Cell Reports
Report

A Polymorphic Enhancer near *GREM1* Influences Bowel Cancer Risk through Differential CDX2 and TCF7L2 Binding

Annabelle Lewis,[1] Luke Freeman-Mills,[1] Elisa de la Calle-Mustienes,[2] Rosa María Giráldez-Pérez,[2] Hayley Davis,[1] Emma Jaeger,[1] Martin Becker,[3] Nina C. Hubner,[4] Luan N. Nguyen,[4] Jorge Zeron-Medina,[5] Gareth Bond,[5] Hendrik G. Stunnenberg,[4] Jaime J. Carvajal,[2] Jose Luis Gomez-Skarmeta,[2] Simon Leedham,[1] and Ian Tomlinson[1,*]
[1]Molecular and Population Genetics Laboratory, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK
[2]Centro Andaluz de Biología del Desarrollo, CSIC-Universidad Pablo de Olavide-Junta de Andalucía, Carretera de Utrera Km1, 41013 Sevilla, Spain
[3]Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, the Netherlands
[4]Department of Molecular Biology, Radboud Institute for Molecular Life Science, Geert Grooteplein 26/28, 6525 GA Nijmegen, the Netherlands
[5]Ludwig Institute for Cancer Research, Ltd., Nuffield Department of Clinical Medicine, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford OX3 7DQ, UK
*Correspondence: iant@well.ox.ac.uk
http://dx.doi.org/10.1016/j.celrep.2014.07.020

SUMMARY

A rare germline duplication upstream of the bone morphogenetic protein antagonist *GREM1* causes a Mendelian-dominant predisposition to colorectal cancer (CRC). The underlying disease mechanism is strong, ectopic *GREM1* overexpression in the intestinal epithelium. Here, we confirm that a common *GREM1* polymorphism, rs16969681, is also associated with CRC susceptibility, conferring ~20% differential risk in the general population. We hypothesized the underlying cause to be moderate differences in *GREM1* expression. We showed that rs16969681 lies in a region of active chromatin with allele- and tissue-specific enhancer activity. The CRC high-risk allele was associated with stronger gene expression, and higher *Grem1* mRNA levels increased the intestinal tumor burden in Apc^{Min} mice. The intestine-specific transcription factor CDX2 and Wnt effector TCF7L2 bound near rs16969681, with significantly higher affinity for the risk allele, and CDX2 overexpression in CDX2/GREM1-negative cells caused re-expression of *GREM1*. rs16969681 influences CRC risk through effects on Wnt-driven *GREM1* expression in colorectal tumors.

INTRODUCTION

Bone morphogenetic proteins (BMPs) are critical for normal intestinal homeostasis, counteracting Wnt signaling toward the top of crypts, and allowing differentiation (Scoville et al., 2008). BMP antagonists, notably *GREM1*, are expressed in the mesenchyme at the crypt base, restricting BMP activity, potentiating Wnt signaling, and maintaining the epithelial stem cell compartment. Germline mutations in BMP pathway members occur in juvenile polyposis and hereditary mixed polyposis (HMPS) (Hardwick et al., 2008; Jaeger et al., 2008). HMPS is a rare, autosomal-dominant condition in which patients develop multiple types of colorectal polyp and cancer. HMPS results from a 40 kb duplication upstream of *GREM1,* which causes dramatically increased, ectopic *GREM1* expression in the normal epithelium (Jaeger et al., 2012). In the general population, moreover, single-nucleotide polymorphisms (SNPs) near BMP pathway genes have been found by genome-wide association studies to predispose to colorectal cancer (CRC) (Jaeger et al., 2012; Tomlinson et al., 2011).

rs4779584, near *GREM1*, was one of the first CRC SNPs to be found. Interestingly, rs4779584 does not show an association with predisposition to benign colorectal tumors (Carvajal-Carmona et al., 2013), suggesting that it may act after tumor initiation. To uncover functional variation tagged by rs4779584, we had previously refined its association signal using direct genotyping, imputation, and conditional logistic regression (Tomlinson et al., 2011). This analysis suggested that rs4779584 captured two independent association signals represented by SNPs rs16969681 and rs11632715 upstream of *GREM1*. In this study, we aimed to identify the functional variation tagged by rs16969681 and characterize the mechanism whereby it influenced CRC risk.

RESULTS AND DISCUSSION

rs16969681 Lies in a Region with Features of an Enhancer Element

An updated genetic fine-mapping analysis of SNPs upstream of *GREM1* using current SNP reference panels confirmed the existence of two independent CRC risk SNPs (details not shown). One of the signals was still represented by rs16969681. Other

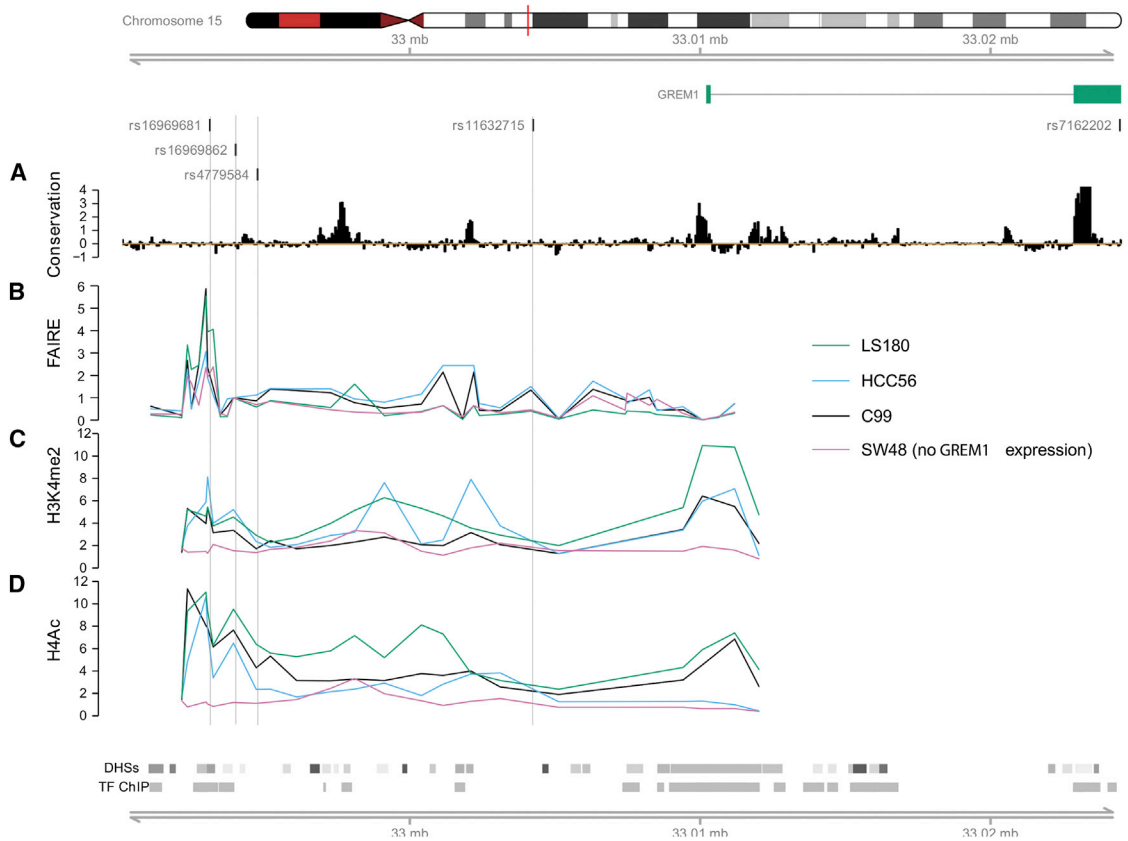

Figure 1. Chromatin Profiling Upstream of *GREM1*

(A) Diagram showing position on chromosome 15 and the main features in the 25 kb region upstream of the promoter, TSS, and transcribed region of *GREM1* (chr15:32.9–33.025 Mb; hg37), including exons (green boxes) and SNPs (black marks). Regional interspecific conservation (University of California, Santa Cruz phyloP100way all [vert.]), DNase1 hypersensitivity (wgEncodeReg Dnase Clustered V2), and TF-binding (wgEncodeReg Tfbs Clustered V3) sites are also shown in the bottom panel. TFs that bind at or adjacent to rs16969681 in ENCODE nonintestinal cell lines include HDAC2, BHLHE40, FOXA1, FOXA2, MAFK, HNF4G, SP1, RXRA, GATA3, MYBL2, p300, HNF4A, and TCF7L2.

(B) FAIRE assay on four CRCCLs: LS180 (*GREM1*-high; rs16969681 C/C), C99 (*GREM1*-high; T/T), HCC56 (*GREM1*-low; T/T), and SW48 (*GREM1* null; C/C). The graph shows relative enrichment at each primer set as determined by the ΔΔCt method, first by using noncrosslinked genomic DNA to normalize the results from nucleosome-depleted DNA and then by expressing each primer set relative to primers at chr15:32,993,925, telomeric to the rs16969681 signal.

(C and D) ChIP assays with anti-H3K4Me2 and anti-H4Ac antibodies, respectively, analyzed by SYBR green qPCR on four CRC cell lines as in (B). Relative enrichment was determined by the ΔΔCt method using sonicated input DNA to normalize the values from immunoprecipitated DNA and expressing this ratio relative to that seen at the *Rhodopsin* promoter as a negative control.

(B), (C), and (D) all show an area of active chromatin in the region surrounding rs16969681 with rs16969862 at its telomeric edge. Chromatin at rs4779584 shows no enrichment for activating marks. See also Table S1 and Figures S1 and S2.

SNPs in strong linkage disequilibrium (LD) with rs16969681 showed an ~20-fold lower likelihood of association with disease (Table S1). Because rs16969681 tags SNPs in intergenic DNA, its association is most likely caused by modifying a gene regulatory element. We therefore investigated the epigenetic characteristics of the rs16969681 region (Figure 1). ENCODE shows an area of DNase hypersensitivity and transcription factor binding close to rs16969681. However, the cell lines used in these studies are not of intestinal origin, so we wished to expand these studies in appropriate cell types. Although *GREM1* is principally expressed by mesenchymal cells in normal intestine, we found *GREM1* mRNA in 11/32 CRC cell lines (CRCCLs), which are all of epithelial origin (Figure S1). We therefore performed formaldehyde-assisted identification of regulatory elements (FAIRE) in four CRCCLs to identify areas of nucleosome-depleted chromatin in the 25 kb region upstream of the *GREM1* promoter, including rs4779584 and rs16969681. The highest signal in *GREM1*-expressing lines was a 1.3 kb peak centered on rs16969681 (Figure 1B). Chromatin immunoprecipitation (ChIP) using antibodies against H3K4Me2 (marker of promoters and enhancers) and panH4Ac (marker of active chromatin) also showed signals at rs16969681 and the *GREM1* transcription start site (TSS), specifically in *GREM1*-expressing lines (Figures 1C and 1D). The chromatin signature was indicative of a regulatory element, probably an enhancer, located at or near rs16969681. In support of this, we demonstrated an interaction between the region containing rs16969681 and *GREM1*'s promoter using chromosome conformation capture (Jaeger et al., 2012; Figure S2).

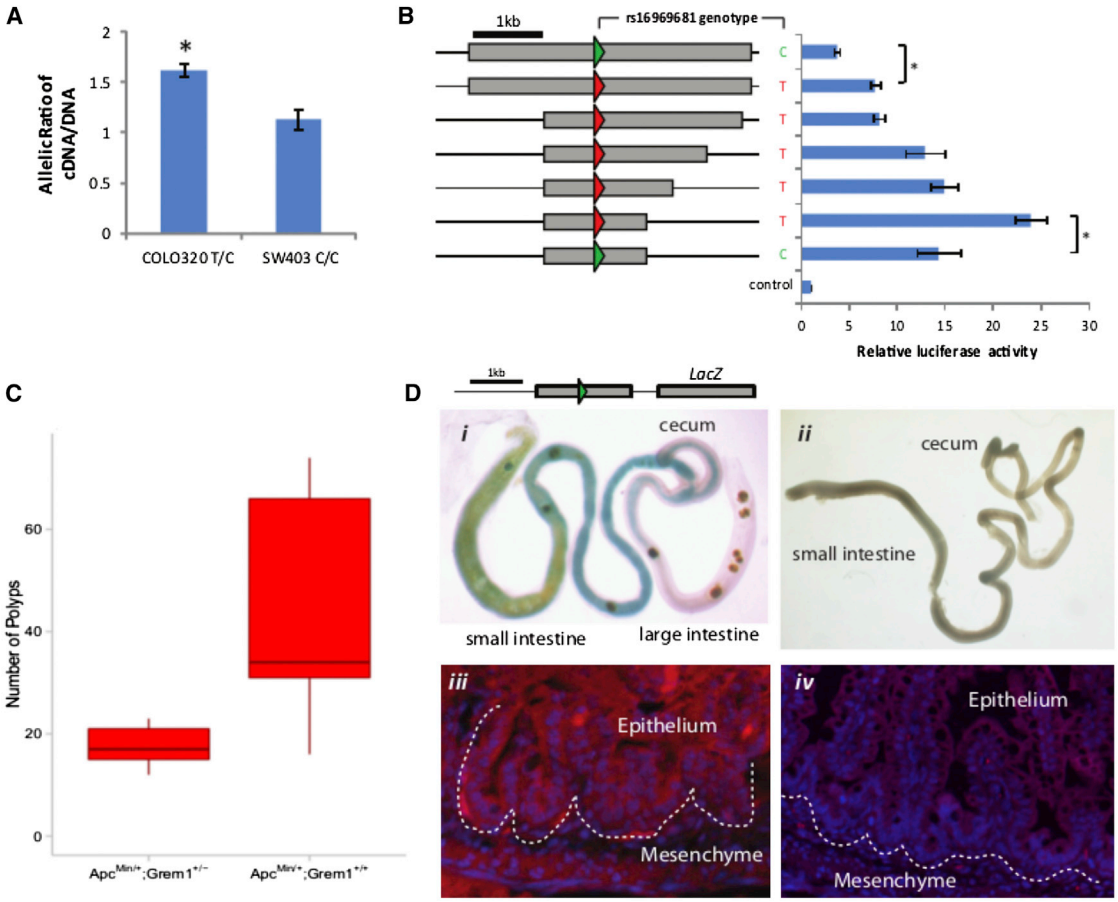

Figure 2. Allele-Specific Effects of an Enhancer at rs16969681

(A) rs16969681 heterozygote COLO320 shows allele-specific expression of *GREM1*. cDNA isolated from COLO320 cells was analyzed to measure ASE of SNP rs7162202 (*GREM1* 3′ UTR). The chart shows the results after normalization with COLO320 genomic DNA. There was a mean G/T ratio of 1.62 ($p = 0.001$; t test). rs16969681 homozygous cell line SW403 showed a normalized ratio of 1.13 ($p = 0.11$).

(B) Luciferase reporter assays to analyze the enhancer activity of 1.0–4.1 kb fragments containing rs16969681 upstream of *GREM1* using the pGL3 promoter vector in SW948 cells. Scale diagrams are drawn for each fragment. Green arrows represent the low-risk and red arrows the high-risk allele. The bar chart shows the average luciferase activity of triplicate experiments normalized using the pGL4 Renilla activity and expressed relative to pGL3 containing a neutral stretch of DNA (control). The high-risk allele causes a significant increase in enhancer activity ($p < 0.0002$). The smallest (1.4 kb) fragment containing rs16969681 shows high activity, which is increased nearly 2-fold in the presence of the risk allele ($p = 0.029$).

(C) Box plot showing significant reduction in polyp numbers in $Grem1^{+/-}$;$Apc^{Min/+}$ mice compared to $Grem1^{+/+}$;$Apc^{Min/+}$ littermates. In the former, mean polyp number at 5 to 6 months was 18 ($n = 5$; SD = 4.4); in the latter, mean polyp number was 43 ($n = 9$; SD = 22; $p = 0.0162$; Wilcoxon test).

(D) Gastrointestinal expression of a rs16969681-*LacZ* reporter construct. Intestines and stomachs were removed from transient transgenic mice (16 dpc). The upper panels show β-galactosidase expression in the reporter animal (i), but not in wild-type controls (ii). The lower panels show anti-β-galactosidase immunofluorescence (red). Reporter expression is present in both the mesenchymal (lower) and epithelial compartments (upper) of the intestine in transgenic mice (iii), but not their wild-type littermates (iv).

Error bars represent SEM.

The rs16969681 Enhancer Regulates Gene Expression in an Allele-Specific Manner

We next determined whether the rs16969681 genotype influenced *GREM1* mRNA. rs16969344, a SNP in perfect LD with rs16969681, was not associated with total *GREM1* mRNA levels in The Cancer Genome Atlas CRC samples. However, differences in gene-expression levels among cancers may have many different causes, and we therefore searched for allele-specific expression (ASE) differences. One CRCCL in our panel (COLO320) was heterozygous at rs16969681, had two copies of chromosome 15q, expressed *GREM1*, and was heterozygous for a suitable SNP in the *GREM1* transcript (rs7162202 in the 3′ UTR). We found that the normalized allelic ratio (G:T) of rs7162202 in COLO320 cDNA was 1.62 ($p = 0.001$; Figure 2A) compared with 1.1 ($p = 0.11$) in a control CRCCL homozygous for rs16969681 (SW403), consistent with ASE in COLO320. Due to the large distance (>30 kb) between rs16969681 and rs7162202 and no suitable cell line for copy-number-based phasing, we could not determine whether the high-risk allele was associated with higher expression in *cis*. We therefore carried

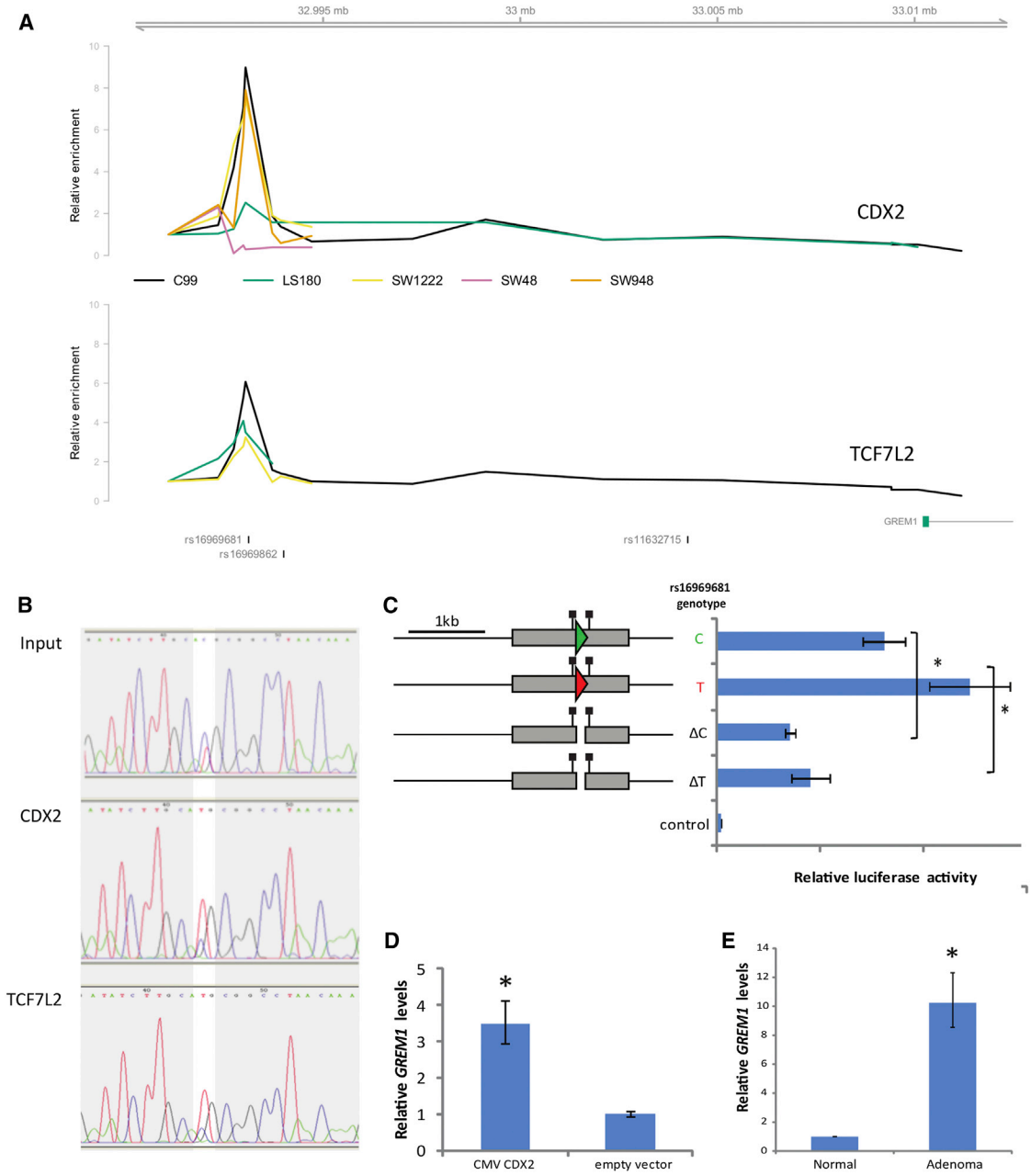

Figure 3. CDX2 Binding at rs16969681 in *GREM1*-Expressing CRCCLs and Its Effect on *GREM1* Expression

(A) ChIP assays with anti-CDX2 and anti-TCF7L2 antibodies analyzed by SYBR green qPCR on a number of CRC cell lines (C99 T/T, LS180 C/C, SW1222 C/T, SW48 C/C, and SW948 C/C). The graphs show enrichment around rs16969681 in *GREM1*-expressing cell lines irrespective of genotype. Relative enrichment was determined by the ΔΔCt method, using sonicated input DNA to normalize the enrichment seen in immunoprecipitated DNA and then expressing each primer set relative to primers at chr15: 32,993,925. Despite repeated attempts, COLO320 (C/T) chromatin did not immunoprecipitate efficiently enough for use in ChIP assays.

(B) Sanger sequence traces of input, CDX2 ChIP, and TCF7L2 ChIP output from SW1222 highlighting rs16969681. A bias toward the risk T allele (red) at rs16969681 in the ChIP output is not seen in the input. This was confirmed by formal allele counting.

(C) Luciferase reporter assays to analyze the enhancer activity of *GREM1* upstream fragments containing the 1.4 kb around rs16969681 cloned into pGL3 promoter vector carried out in SW948 cells. Green arrows represent the low-risk alleles and red arrows the high risk. Black squares denote CDX2 core-binding motifs. The graph shows the average luciferase activity of three experiments (where each construct was analyzed in triplicate) normalized using the pGL4 Renilla activity and expressed relative to pGL3 promoter containing a neutral stretch of DNA (control). Deletion of a 71 bp EcoR1/ZraI fragment removing rs16969681 and disrupting bases adjacent to CDX2-binding motifs caused a significant reduction in enhancer activity.

(D) CDX2 overexpression experiment to analyze the effect on endogenous *GREM1* mRNA expression. Two CMV-CDX2 expression plasmids and an empty vector were each transfected in triplicate into the SW480 CRCCL, chosen for its low endogenous *CDX2* and *GREM1* expression and its consistently high levels of *CDX2*

(legend continued on next page)

out luciferase reporter assays to investigate allele-specific enhancer activity of the rs16969681 region in vitro in SW948 cells (Figure 2B). Constructs containing the low-risk allele showed moderate enhancing activity, but this was significantly increased when rs16969681 was mutated to the high-risk allele ($p < 0.001$). A series of deletion constructs around rs16969681 showed increasing enhancer activity (Figure 2B). The smallest construct, with the greatest enhancer activity, showed rs16969681's high-risk allele to have almost double the activity of the low-risk allele ($p = 0.029$). Deletion of the rs16969681 region, leaving rs16969862 alone, showed no enhancing effect. Thus, we found that the region containing rs16969681 carries the hallmarks of an allele-specific transcriptional enhancer.

Grem1 Levels Modify Intestinal Tumorigenesis in Mice

We wished to test whether modest changes in *GREM1* levels, such as those resulting from the rs16969681 genotype, could affect intestinal tumorigenesis in vivo. Because the region around rs16969681 is not highly conserved, we used constitutionally heterozygous *Grem1*$^{+/-}$ mice (Gazzerro et al., 2007) as a model for the effects of rs16969681. We confirmed that these animals had ~50% reduced *Grem1* mRNA expression relative to wild-type mice and showed no intestinal abnormalities or polyposis phenotype. *Grem1*$^{+/-}$;*Apc*$^{Min/+}$ animals had a >50% reduction in adenomas compared with their well-characterized *Grem1*$^{+/+}$;*Apc*$^{Min/+}$ littermates ($p = 0.0162$; Wilcoxon test; Figure 2C; Su et al., 1992). We examined tumors from littermates for differences in proliferation (Ki67 immunohistochemistry), apoptosis (Casp3), Wnt signaling (β-catenin), stem cell numbers (Sox9), and canonical BMP pathway activity (phospho-Smad1,5,8). We observed increased apoptosis in *Grem1*$^{+/-}$;*Apc*$^{Min/+}$ tumors ($p = 0.044$; $n = 12$; t test) and no other differences between the mouse genotypes.

The rs16969681 Enhancer Shows Tissue-Specific Activity In Vivo

In order to confirm that the rs16969681 enhancer is functional in vivo, we generated transgenic mice carrying a 2 kb fragment containing rs16969681 upstream of a *LacZ* reporter gene. β-galactosidase expression was found specifically in the intestines and stomach and was absent from other tissues (Figure 2D). Interestingly, reporter expression was seen in both mesenchymal and epithelial cells of the intestine (Figure 2Diii), though *GREM1* is not normally expressed in the epithelium. This suggests an epithelial role for the enhancer.

CDX2 and TCF7L2 Bind the Enhancer in an Allele-Specific Fashion and CDX2 Drives *GREM1* Expression

Next, we sought to identify the mechanism by which rs16969681 genotype could modify enhancer activity by identifying transcription factors (TFs) that bound to the surrounding region. We used biotin-labeled electrophoretic mobility shift assay probes, a development of Butter et al. (2012) to pull down DNA-protein complexes from SW948 and C99 nuclear extracts, followed by quantitative mass-spectrometry (N.C.H., L.N.N., N.C. Hornig, and H.G.S., unpublished data). We combined these data, in silico predictions from Biobase/TRANSFAC in a 50 bp region around rs16969681, and mining of public ChIP sequencing (ChIP-seq) databases (Mokry et al., 2010; Verzi et al., 2010) to generate a list of candidate TFs. We assessed these proteins using ChIP in *GREM1*-expressing cells (C99 rs16969681 T/T, LS180 C/C, SW1222 C/T, and SW948 C/C) and a nonexpressing line (SW48 C/C). CDX2 and TCF7L2 (also known as TCF4) bound to the rs16969681 region (Figure 3A), as did p300, RUNX1, and, less reliably, HNF4A (Figure S3). TCF7L2, HNF4A, and p300 are also listed as binding in published ENCODE data.

In order to investigate allele-specific TF binding, we quantified rs16969681 alleles in the ChIP output from the heterozygous cell line SW1222. p300 and RUNX1 showed no evidence of allele-specific binding ($p > 0.5$). HNF4A binding was not present in all heterozygous cell lines. However, CDX2 (odds ratio [OR] = 4.7; $p = 0.001$; $n = 81$; Fisher's exact test) and TCF7L2 (OR = 2.0; $p = 0.028$; $n = 63$) showed a significant bias toward the risk allele (Figure 3B). We also found a correlation between *GREM1* and *CDX2* mRNA expression in our panel of CRC cell lines ($p = 0.0088$; $n = 34$; Spearman's rank correlation). These data again indicate that the enhancer at rs16969681 is important in cells of epithelial origin.

The core consensus CDX2-binding motif, ATAAA (Verzi et al., 2010), occurs both 14 bp upstream and 65 bp downstream of rs16969681. To confirm that the CDX2-binding region of the rs16969681 enhancer is important for its activity, we deleted a 71 bp fragment from our luciferase reporter constructs (Figure 2B), removing rs16969681 and disrupting the bases adjacent to the core CDX2-binding motifs. Figure 3C shows the resultant significant reduction in enhancer activity and loss of allele-specific expression in SW948 cells. To demonstrate that CDX2 protein can directly regulate *GREM1* transcription, we overexpressed CDX2 in SW480, a CRCCL with low endogenous CDX2 and *GREM1*. We achieved consistent CDX2 overexpression using two separate constructs, causing an ~3.5-fold increase in *GREM1* expression ($p = 0.0005$; Figure 3D). By comparison, in cell lines with already high levels of GREM1 or CDX2, no effect or less consistent results were seen.

Epithelial *GREM1* Expression Occurs in Sporadic Colorectal Adenomas

Several lines of evidence presented above suggest that the rs16969681 genotype might have its effects in dysplastic colorectal epithelium, that is, at a stage after tumor initiation (Carvajal-Carmona et al., 2013). We noted that CDX2 expression

following transfection (confirmed by both qRT-PCR and western blot). mRNA was extracted and analyzed by Taqman, also in triplicate. The chart shows the relative expression of *GREM1* mRNA in the overexpression samples compared to empty vector calculated using the ΔΔCt method, normalized to *GAPDH*. Each overexpression plasmid individually caused a similar ~3.5-fold increase in *GREM1* expression. The chart shows the combined results from both ($p = 0.0005$; t test). (E) *GREM1* expression in three sporadic adenomas relative to paired normal tissue. Two to six crypts were isolated from each sample and mRNA levels analyzed by Taqman qRT-PCR. The chart shows relative *GREM1* levels calculated by ΔΔCt, normalized to *GAPDH*. Adenomas had a mean of 10.3 times more *GREM1* mRNA than paired normal crypts ($p = 0.006$; t test).
See also Figure S3. Error bars represent SEM.

occurs in almost all colorectal adenomas but is lost in some CRCs (Qualtrough et al., 2002), and we therefore wondered whether *GREM1* is expressed in adenomas as well as CRCs. We assayed *GREM1* mRNA expression in 15 isolated crypts from three sporadic adenomas and found a 10-fold increase in epithelial expression compared with surrounding normal epithelial tissue ($p = 0.002$; ANOVA; Figure 3E).

Concluding Remarks and Proposed Model of *GREM1* Regulation by CDX2 and TCF7L2 Binding to the rs16969681 Enhancer

We have found that the *GREM1* locus harbors at least two SNPs that independently influence the risk of CRC. Studies of *Grem1* expression in the developing limb bud (Zuniga et al., 2012) have shown that proximal and distal global control regions (GCRs) near *GREM1* regulate gene expression in a complex fashion. The SNP rs16969681 lies upstream of *GREM1* in a region distinct from the known GCRs and is close to a regulatory element that we have shown to act as an allele-specific *GREM1* enhancer. Mouse models confirm that relatively modest changes in *Grem1* expression influence intestinal tumor burden, possibly through variation in apoptosis. The rs16969681 region differentially binds the intestine-specific TF CDX2 and the Wnt effector TCF7L2. Both these transcription factors play important roles in intestinal function and often co-occupy binding sites (Verzi et al., 2010, 2013). CDX2 is described as a master TF that establishes intestinal epithelial identity (Gao et al., 2009; Qualtrough et al., 2002) and maintains open chromatin (Verzi et al., 2013) for TFs such as TCF7L2. Although further investigation is required, we hypothesize (Figure 4) that, with the possible exception of the stem cell compartment, CDX2 allows TCF7L2 to act as repressor in the Wnt-low environment of the normal epithelium. However, with high levels of Wnt, such as colorectal tumors, TCF7L2 complexes with β-catenin and drives *GREM1* expression, promoting tumorigenesis.

EXPERIMENTAL PROCEDURES

Detailed methodology is presented in Supplemental Information. All error bars represent the SEM.

CRC Patients

Sample sets are those reported in Whiffin et al. (2013), plus UK2 and Scotland2 reported in Tomlinson et al. (2011).

Formaldehyde-Assisted Identification of Regulatory Elements

FAIRE was carried out using the method adapted from Giresi et al. (2007). Briefly, ~10^8 cells were crosslinked for 5 min with 1% formaldehyde, neutralized, washed, and scraped. Uncrosslinked cells were used as a control. Cells were lysed and the chromatin sonicated and extracted with phenol:chloroform. Purified DNA was analyzed by SYBR green quantitative PCR (qPCR) using primers from Table S2.

Native ChIP

Native ChIP was carried out as previously described (Umlauf et al., 2004). Approximately 10^8 cells were collected, washed in PBS, and lysed, and chromatin was purified using a sucrose cushion. Chromatin was fragmented with MNase1, and immunoprecipitation (H3K4Me2, 07-030 Millipore; H4Ac, 06-866 Millipore) was carried out overnight. Antibody complexes were captured with protein A or G Dynabeads (Invitrogen) and washed. DNA was eluted and purified. Purified DNA was analyzed by SYBR green qPCR as above.

Crosslinked ChIP

Approximately 10^8 cells were crosslinked for 10 min with 1% formaldehyde, neutralized, washed, and scraped. Cells were lysed and the chromatin sonicated prior to overnight immunoprecipitation (anti-CDX2, Bethyl Laboratories; anti-TCF7L2, Santa Cruz). Antibody complexes were incubated with protein A or G beads, washed, and eluted. Purified DNA was analyzed by SYBR green qPCR as above. For allelic analysis, PCR products were TA cloned using pGem TAeasy (Promega) followed by colony PCR and sequencing.

Sequenom MassARRAY

For ASE analysis, genomic DNA and RNA were extracted from CRCCLs COLO320 and SW403 and cDNA generated. Paired cDNA and genomic DNA samples, in triplicate, were processed using the Iplex assay and detected by Sequenom MALDI-TOF mass spectrometry. The products were run using the Sequenom genotyping protocol, visualized using the MassARRAY Typer 4.0, and the areas under the curves used to calculate the allelic ratios of each sample.

Luciferase Reporter Assays

Luciferase reporter assays were carried out as described (Jaeger et al., 2012). Briefly, putative enhancer fragments in the pGL3 promoter vector (Promega) were cotransfected into SW948 cells with 1/10 concentration of pGL4 Renilla vector. Luciferase and Renilla activity was measured using the Dual Luciferase Assay system (Promega) and the Avent luminometer. Readings were normalized and expressed relative to a control plasmid.

Transgenic Reporter Assays

All work was performed according to Spanish regulations for animal experimentation and approved by the Consejo Superior de Investigaciones Cientificas ethics committee. Two-kilobase constructs containing rs16969681 and rs16969862, the minimal beta-globin promoter, *LacZ*, and a SV40 polyadenylation signal were linearized, diluted to 5 ng/ml in microinjection buffer, and used to inject fertilized mouse eggs from CBA/Ca × C57Bl/6 crosses. F0 embryos of 9.5–13 days postcoitum (dpc) stages and tissues from F0 animals at postnatal day 1 (P1) and P35 were harvested. Embryos and dissected tissues were fixed overnight, washed and placed in 10 ml of X-gal solution for 2–20 hr at 37°C, and then postfixed.

ChIP-Seq Data Mining

CDX2 ChIP-seq data from CACO2 cells (Gene Expression Omnibus series GSE23436; Verzi et al., 2010) and TCF7L2 data from LS147T cells (series GSE18481; Mokry et al., 2010) were downloaded and mined for signals in the *GREM1* gene and surrounding region.

CDX2 Overexpression Assays

SW480 cells were plated out and transfected with CMVCDX2 (Addgene) and CMVFLAG-CDX2 (Barros et al., 2011). Cells were lysed, RNA and cDNA generated, and *CDX2* and *GREM1* mRNA levels measured by quantitative RT-PCR (qRT-PCR) using the Taqman system (Applied Biosystems).

Adenoma GREM1 mRNA Expression

Individual crypts from normal and adenomatous patient epithelium were isolated as described previously (Jaeger et al., 2012) to ensure exclusion of any pericryptal myofibroblasts. cDNA was generated and analyzed by qRT-PCR as above.

Transgenic Mouse Crosses

$Grem1^{+/flox}$ mice, a kind gift of E. Canalis, were crossed with *Pgk-Cre* mice to inactivate one copy of *Grem1* in the germline. These mice were crossed with $Apc^{Min/+}$ mice to generate $Grem1^{+/-}$; $Apc^{Min/+}$ and $Grem1^{+/+}$; $Apc^{Min/+}$ animals. Mice were aged for 5 to 6 months and litters sacrificed when one member showed symptoms of intestinal tumors. Intestines were dissected, fixed, and stained and polyps counted. All procedures were carried out in

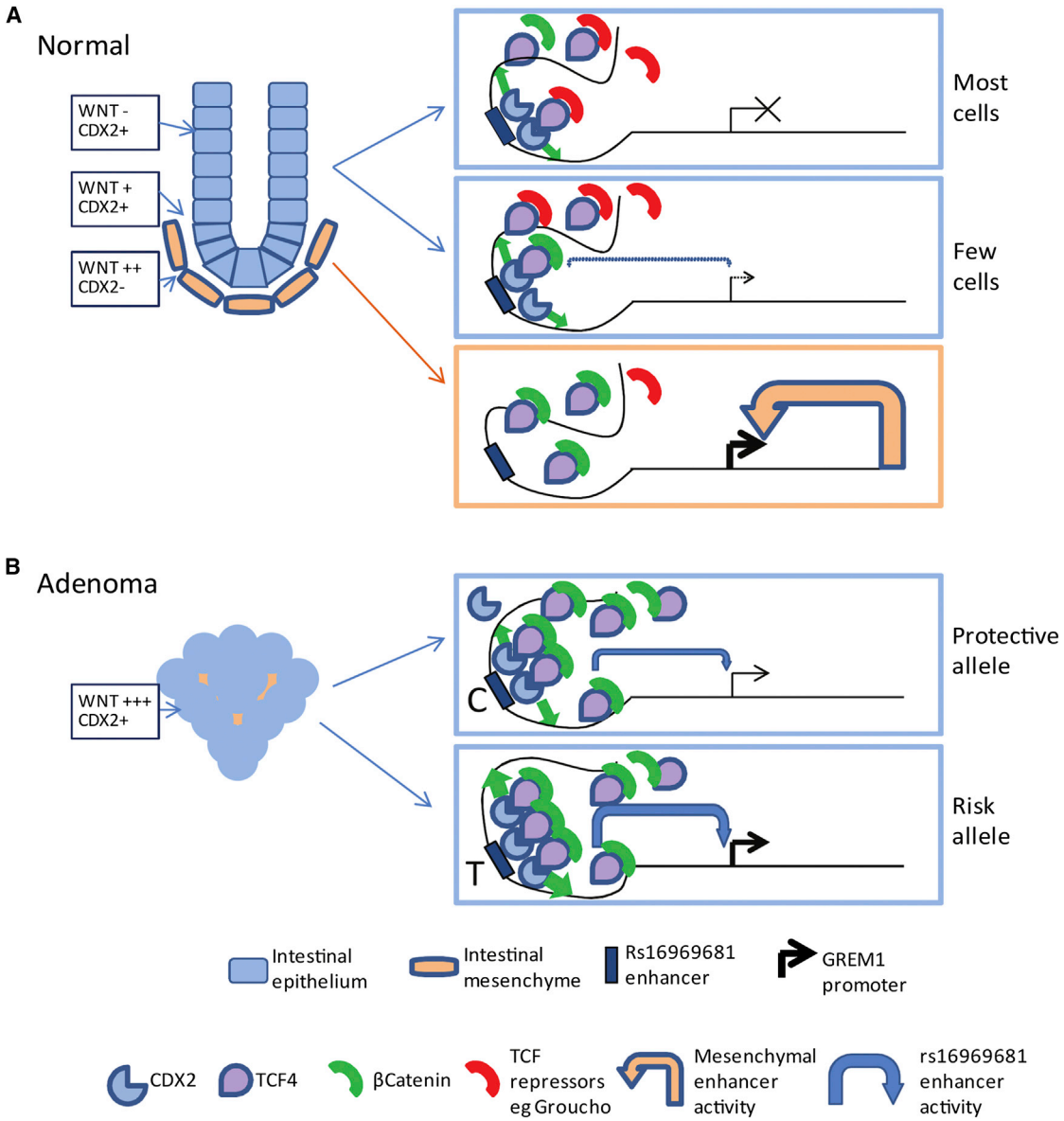

Figure 4. Proposed Model of Action of CDX2 and TCF7L2 on *GREM1* Expression through Binding Close to rs16969681

The model describes how the rs16969681 enhancer might regulate *GREM1* transcription in (A) normal intestines and (B) adenomas based on the previously proposed model of CDX2 acting as a pioneer TF for TCF7L2.

(A) The normal crypt is subjected to a Wnt gradient with high levels at the base and low/absent levels at the neck. The top panel (blue) shows the situation for most normal crypt epithelium cells. Here, CDX2 binds to the enhancer, partially disrupting the repressive chromatin, but Wnt signaling, and therefore active β-catenin, is at low or moderate levels, so the majority of TCF7L2 is bound by repressors such as Groucho (Cavallo et al., 1998; Roose et al., 1998). Allele-specific effects of CDX2 binding are not apparent because *GREM1* expression is at a very low basal level and essentially repressed. There is no enhancer activity. In the middle panel, Wnt levels are higher, as for example in the stem cell zone at the crypt base. Here, activated TCF7L2/β-catenin complexes bind together with CDX2 and low levels of enhancer activity result; this is consistent with evidence that intestinal stem cells produce low levels of secreted BMP antagonists (He et al., 2004). The bottom panel (orange) shows the mesenchyme, where no CDX2 is present, although there are high levels of Wnt. *GREM1* expression is activated by unknown mesenchymal enhancers or mesenchyme-specific TFs binding at rs16969681.

(B) In the pathological setting of an adenoma, *APC* or *CTNNB1* mutations cause constitutive Wnt activation and high levels of Wnt target gene expression. CDX2 opens up the chromatin, allowing high levels of activated TCF7L2/β-catenin to cause *GREM1* transcription. On the rs16969681 risk allele (bottom panel), extra CDX2 binding results in higher enhancer activity. The downstream effects of the increased *GREM1* may include an increase in stem-like cells and angiogenesis.

accordance with Home Office UK regulations and the Animals (Scientific Procedures) Act 1986. Mice were housed at the Functional Genomics Facility, Wellcome Trust Centre for Human Genetics, Oxford.

Immunohistochemistry

Immunohistochemistry was performed as described in Jaeger et al. (2012) using anti-CASP3 (R&D Systems).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.07.020.

AUTHOR CONTRIBUTIONS

A.L., S.L., and I.T. conceived the study. A.L., L.F.-M., E.d.I.C.-M., R.M.G.-P., H.D., E.J., M.B., N.C.H., and L.N.N. carried out the experiments. A.L., L.F.-M., N.C.H., J.Z.-M., S.L., and I.T. analyzed the data. J.Z.-M., G.B., H.G.S., J.J.C., J.L.G.-S., S.L., and I.T. gave technical and conceptual advice and supervised the experiments. A.L., L.F.-M., and I.T. wrote the manuscript.

ACKNOWLEDGMENTS

Funding was provided from Cancer Research UK grant A/16459, an EU FP7 SYSCOL Consortium grant, and the EU COST colorectal cancer initiative. Core funding to the Wellcome Trust Centre for Human Genetics was provided from the Wellcome Trust (090532/Z/09/Z). J.L.G.-S. and J.J.C. were supported by the Spanish/FEDER government grants BFU2010-14839 and BFU2011-2292. We thank R. Almeida for the CMV-CDX2 overexpression plasmid, E. Canalis for *Grem1* mice, and the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics for Sequenom data.

Received: April 4, 2014
Revised: June 9, 2014
Accepted: July 15, 2014
Published: August 14, 2014

REFERENCES

Barros, R., da Costa, L.T., Pinto-de-Sousa, J., Duluc, I., Freund, J.N., David, L., and Almeida, R. (2011). CDX2 autoregulation in human intestinal metaplasia of the stomach: impact on the stability of the phenotype. Gut *60*, 290–298.

Butter, F., Davison, L., Viturawong, T., Scheibe, M., Vermeulen, M., Todd, J.A., and Mann, M. (2012). Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. PLoS Genet. *8*, e1002982.

Carvajal-Carmona, L.G., Zauber, A.G., Jones, A.M., Howarth, K., Wang, J., Cheng, T., Riddell, R., Lanas, A., Morton, D., Bertagnolli, M.M., and Tomlinson, I.APC Trial Collaborators; APPROVe Trial Collaborators; CORGI Study Collaborators; Colon Cancer Family Registry Collaborators; CGEMS Collaborators (2013). Much of the genetic risk of colorectal cancer is likely to be mediated through susceptibility to adenomas. Gastroenterology *144*, 53–55.

Cavallo, R.A., Cox, R.T., Moline, M.M., Roose, J., Polevoy, G.A., Clevers, H., Peifer, M., and Bejsovec, A. (1998). Drosophila Tcf and Groucho interact to repress Wingless signalling activity. Nature *395*, 604–608.

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

Gazzerro, E., Smerdel-Ramoya, A., Zanotti, S., Stadmeyer, L., Durant, D., Economides, A.N., and Canalis, E. (2007). Conditional deletion of gremlin causes a transient increase in bone formation and bone mass. J. Biol. Chem. *282*, 31549–31557.

Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R., and Lieb, J.D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res. *17*, 877–885.

Hardwick, J.C., Kodach, L.L., Offerhaus, G.J., and van den Brink, G.R. (2008). Bone morphogenetic protein signalling in colorectal cancer. Nat. Rev. Cancer *8*, 806–812.

He, X.C., Zhang, J., Tong, W.G., Tawfik, O., Ross, J., Scoville, D.H., Tian, Q., Zeng, X., He, X., Wiedemann, L.M., et al. (2004). BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling. Nat. Genet. *36*, 1117–1121.

Jaeger, E., Webb, E., Howarth, K., Carvajal-Carmona, L., Rowan, A., Broderick, P., Walther, A., Spain, S., Pittman, A., Kemp, Z., et al.; CORGI Consortium (2008). Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat. Genet. *40*, 26–28.

Jaeger, E., Leedham, S., Lewis, A., Segditsas, S., Becker, M., Cuadrado, P.R., Davis, H., Kaur, K., Heinimann, K., Howarth, K., et al. (2012). Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. Nat. Genet. *44*, 699–703.

Mokry, M., Hatzis, P., de Bruijn, E., Koster, J., Versteeg, R., Schuijers, J., van de Wetering, M., Guryev, V., Clevers, H., and Cuppen, E. (2010). Efficient double fragmentation ChIP-seq provides nucleotide resolution protein-DNA binding profiles. PLoS ONE *5*, e15092.

Qualtrough, D., Hinoi, T., Fearon, E., and Paraskeva, C. (2002). Expression of CDX2 in normal and neoplastic human colon tissue and during differentiation of an in vitro model system. Gut *51*, 184–190.

Roose, J., Molenaar, M., Peterson, J., Hurenkamp, J., Brantjes, H., Moerer, P., van de Wetering, M., Destrée, O., and Clevers, H. (1998). The Xenopus Wnt effector XTcf-3 interacts with Groucho-related transcriptional repressors. Nature *395*, 608–612.

Scoville, D.H., Sato, T., He, X.C., and Li, L. (2008). Current view: intestinal stem cells and signaling. Gastroenterology *134*, 849–864.

Su, L.K., Kinzler, K.W., Vogelstein, B., Preisinger, A.C., Moser, A.R., Luongo, C., Gould, K.A., and Dove, W.F. (1992). Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. Science *256*, 668–670.

Tomlinson, I.P., Carvajal-Carmona, L.G., Dobbins, S.E., Tenesa, A., Jones, A.M., Howarth, K., Palles, C., Broderick, P., Jaeger, E.E., Farrington, S., et al.; COGENT Consortium; CORGI Collaborators; EPICOLON Consortium (2011). Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. PLoS Genet. *7*, e1002105.

Umlauf, D., Goto, Y., and Feil, R. (2004). Site-specific analysis of histone methylation and acetylation. Methods Mol. Biol. *287*, 99–120.

Verzi, M.P., Hatzis, P., Sulahian, R., Philips, J., Schuijers, J., Shin, H., Freed, E., Lynch, J.P., Dang, D.T., Brown, M., 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*, 15157–15162.

Verzi, M.P., Shin, H., San Roman, A.K., Liu, X.S., and Shivdasani, R.A. (2013). Intestinal master transcription factor CDX2 controls chromatin access for partner transcription factor binding. Mol. Cell. Biol. *33*, 281–292.

Whiffin, N., Dobbins, S.E., Hosking, F.J., Palles, C., Tenesa, A., Wang, Y., Farrington, S.M., Jones, A.M., Broderick, P., Campbell, H., et al. (2013). Deciphering the genetic architecture of low-penetrance susceptibility to colorectal cancer. Hum. Mol. Genet. *22*, 5075–5082.

Zuniga, A., Laurent, F., Lopez-Rios, J., Klasen, C., Matt, N., and Zeller, R. (2012). Conserved cis-regulatory regions in a large genomic landscape control SHH and BMP-regulated Gremlin1 expression in mouse limb buds. BMC Dev. Biol. *12*, 23.