The following full text is a publisher’s version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/131550

Please be advised that this information was generated on 2018-11-19 and may be subject to change.
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,2 Nina C. Hubner,4 Luan N. Nguyen,4 Jorge Zeron-Medina,2 Gareth Bond,1 Hendrik G. Stunnenberg,4 Jaime J. Carvajal,2 Jose Luis Gomez-Skarmeta,2 Simon Leedham,1 and Ian Tomlinson1,*

1Molecular and Population Genetics Laboratory, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK
2Centro Andaluz de Biología del Desarrollo, CSIC-Universidad Pablo de Olavide-Junta de Andalucía, Carretera de Utrera Km1, 41013 Sevilla, Spain
3Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, the Netherlands
4Department of Molecular Biology, Radboud Institute for Molecular Life Science, Geert Grootplein 26/28, 6525 GA Nijmegen, the Netherlands
5Ludwig 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
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

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 ApcMin 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 reexpression 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
SNPs in strong linkage disequilibrium (LD) with rs16969681 showed a 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 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.
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 upstream of GREM1 using the pGL3 promoter vector in SW408 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<sup>-/-</sup>;Apc<sup>Min</sup> mice compared to Grem1<sup>+/+</sup>;Apc<sup>Min</sup> 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.
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 \( \Delta \Delta CT \) 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)
Grem1+/+ rs16969681 is not highly conserved, we used constitutionally such as those resulting from the rs16969681 genotype, could phenotype.

had for the effects of rs16969681. We confirmed that these animals 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 (Gazzero 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+/-:ApcMin/+ animals had a >50% reduction in adenomas compared with their well-characterized Grem1+/-; ApcMin/+ 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+/-:ApcMin/+ 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 2Dii), 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 CRC cell line 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 ($p = 0.0066$; t test). See also Figure S3. Error bars represent SEM.
occurs in almost all colorectal adenomas but is lost in some CRCs (Quailtough 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).

**Experimental Procedures**

**CRC Patients**

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

**Formaldehyde-Assisted Identification of Regulatory Elements**

FAIRE was carried out using the method adapted from Giresi et al. (2007). Briefly, ~10^5 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 10^5 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; 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.

**Crosslinked ChIP**

Approximately 10^5 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 (GSE18481; Mokry et al., 2010) were downloaded and mined for signals in the GREM1 gene and surrounding region.

**Transgenic Mouse Crosses**

Grem1+/fox mice, a kind gift of E. Canalis, were crossed with Ptgk-Cre mice to inactivate one copy of Grem1 in the germine. 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
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).

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.
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


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.J.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 Greml1 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


Supplimentary 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.