fluorescence microscope with Applied Imaging CytoVision 3.0 software. Red (Chr13q) and green (chr10cen) FISH signals were counted for at least 20 cells per case, and chromosome 13 gain was defined by a red/green ratio >1.5. For IHC, we used a mouse monoclonal antibody directed against *CDX2* (Biogenex) or β -catenin (Cell Marque). *CDX2* staining was scored from 0 (no nuclear staining) to 3, indicating an increasing extent of nuclear staining. β -Catenin staining was scored as membranous only (1), membranous and cytoplasmic (2), or nuclear (3–6, indicating <5%, 5–10%, or >50% of tumor cell nuclei staining positive, respectively).

shRNA and cDNA Transductions. shRNA constructs in pLKO.1 vectors (The RNAi Consortium) were packaged into lentiviral particles, and target cell lines were transduced and cultured under puromycin selection. Knockdown was confirmed by Western blot analysis. All functional experiments were performed with transduced cells that had undergone fewer than three passages after puromycin selection (to minimize any selective pressure to revert knockdown). Lentiviral expression vectors pLOC-RFP and pLOC-*CDX2* were obtained from Open Biosystems and were transduced as above.

Functional Studies. Cell proliferation was quantified by colorimetry based on the metabolic cleavage of the tetrazolium salt WST-1 in viable cells (Roche). Apoptosis was assayed by annexin V staining, quantified by flow cytometry using the Vybrant Apoptosis Assay kit (Invitrogen). Cell-cycle distribution was quantified by flow cytometry using the BrdU-FITC Flow kit (BD Biosciences). For anchorage-independent (soft agar colony) growth, cells were plated in a layer of 0.18% (wt/vol) agarose on top of a layer of 0.36% (wt/vol) agarose. For colony growth on plastic (liquid medium colony growth), cells were plated sparsely in six-well plates. Colonies were stained after 2–3 wk with neutral red (soft agar) or Giemsa (plastic) and were counted using ImageJ

software (National Institutes of Health). Loss of contact inhibition was assayed by transduction of NIH 3T3 cells and counting visible foci 3 wk later. For β -catenin/TCF reporter assays, cells were transfected with either TOPflash or FOPflash firefly luciferase reporter constructs, with a *Renilla* luciferase reporter construct (pRL-TK; Promega) serving as an internal transfection control. Firefly and *Renilla* luciferase activities were measured using the Dual-Luciferase Reporter assay system (Promega), and the ratio was reported. All the above assays were performed as biological triplicates, and average values (± 1 SD) are reported. All experimental results were reproduced at least once.

ACKNOWLEDGMENTS. We thank the Stanford Functional Genomics Facility for microarray manufacture; the Stanford Microarray Database for database

support; Ilana Galperin, Dana Bangs (Stanford Cytogenetics Laboratory), and Kelli Montgomery for assistance with FISH analysis; Jennifer Lilla for assistance with immunofluorescence; Ziming Weng, Phil Lacroute, and Arend Sidow for Illumina sequencing support; Janet Bueno and the Stanford Tissue Bank staff for providing tissue specimens; and Rene van Amerongen and Roel Nusse for TOPflash and FOPflash reporter gene constructs. This work was supported in part by National Institutes of Health Grant CA112016 (to J. R.P.), by a grant from the Stanford Digestive Disease Center (to J.R.P.), and by fellowship/training grants from the Paul and Daisy Soros Foundation (to K.S.), Department of Defense (to K.S.), Medical Scientist Training Program (to K.S. and M.E.S.), Stanford Genome Training Program (to C.P.G.), and Stanford Cancer Biology Training Program (to M.E.K.). A.Y.L. was supported, in part, by the Valley Foundation.

- American Cancer Society (2011) *Cancer Facts & Figures 2011* (American Cancer Society, Atlanta).
- Vogelstein B, et al. (1988) Genetic alterations during colorectal-tumor development. *N Engl J Med* 319(9):525–532.
- Lengauer C, Kinzler KW, Vogelstein B (1997) Genetic instability in colorectal cancers. *Nature* 386(6625):623–627.
- Walther A, Houlston R, Tomlinson I (2008) Association between chromosomal instability and prognosis in colorectal cancer: A meta-analysis. *Gut* 57(7):941–950.
- Popat S, Houlston RS (2005) A systematic review and meta-analysis of the relationship between chromosome 18q genotype, DCC status and colorectal cancer prognosis. *Eur J Cancer* 41(14):2060–2070.
- Grady WM, Carethers JM (2008) Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology* 135(4):1079–1099.
- Rajagopalan H, et al. (2002) Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. *Nature* 418(6901):934.
- Sjöblom T, et al. (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* 314(5797):268–274.
- Wood LD, et al. (2007) The genomic landscapes of human breast and colorectal cancers. *Science* 318(5853):1108–1113.
- Camps J, et al. (2009) Integrative genomics reveals mechanisms of copy number alterations responsible for transcriptional deregulation in colorectal cancer. *Genes Chromosomes Cancer* 48(11):1002–1017.
- Douglas EJ, et al. (2004) Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. *Cancer Res* 64(14):4817–4825.
- Kim MY, et al. (2006) Recurrent genomic alterations with impact on survival in colorectal cancer identified by genome-wide array comparative genomic hybridization. *Gastroenterology* 131(6):1913–1924.
- Nakao K, et al. (2004) High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. *Carcinogenesis* 25(8):1345–1357.
- Garraway LA, Sellers WR (2006) Lineage dependency and lineage-survival oncogenes in human cancer. *Nat Rev Cancer* 6(8):593–602.
- Firestein R, et al. (2008) CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. *Nature* 455(7212):547–551.
- Veeriah S, et al. (2010) Somatic mutations of the Parkinson's disease-associated gene PARK2 in glioblastoma and other human malignancies. *Nat Genet* 42(1):77–82.
- Beroukhi R, et al. (2007) Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma. *Proc Natl Acad Sci USA* 104(50):20007–20012.
- Slattery ML, et al. (2004) Associations among IRS1, IRS2, IGF1, and IGFBP3 genetic polymorphisms and colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 13(7):1206–1214.
- Jonsson J, Carlsson L, Edlund T, Edlund H (1994) Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* 371(6498):606–609.
- Offield MF, et al. (1996) PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* 122(3):983–995.
- Beroukhi R, et al. (2010) The landscape of somatic copy-number alteration across human cancers. *Nature* 463(7283):899–905.
- Rhodes DR, et al. (2007) Oncomine 3.0: Genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* 9(2):166–180.
- Chawengsaksophak K, James R, Hammond VE, Köntgen F, Beck F (1997) Homeosis and intestinal tumours in Cdx2 mutant mice. *Nature* 386(6620):84–87.
- Aoki K, Tamai Y, Horike S, Oshima M, Taketo MM (2003) Colonic polyposis caused by mTOR-mediated chromosomal instability in Apc+/Delta716 Cdx2+/- compound mutant mice. *Nat Genet* 35(4):323–330.
- Bonhomme C, et al. (2003) The Cdx2 homeobox gene has a tumour suppressor function in the distal colon in addition to a homeotic role during gut development. *Gut* 52(10):1465–1471.
- Wicking C, et al. (1998) CDX2, a human homologue of Drosophila caudal, is mutated in both alleles in a replication error positive colorectal cancer. *Oncogene* 17(5):657–659.
- da Costa LT, et al. (1999) CDX2 is mutated in a colorectal cancer with normal APC/ beta-catenin signaling. *Oncogene* 18(35):5010–5014.
- Yagi OK, Akiyama Y, Yuasa Y (1999) Genomic structure and alterations of homeobox gene CDX2 in colorectal carcinomas. *Br J Cancer* 79(3–4):440–444.
- Woodford-Richens KL, et al. (2001) CDX2 mutations do not account for juvenile polyposis or Peutz-Jeghers syndrome and occur infrequently in sporadic colorectal cancers. *Br J Cancer* 84(10):1314–1316.
- Subramanian A, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102(43):15545–15550.
- Mikels AJ, Nusse R (2006) Purified Wnt5a protein activates or inhibits beta-catenin-TCF signaling depending on receptor context. *PLoS Biol* 4(4):e115.
- Wang HQ, Xu ML, Ma J, Zhang Y, Xie CH (2012) Frizzled-8 as a putative therapeutic target in human lung cancer. *Biochem Biophys Res Commun* 417(1):62–66.
- Gao Y, Wang HY (2006) Casein kinase 2 is activated and essential for Wnt/beta-catenin signaling. *J Biol Chem* 281(27):18394–18400.
- Berger MF, et al. (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* 133(7):1266–1276.
- Traber PG, Wu GD, Wang W (1992) Novel DNA-binding proteins regulate intestine-specific transcription of the sucrase-isomaltase gene. *Mol Cell Biol* 12(8):3614–3627.
- Suh E, Traber PG (1996) An intestine-specific homeobox gene regulates proliferation and differentiation. *Mol Cell Biol* 16(2):619–625.
- Eda A, et al. (2003) Aberrant expression of CDX2 in Barrett's epithelium and inflammatory esophageal mucosa. *J Gastroenterol* 38(1):14–22.
- Hinoi T, et al. (2002) CDX2 regulates liver intestine-cadherin expression in normal and malignant colon epithelium and intestinal metaplasia. *Gastroenterology* 123(5):1565–1577.
- Matys V, et al. (2003) TRANSFAC: Transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 31(1):374–378.
- Verzi MP, et al. (2010) TCF4 and CDX2, major transcription factors for intestinal function, converge on the same cis-regulatory regions. *Proc Natl Acad Sci USA* 107(34):15157–15162.
- 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(35):31909–31917.
- Benoit YD, et al. (2010) Cooperation between HNF-1alpha, Cdx2, and GATA-4 in initiating an enterocytic differentiation program in a normal human intestinal epithelial progenitor cell line. *Am J Physiol Gastrointest Liver Physiol* 298(4):G504–G517.
- Guo RJ, Funakoshi S, Lee HH, Kong J, Lynch JP (2010) The intestine-specific transcription factor Cdx2 inhibits beta-catenin/TCF transcriptional activity by disrupting the beta-catenin-TCF protein complex. *Carcinogenesis* 31(2):159–166.
- Bass AJ, et al. (2009) SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. *Nat Genet* 41(11):1238–1242.
- Network TCGA; Cancer Genome Atlas Network (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487(7407):330–337.
- Greenman C, et al. (2007) Patterns of somatic mutation in human cancer genomes. *Nature* 446(7132):153–158.
- Blyth K, Cameron ER, Neil JC (2005) The RUNX genes: Gain or loss of function in cancer. *Nat Rev Cancer* 5(5):376–387.
- 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(20):7641–7645.
- Beck F, Chawengsaksophak K, Waring P, Playford RJ, Furness JB (1999) Reprogramming of intestinal differentiation and intercalary regeneration in Cdx2 mutant mice. *Proc Natl Acad Sci USA* 96(13):7318–7323.
- Guo RJ, Suh ER, Lynch JP (2004) The role of Cdx proteins in intestinal development and cancer. *Cancer Biol Ther* 3(7):593–601.
- Mutoh H, et al. (2002) Conversion of gastric mucosa to intestinal metaplasia in Cdx2-expressing transgenic mice. *Biochem Biophys Res Commun* 294(2):470–479.
- Silberg DG, et al. (2002) Cdx2 ectopic expression induces gastric intestinal metaplasia in transgenic mice. *Gastroenterology* 122(3):689–696.
- Rawat VP, et al. (2004) Ectopic expression of the homeobox gene Cdx2 is the transforming event in a mouse model of t(12;13)(p13;q12) acute myeloid leukemia. *Proc Natl Acad Sci USA* 101(3):817–822.
- Visakorpi T, et al. (1995) In vivo amplification of the androgen receptor gene and progression of human prostate cancer. *Nat Genet* 9(4):401–406.
- Garraway LA, et al. (2005) Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* 436(7047):117–122.
- Kendall J, et al. (2007) Oncogenic cooperation and coamplification of developmental transcription factor genes in lung cancer. *Proc Natl Acad Sci USA* 104(42):16663–16668.
- Weir BA, et al. (2007) Characterizing the cancer genome in lung adenocarcinoma. *Nature* 450(7171):893–898.
- Kwei KA, et al. (2008) Genomic profiling identifies TTF1 as a lineage-specific oncogene amplified in lung cancer. *Oncogene* 27(25):3635–3640.

59. Reya T, Clevers H (2005) Wnt signalling in stem cells and cancer. *Nature* 434(7035):843–850.
60. Thorstensen L, et al. (2005) Genetic and epigenetic changes of components affecting the WNT pathway in colorectal carcinomas stratified by microsatellite instability. *Neoplasia* 7(2):99–108.
61. Kobayashi M, et al. (2000) Nuclear translocation of beta-catenin in colorectal cancer. *Br J Cancer* 82(10):1689–1693.
62. Pollack JR, et al. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 23(1):41–46.
63. Pollack JR, et al. (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci USA* 99(20):12963–12968.
64. Olshen AB, Venkatraman ES, Lucito R, Wigler M (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5(4):557–572.
65. Reich M, et al. (2006) GenePattern 2.0. *Nat Genet* 38(5):500–501.
66. Valouev A, et al. (2008) Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat Methods* 5(9):829–834.