Published in final edited form as:

Gastroenterology. 2020 January ; 158(1): 238–252. doi:10.1053/j.gastro.2019.09.023.

Mutated CEACAMs Disrupt Transforming Growth Factor beta Signaling and Alter the Intestinal Microbiome to Promote Colorectal Carcinogenesis

Shoujun Gu^{1,#}, Sobia Zaidi^{1,#}, Md. Imtaiyaz Hassan², Taj Mohammad², Tathiane M. Malta³, Houtan Noushmehr³, Bryan Nguyen⁴, Keith A. Crandall⁴, Jigisha Srivastav⁵, Vincent Obias⁶, Paul Lin⁶, Bao-Ngoc Nguyen¹, Michael Yao⁷, Ren Yao⁸, Charles Hadley King⁸, Raja Mazumder⁸, Bibhuti Mishra¹, Shuyun Rao^{1,*}, Lopa Mishra^{1,7,*}

¹Center for Translational Medicine, Department of Surgery, The George Washington University, Washington, DC, USA

²Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, Jamia Nagar, New Delhi, India

³Department of Neurosurgery, Henry Ford Health System, Detroit, MI, USA

⁴Computational Biology Institute and Department of Biostatistics & Bioinformatics, Milken Institute School of Public Health, The George Washington University, Washington, DC, USA

⁵University of Toledo College of Medicine, Toledo, OH, USA

⁶Department of Surgery, The George Washington University, Washington, DC, USA

⁷Department of Gastroenterology, Veterans Affairs Medical Center, Washington DC, USA

⁸Department of Biochemistry and Molecular Medicine, The George Washington University, Washington, DC, USA.

Abstract

BACKGROUND AND AIMS—We studied interactions among proteins of the carcinoembryonic antigen related cell adhesion molecule (CEACAM) family, which interact with microbes, and transforming growth factor beta (TGFB) signaling pathway, which is often altered in colorectal cancer cells. We investigated mechanisms by which CEACAM proteins inhibit TGFB signaling and alter the intestinal microbiome to promote colorectal carcinogenesis.

METHODS—We collected data on DNA sequences, mRNA expression levels, and patient survival times from 456 colorectal adenocarcinoma cases, and a separate set of 594 samples of colorectal adenocarcinomas, in the Cancer Genome Atlas. We performed shotgun metagenomic

*Correspondence: Lopa Mishra, M.D., Director, Center for Translational Medicine, Department of Surgery, Professor of Surgery, The George Washington University, Office Phone: +1 202-994-4629, lmishra@email.gwu.edu.

#Authors contributed equally

Author contributions

LM, SR- supervised the study with input from BM, VO, PL, BNN, MY; SG, SZ, TMM, HN, BN, KAC, RY, CHK, RM, JS-data analytics, manuscript preparation; SG-cellular experiments; SR-microbiome experiments; IH, TM, SZ-3D structure modeling.

Conflicts of interest

None

sequencing analyses of feces from wild-type mice and mice with defects in TGFB signaling (*Sptbn1*^{+/-} and *Smad4*^{+/-}/*Sptbn1*^{+/-}) to identify changes in microbiota composition before development of colon tumors. CEACAM protein and its mutants were overexpressed in SW480 and HCT116 colorectal cancer cell lines, which were analyzed by immunoblotting and in proliferation and colony formation assays.

RESULTS—In colorectal adenocarcinomas, high expression levels of genes encoding CEACAM proteins, especially *CEACAM5*, were associated with reduced survival times of patients. There was an inverse correlation between expression of CEACAM genes and expression of TGFB pathway genes (*TGFB1*, *TGFB2*, and *SMAD3*). In colorectal adenocarcinomas, we also found an inverse correlation between expression of genes in the TGFB signaling pathway and genes that regulate stem cell features of cells. We found mutations encoding L640I and A643T in the B3 domain of human *CEACAM5* in colorectal adenocarcinomas; structural studies indicated that these mutations would alter the interaction between CEACAM5 and TGFB1. Overexpression of these mutants in SW480 and HCT116 colorectal cancer cell lines increased their anchorage-independent growth and inhibited TGFB signaling to a greater extent than overexpression of wild-type CEACAM5, indicating that they are gain of function mutations. Compared with feces from wild-type mice, feces from mice with defects in TGFB signaling had increased abundance of bacterial species that have been associated with the development of colon tumors, including *Clostridium septicum*, and decreased amounts of beneficial bacteria such as *Bacteroides vulgatus* and *Parabacteroides distasonis*.

CONCLUSION—We found expression of CEACAMs and genes that regulate stem cell features of cells to be increased in colorectal adenocarcinomas and inversely correlated with expression of TGFB pathway genes. We found colorectal adenocarcinomas to express mutant forms of CEACAM5 that inhibit TGFB signaling and increase proliferation and colony formation. We propose that CEACAM proteins disrupt TGFB signaling, which alters the composition of the intestinal microbiome to promote colorectal carcinogenesis.

Keywords

tumor suppressor; signal transduction; TCGA; microbiome

Advanced colorectal cancer (CRC) is the third most commonly diagnosed cancer in the world, with 10–15% of CRC cases in the U.S. arising in men and women under 50 years of age.¹ This statistic is alarming, because screening, which has reduced the incidence of CRC, is generally not performed until after the age of 50.¹ Consequently, CRC occurring in persons under 50 remains undiagnosed until presentation at advanced stages. Several risk factors have been implicated in CRC: diets high in fat and processed foods, obesity, genetic or epigenetic alterations in critical drivers and pathways that control intestinal epithelial homeostasis, and altered function of the immune system.² The pathways associated with CRC include the TGFB, WNT, RAS-MAPK, PI3K, P53, and DNA mismatch-repair pathways.^{3–5}

Altered expression of genes or mutations in genes encoding components of the TGFB pathway occur in more than 80% of hypermutated CRCs.³ TGFB signaling is initiated by the binding of TGFB ligand to the serine-threonine kinase type II TGFB receptors

(TGFB β 2), which phosphorylate the type I receptors (TGFB β 1) and form a large ligand-receptor complex. Activated TGFB β 1 activates Smad-mediated transcriptional regulation and Smad-independent signaling, such as activation of the MAPK pathway mediated by ERK1 and ERK2 and PI3K-AKT signaling. Smad-mediated signaling occurs in response to TGFB β 1-mediated phosphorylation of receptor-regulated SMAD2 and SMAD3, enabling the formation of SMAD2-SMAD4 and SMAD3-SMAD4 complexes. The resulting SMAD complexes translocate into the nucleus and interact with other transcription factors in a cell-specific manner to regulate the transcription of a multitude of TGFB-responsive genes.⁶

Depending on the context, the TGFB pathway can suppress tumorigenesis or can promote metastasis.⁷⁻⁹ Current data support a role for TGFB signaling as tumor suppressor in early CRC.^{10, 11} In later disease, metastatic CRCs escape the tumor-suppressing effects of TGFB signaling, becoming resistant to TGFB-induced growth inhibition and responsive to the pro-tumorigenic activities of TGFB.¹² The importance of this pathway is also evident from the reduction in the activity of the TGFB pathway in many CRCs with a cancer stem cell signature.¹³ We have generated double heterozygous mice (*Smad4*^{+/-}/*Sptbn1*^{+/-}) which have impaired Smad-mediated TGFB signaling through a pathway involving Smad3 and Sptbn1 (also known as β 2-spectrin). TGFB deficient mice develop CRC,⁵ but not when housed in a germ-free environment.¹⁴ Thus, these mice represent useful models for assessing the role of TGFB signaling in CRC development and progression and investigating the role of the microbiota in this process.

Previously, we reported correlations between genetic alterations in TGFB pathway components, reduced TGFB signaling, and increased *CEACAM5* expression in 23% of colonic adenomas.⁴ Additionally, we determined that CEACAM5 binds and inhibits the TGFB receptor (TGFB β 1) to interfere with TGFB signaling.⁴ CEACAM5 is a biomarker for recurrent disease in CRC, a biomarker for pancreatic and thyroid cancers, and biomarker for and facilitator of metastatic liver cancer.¹⁵

The CEACAM family consists of both transmembrane and glycosylphosphatidylinositol (GPI)-linked membrane glycoproteins and are a subgroup of the immunoglobulin superfamily.¹⁶ CEACAM functions are diverse: they serve as microbial receptors for viruses and bacteria,^{17, 18} modulate gut mucosal immunity and homeostasis,¹⁹ and promote intercellular adhesion,¹⁵ T cell proliferation,²⁰ and neovascularization.²¹ In the gastrointestinal tract, CEACAM members serve as mucosal receptors that enable adherent microbes, including invasive *Escherichia coli* and *Helicobacter pylori*, to colonize gut mucosa, which leads to a persistent alteration in the gut microbiome and inflammation.^{22, 23}

Gut microorganisms participate in cancer development and drug resistance. Because our TGFB signaling-deficient mouse mutant spontaneously develops adenomas that progress to CRC and this does not occur in germ-free conditions, we hypothesized that the TGFB pathway, CEACAMs, and gut microbes may be connected through a regulatory network and contribute to CRC development. Here, we used *Smad4*^{+/-}/*Sptbn1*^{+/-} and *Sptbn1*^{+/-} mice to investigate this relationship in CRC. We analyzed interactions between CEACAMs, their mutations, and the TGFB pathway, and we examined differences in the gut microbiomes between wild-type mice and the TGFB pathway-deficient animals prior to the development

of CRC. We also used The Cancer Genome Atlas (TCGA) to explore the relationship between TGFB pathway activity and CEACAMs and patient survival. Our results indicated that there is an inverse relationship between the tumor-suppressing activity of TGFB signaling and several CEACAM-encoding genes in CRC patients. The mouse data indicated that impaired TGFB signaling results in gut microbiome alterations resembling those associated with CRC in humans.

Materials and Methods

Microbiome DNA extraction, shotgun metagenomic sequencing, and bioinformatic analysis

The DNA from fecal samples collected from wild-type mice (n=8) or TGFB signaling-deficient mice *Smad4*^{+/-}*Sptbn1*^{+/-} and *Sptbn1*^{+/-} (n=8) between 2–5 months old before CRC development, was extracted using the QIAmp PowerFecal DNA Extraction Kit (Qiagen, 12830–50). Double-stranded DNA (dsDNA) concentration was assessed by Qubit 3.0 dsDNA HS DNA kit (Thermo Fisher, NY, USA), and the quality was evaluated by the DS-11 spectrophotometer (DeNovix, Wilmington, DE, USA). Samples were prepared for Illumina sequencing following the manufacturer's protocol using the Nextera XT DNA Library Preparation Kit (Illumina, FC-131–1096). Paired-end sequencing was performed using a Mid Output v 2.5 (300 cycles) kit (20024905; Illumina) on a NextSeq 500 (Illumina, San Diego, CA), with dual indexing. Each dsDNA molecule was sequenced 150 bases from the end of each strand. Each strand also had an 8 base pair molecular barcode that was sequenced. Two analysis tools exclusive to the HIVE (High-performance Integrated Virtual Environment) platform (<https://hive.biochemistry.gwu.edu/dna.cgi?cmd=main>), CensuScope24 and HIVE-Hexagon25, were used to determine the microbial relative abundance. HIVE is a cloudbased platform for the storage and analysis of extra-large data, such as biomedical data, clinical data, next-generation sequencing (NGS) data, and mass spectrometry files. There are currently several HIVE instances in both the public and private domains. The U.S. Food and Drug Administration (FDA) is home to two separate HIVE installations, and the public domain of HIVE is hosted at The George Washington University (GW) where any scientist around the world can access the resources through the web-portal, once registered for an account, The CensuScope tool uses Filtered-nt, one of the most comprehensive reference databases available. Filtered-nt (<https://hive.biochemistry.gwu.edu/filterednt/about>) was generated from NCBI nonredundant nucleotide database (nt) by filtering all the taxonomy in the blacklist file. Filtered_NTv6.0 was used here.²⁶ After normalization, the relative abundance calculated in each sample was plotted using GraphPad Prism 7.0, and unpaired *t*-test was used for statistical analysis.

Cell culture and transfection

SW480, HCT116, and FET cells were cultured in complete culture medium: DMEM/F12 medium (Sigma-Aldrich, D5671) supplemented with 1% Streptomycin-Penicillin and 10% FBS (Sigma-Aldrich, F2442). Cells were transfected with hemagglutinin-tagged (HA)-CEACAM5 (HA-CEA), HA-CEA-L640I, or HA-CEA-A643T plasmids using XtremeGENE 9 (Sigma-Aldrich, 6365779001) according to the manufacturer's protocols. For

TGFB treatment, cells were incubated in serum-free DMEM/F12 medium for 18 hours then TGFB1 (Sigma-Aldrich, T1654) was added to a final concentration of 200 pM.

For the 5-FU cytotoxicity experiment, siRNA targeting SMAD4 or SPTBN1 (Dharmacon, Lafayette, CO) were transfected into HCT116 cells by Lipofectamine RNAiMAX (ThermoFisher, #13778150). After 24 hours, $1-2 \times 10^4$ HCT116 cells were plated in a 12-well plate, treated with vehicle control or 5-FU at a final concentration of 1 μ M, 10 μ M, or 100 μ M. Two days later, the viable cells based on trypan blue exclusion were counted.

Immunoblotting

Cells were lysed with lysis buffer (20 mM Tris, 100 mM NaCl, 0.5% Nonidet P-40, 0.5% Triton X-100, 1 mM EDTA). Protease/phosphatase inhibitor cocktail (Sigma-Aldrich, PPC1010) was added to the lysis buffer immediately before harvesting cells. Protein concentration was analyzed with DC Protein Assay Reagent (BioRad, 5000114), and equal amounts of protein were subjected to immunoblot analysis. Antibodies used were anti-HA (ThermoFisher, #71-5500), anti-p-Smad3 (Cell Signaling, #9520), anti-Smad3 (Cell Signaling, #9523), and anti-CEACAM5 (ThermoFisher, #MA5-14675).

Sphere formation assay

Before seeding the cells, 2 ml/well of bottom-layer medium (0.5% agar in complete medium) were added to 6-well cell culture plates (Sarstedt, 83.392). After bottom-layer medium solidified, 1 ml of a 1:1 mixture of cells resuspended in complete medium and top-layer medium (0.75% agar in complete medium) was added on top of the solid bottom layer. Cells were incubated for 2 weeks before analysis. A colony was defined as an aggregate of >40 cells. Colonies were counted under the inverted microscope.

Data processing and analysis

For the gene alteration assay, the Colorectal Adenocarcinoma (TCGA, PanCancer Atlas) dataset from cBioportal (<http://www.cbioportal.org>) was used.

For mRNA abundance analysis, a default mRNA expression Z score of (± 2.0) was applied. For survival and expression analysis, 456 cases in the TCGA-COAD (colorectal adenocarcinoma) project was used. RNASeq counts were combined and normalized using median of ratios method.²⁷ The average normalized counts from ‘Solid Tissue Normal’ samples were used as a threshold to filter the samples. Only samples with counts above the threshold were used for survival analysis. Survival plots were made by comparing the top 30% filtered samples (high expression group) and bottom 30% filtered samples (low expression group). The p-values were calculated by log-rank test.

For TGFB pathway activity analysis, scores were obtained from the previous study.⁹ *CEACAM5* mutation analysis was performed with five datasets in the category ‘Colorectal adenocarcinoma’ from cBioportal with the default settings in cBioPortal.^{9, 13}

To calculate a stemness index (si) based on mRNA expression (mRNA_{si}), we applied the same approach described previously.¹³ Briefly, we built a predictive model using one-class logistic regression (OCLR) on the pluripotent stem cell samples (embryonic stem cell and

induced pluripotent stem cell) from the Progenitor Cell Biology Consortium (PCBC) dataset. The training matrix based on mRNA expression-based signatures, contained 12,945 mRNA expression values measured across all available PCBC samples. Having validated the signature using cross-validation and external stem cell data, we then applied the signature to determine a Spearman correlation for the signature in the TCGA PanCancer cohort, including COAD and READ (rectal adenocarcinoma). The mRNAsi was subsequently mapped to the [0,1] range using a linear transformation that subtracted the minimum and divided by the maximum.¹³

Protein structure

Atomic coordinates of the human TGFBR1 were obtained from Protein Data Bank (PDB ID: 1B6C).²⁸ The model of the B3 domain of CEACAM5 was based on V-type, I-type and C2-type immunoglobulin folds (PDB ID: 1E07).²⁹ This starting structure represented 94% query coverage and 100% sequence identity to the B3 domain model that we needed for the docking studies. TGFBR1 from PDB and the modeled structures of CEACAM5 and its mutants (A643T and L640I) were used for docking studies with PatchDock webserver. PyMOL was used to visualize the docked proteins and analyze the interactions between CEACAM5 and TGFBR1.

Results

TGFB deficient mice that develop CRC have altered distributions of gut microbiome species

TGFB signaling-deficient mice (*Smad4*^{+/-}*Sptbn1*^{+/-}) spontaneously develop adenomas and CRC (Figure 1A). Increasing data reveal a critical role for gut microorganisms in cancer development and therapy outcome,^{30–34} thus we performed shotgun metagenomics sequencing using fecal samples from wild-type mice and age-matched TGFB signaling-deficient mice (Figure 1B–D) to identify changes in gut microbiota composition (Phylum, Class, Family, Genus, Species) in mice with impaired TGFB signaling throughout the body, including in the gastrointestinal tract and immune system. By analyzing the abundance of known human CRC-associated bacteria species, we found *Clostridium septicum*³⁵ (Figure 1C) is a significantly increased bacterial species in TGFB signaling-deficient mice. Consistent with enriched microbiome data from human CRC, we also observed a trend toward increased amounts of *Fusobacterium nucleatum*³⁰ and *Enterococcus faecalis*³⁶ in the TGFB signaling-deficient mice that develop CRC (Figure 1C). Additionally, two species of commensal gut microbes associated with a healthy microbiome, *Bacteroides vulgatus*³⁷, and *Parabacteroides distasonis*³⁸, were significantly decreased in TGFB signaling-deficient mice with CRC (Figure 1D).

5-Fluorouracil (5-FU) treatment significantly reduces the abundance of *B. vulgatus* in mice,³⁹ and 5-FU is commonly used to treat CRC. Therefore, we evaluated the effect of 5-FU on HCT116 cells in which we knocked down either SPTBN1 or both SPTBN1 and SMAD4. We found these cells were more susceptible to the cytotoxic effects of 5-FU (Supplementary Figure 1), which is consistent with previous work on the effect of DNA cross-linking agents on *Sptbn1* deficient cells.⁴⁰

An inverse relationship between the TGFB pathway and CEACAMs extends beyond CEACAM5 in human colon cancer

Both TGFB signaling and CEACAMs have been associated with various types of cancer. Additionally, CEACAM5 interferes with TGFB signaling.^{4, 10} Furthermore, CEACAMs are receptors for host-specific bacteria, and TGFB signaling regulates both immune responses and tissue homeostasis. CEACAM1, 5, and 6 are present in the gastrointestinal tract, function as microbial receptors, and have been associated with CRC.^{15, 16} *CEACAM4, 8, 16 and 19* are not normally expressed in the gastrointestinal tract and have not yet been associated with CRC. This set of *CEACAM* genes thus includes those with well-established relevance in the gut microbial system and those without any connection to this system. We investigated the relationship between TGFB signaling, expression of each member of this set of CEACAM-encoding genes, and patient outcome using the human colon cancer dataset from TCGA.

To evaluate the relationship between the CEACAM family members and patient survival, we sorted the 456 patient cases from TCGA colon adenocarcinoma (COAD) with survival data into those with high expression of individual *CEACAM* genes and those with low expression of these genes. Note that only those CEACAM members with sufficient survival data are shown, thus there are no survival curves presented for CEACAM1. We observed a similar relationship with several CEACAM-encoding genes: Patients with high expression of *CEACAM6, 8, or 16*, but not *CEACAM4 or 19*, had reduced long-term survival compared with the patients with low expression of these CEACAM-encoding genes (Figure 2A). Among them, *CEACAM5* (also known as *CEA*) overexpression was associated with poor overall survival (log-rank $p=0.0242$), which is consistent with the ability of *CEACAM5* to inhibit tumor-suppressive TGFB signaling⁴ and from the increased expression of *CEACAM5* in colon adenoma samples⁴. We also noted that the effects on survival were most evident at the later times, which is consistent with the association of CEACAM5 with metastatic colon cancer.^{41, 42}

To investigate alterations in the CEACAM-encoding genes or their expression in colon adenocarcinomas, we analyzed another set of 594 samples of colorectal adenocarcinoma samples (TCGA, PanCancer Atlas) using the cBioportal database. From the 7 CEACAM genes we tested, we selected those with alterations in >20% of cases for evaluation (Figure 2B). We compared the alterations in this set of CEACAM genes with alterations in TGFB-Smad3 pathway genes, *TGFBR1*, *TGFBR2*, and *SMAD3*. Whereas most of the CEACAM alterations would increase the abundance of the encoded protein; the TGFB signaling component alterations would decrease the activity of this pathway. The alterations in the CEACAMs and the alterations in three TGFB pathway components were not mutually exclusive, indicating that each could contribute to cancer progression.

Analysis of the RNASeq data from 456 patient cases in TCGA-COAD data showed that cases with high TGFB pathway activity, as indicated by a score based on *SMAD3* (TGFB signal transducer) and *SERPINE1* (SMAD3 target gene) expression, had low expression of each CEACAM-encoding gene except *CEACAM21* and *CEACAM4* (Figure 2C). Thus, the inverse relationship between the TGFB pathway and CEACAMs extends beyond *CEACAM5*. This inverse relationship between TGFB pathway activity and CEACAMs may

contribute to the changes in the gut microbiome that we observed in the TGFB signaling-deficient mice.

Previously, we observed CEACAM5 overexpression in 23% of adenoma samples with a concomitant reduction in TGFB signaling members.⁴ Here, we confirm these findings (Supplementary Figure 2A) and further evaluated the effect of crosstalk between CEACAM5 with TGFBR1 and TGFBR2 on survival of colorectal cancer patients. We observed that CEACAM5 amplification or overexpression, when combined with mutations or loss of expression of TGFBR1 and TGFBR2, the survival of patients is far lower (log-rank $p=9.67e-3$) (Supplementary Figure 2B). We extended our study to other two microbe-sensing intestinal epithelial CEACAMs (CEACAM1 and CEACAM6) in adenoma samples ($n = 20$). When compared with normal tissues, <10% of adenoma samples had high *CEACAM1* or *CEACAM6* expression together with low *TGFBR2*, *SMAD4*, or *SPTBN1* expression (Figure 2D). Consistent with our previous studies, these data suggested that a defect in TGFB signaling and increase in CEACAM5 occurs at an early stage of tumor development.

TGFB pathway activity is negatively correlated with stemness in human colon cancer

Previous studies have indicated a link between stem cell transformation and colon cancer with ~50% of CRC arising from foci of aberrant cells in the crypt or crypt (stem) cells that acquire oncogenic mutations.^{43, 44} TGFB signaling is important for tissue homeostasis and is a regulator of stem cell behavior.¹³ Therefore, we calculated a stemness index (mRNAsi) based on gene expression using mRNA data for genes enriched in oncogenic pathways.¹³ To generate this index, we used Molecular Signatures Database (MSigDB) and then we stratified gastrointestinal cancers on the basis of this stemness index and plotted each cancer in each stemness group according to the activity of the TGFB pathway using a pathway activity score based on mRNA expression of the 43 genes⁹ (Figure 3A). Gene set enrichment analysis (GSEA) for cancer-associated pathways also identified low TGFB signaling in two types of CRC: COAD (colon adenocarcinoma) and READ (rectal adenocarcinoma) (Supplementary Figure 3). Consistent with published data¹³, COAD had the highest stem cell index (Figure 3B). Among the various molecular subtypes of COAD, tumors with hypermutations with elevated single nucleotide variants (HM-SNV) had the highest stemness index (Figure 3C). The stemness analysis revealed an inverse correlation between TGFB signaling and stemness, and both GSEA and stemness analysis provided further evidence for a tumor-suppressing role of TGFB signaling pathway in CRC.

Collectively, the CEACAM-TGFB pathway relationship data and the stemness-TGFB pathway relationship data suggested multiple mechanisms for tumor-suppressing effects of TGFB signaling. Impaired TGFB signaling may enable intestinal stem cells to grow abnormally, as well as alter CEACAM-encoding gene expression such that the gut microbiome is altered, creating a pro-inflammatory condition that promotes the progression to cancer (Figure 3D).

Modeling and molecular docking predicts that mutants in CEACAM5 B3 domain, which is critical for TGFBR1 binding, alter the TGFBR1-CEACAM5 interaction

For the remainder of the study, we focused on CEACAM5 because this CEACAM had the greatest association with poor survival (Figure 2A), an inverse relationship with TGFB pathway genes (Figure 2C),⁴ inhibits TGFB signaling,¹⁰ and serves as a microbial receptor in the gastrointestinal tract. Thus, CEACAM5 could connect impaired TGFB signaling with changes in microbial composition.

The B3 domain of CEACAM5, spanning residues 577 to 682, binds to TGFBR1 and this interaction blocks TGFB-induced SMAD3 activity.^{4,10} CEACAM5 has a variable (V)-type, I-type, and six C2-type Ig domains (Figure 4). The V-type domain is referred to as the N domain and is followed by repeated pairs of C2-type domains (A1, B1, A2, B2, A3, B3).

To investigate the mechanism by which the CEACAM5 B3 domain interacted with TGFBR1 and inhibits TGFB signaling, we mapped the interaction site on the B3 domain. We generated 5 truncated versions of the B3 domain fused to glutathione-S-transferase (GST): CEA-B3F1 (residues 577–662), CEA-B3F2 (577–648), CEA-B3F3 (577–634), CEA-B3F4 (577–620), and CEA-BF5 (577–604) (Supplementary Figure 4). Using these GST-tagged B3 truncated peptides, the positive control GST-tagged full-length B3 domain (CEA-B3), and the negative control GST-tagged N-terminal region of CEACAM5 (CEA-N, amino acids 1–144), we performed pull-down assays to test for TGFBR1 interactions. Only the full B3 domain and the truncated CEA-B3 F1 fragment (residues 577–662) bound to TGFBR1, suggesting that the residues between 648–662 are critical for the interaction (Supplementary Figure 4).

Mining the colorectal adenocarcinoma datasets in cBioportal revealed two CEACAM5 mutations (L640I and A643T) located within the B3 domain and adjacent to the region required for interactions with TGFBR1 (Figure 4A). Despite low frequency of CEACAM5 mutations in the CRC samples, we observed a significant co-occurrence with mutations in *MLH1* and *MSH2*, encoding mismatch repair proteins (Supplementary Figure 5). This association suggested that these mutations may be prevalent in tumors with microsatellite instability (MSI).

To predict the function of these L640I and A643T mutants, we generated a model of the structure of the B3 domain (Figure 4B) based on a structural model of CEACAM5 from solution scattering data.²⁹ The starting model included all but residues from S677-G682 of the B3 domain. We generated a complete 106-residue model of the CEACAM5 B3 domain through remodeling with 99.9% confidence and achieved a Dope (Discrete optimized protein energy) score of –6850.52 (Figure 4B–C). Ramachandran analysis of the model showed that 94 (90.4%) residues were inside the favored region (Figure 4C, hardsphere), 5 (4.8%) residues were in an additional allowed region (Figure 4C, overlap), and 5 (4.8%) residues were in outlier region (outside) (Figure 4C). Thus, we used this to generate models of the B3 domain with each of the hotspot mutations L640I and A643T.

The overall structures of the wild-type CEACAM5 B3 domain model and the B3 domain with either of the cancer-associated mutants L640I and A643T were similar, and the B3

domain folding was preserved in the mutants (Figure 4D). The mutations were predicted to reduce CEACAM5 B3 domain stability, according to estimated pseudo G values. L640I reduced the stability of the domain by -0.13 kcal/mol, and A643T reduced the stability by -0.71 kcal/mol. The models predicted that mutated residues have slightly decreased accessibility compared with the wild-type residues (Figure 4E): from 55.9% to 54.4% for L640I and from 7.1% to 6.9% for A643T. These changes correspond to small differences in the modeled depth of the mutated residue within the B3 domain: from 3.5Å to 3.7Å for L640I and from 4.8Å to 5.1Å for A643T (Figure 4E). The models also predicted changes in the electrostatic potential in the B3 domain with each of the mutants (Figure 4F), consequently, the region where the B3 domain interacted with TGFBR1 would be affected. Thus, the molecular modeling predicted that these hotspot mutations cause potentially functionally important alterations in the CEACAM5 B3 domain.

We performed molecular docking simulations to predict the orientation of the CEACAM5 B3 domain in the binding pocket of the human TGFBR1. The simulations indicated that CEACAM5 has several close interaction sites with residues in the binding pocket of TGFBR1 (Figure 4G). To predict how the mutations affected the interaction between the B3 domain and TGFBR1, we selected matching docked conformations of the wild-type B3 domain and those of the mutants. The docked CEACAM5 mutants had an altered binding orientation and made different interactions with the binding pocket of TGFBR1 (Figure 4H–I). Whereas the wild-type CEACAM5 B3 domain fit into the binding cavity of TGFBR1 and formed 10 hydrogen bonds with 9 residues, the mutants had more hydrogen bonding interactions. A643T had 14 hydrogen bonds with 12 residues of TGFBR1, and L640I had 11 hydrogen bonds with 8 residues of TGFBR1. In both mutants, the hydrogen bond between Leu640 of the B3 domain and Lys489 of TGFBR1 is lost, consistent with the altered orientation of the mutants in the binding pocket. Collectively, the modeling and molecular docking data predicted that the hotspot CEACAM5 mutations would alter the interaction between CEACAM5 and TGFBR1.

Overexpression of CEACAM5 hotspot mutants enhances colon cancer cell proliferation

To investigate the functional consequences of the CEACAM5 mutations, we identified colon cancer cell lines with low amounts of endogenous CEACAM5 (Figure 5A). We overexpressed wild-type CEACAM5 and each of the mutants (CEA-L640I, CEA-A643T) in two of the cell lines with low endogenous CEACAM5, SW480 [a microsatellite stable (MSS) cell line] and HCT116 (MSI cell line), and evaluated cell proliferation and anchorage-independent cell growth by soft agar assay. Overexpression of wild-type CEACAM5 significantly enhanced cell proliferation (Figure 5B) and colony formation (Figure 5C–D). The two CEACAM5 mutants, CEA-L640I and CEA-A643T, exhibited gain of function effects, stimulating cell proliferation and anchorage-independent cell growth more than that of wild-type CEACAM5 (Figure 5B–D).

To determine if the mutants affected tumorigenic characteristics in cells with abundant CEACAM5, we transfected the mutants individually into FET cells, which have high amounts of endogenous CEACAM5 (Figure 5A). Overexpression of either of the mutants in FET cells did not inhibit colony formation, ruling out the possibility that these mutants

function as dominant-negative variants (Figure 5C). Consistent with CEACAM5 promoting tumorigenic phenotypes, non-transfected FET cells formed larger colonies than did non-transfected SW480 or HCT116 cells (Figure 5C–D), and knockdown of CEACAM5 reduced the size of colonies formed by FET cells (Figure 5E).

To explore whether these two CEACAM5 mutants inhibited TGFB signaling, we expressed each mutant in HCT116 cells and assessed TGFB–stimulated SMAD3 activation. Consistent with our previous study,^{4, 10} overexpression of wild-type CEACAM5 inhibited TGFB–induced accumulation of phosphorylated SMAD3, and overexpression of CEA-L640I or CEA-A643T inhibited TGFB signaling to a greater extent than overexpression of wild-type CEACAM5 (Figure 5F).

In summary, we found that the cancer-associated CEACAM5 B3 domain mutants exhibit enhanced tumorigenic activity compared to wild-type CEACAM5 and these mutants inhibited TGFB signaling. Furthermore, our results indicated that the colony-forming activity of colon cancer cells depends on CEACAM5 with high amounts of this protein promoting colony formation.

Discussion

CRC is the third most common cause of cancer mortality worldwide. Microorganisms in the gastrointestinal tract play a critical role in cancer development and drug response.^{33, 34} TGFB signaling is important for regulating both intestinal homeostasis and the immune system. Through the crosstalk with NF- κ B signaling and IL6-STAT3 signaling, the TGFB pathway can alter the immune response and microbiome composition.^{45,46} Thus, we examined how impaired TGFB signaling affected the gut microbiome in TGFB signaling-deficient mice and found that the abundance of *Clostridium septicum* increased and *Bacteroides vulgatus* and *Parabacteroides distasonis* decreased. These changes may contribute to the spontaneous development of CRC in the mutant mice and may represent a mechanism by which impaired TGFB signaling contributes to CRC in humans.

Clostridium produces metabolites, such as butyrate, which increase TGFB production by epithelial cells.⁴⁷ In the context of canonical TGFB signaling deficiency, such bacterial metabolites, along with the feedback-mediated increase in the expression of genes encoding the TGFB ligands would stimulate noncanonical TGFB signaling, such as stress-activated protein kinases, to promote carcinogenesis or metastases. In addition, TGFB regulates immune cells, thus, increased production of these ligands could help the cancer cells escape from immune surveillance.⁴⁸

The reduction in *B. vulgatus* and *P. distasonis* may result in loss of protective benefits of these microbes. *P. distasonis* produces succinate and secondary bile acids, which alleviate obesity and metabolic dysfunction.³⁸ Metabolism is emerging as a key regulator of immune function and thus this reduction in *P. distasonis* could influence immune responses in such a way to promote cancer progression. *B. vulgatus* inhibits atherosclerosis and decreases gut microbial lipopolysaccharide production, which suppresses pro-inflammatory immune

responses.³⁷ Because inflammation is a driver of cancer progression, loss of this species could contribute to a higher risk of cancer.

An association between *B. vulgatus* and the chemotherapeutic 5-FU exists. *B. vulgatus* can hydrolyze anti-viral agent, Sorivudine, to a metabolite to inhibit 5-FU degradation and thereby increases 5-FU toxicity.⁴⁹ In mice, 5-FU treatment significantly reduces the abundance of *B. vulgatus*.³⁹ 5-FU is a common therapy for CRC, but resistance frequently develops. Here, we extended our previous work^{40, 50} showing that impairment of TGFB signaling enhances the toxicity of 5-FU in CRC cells in culture. Future studies will address if there is a role for the altered microbiome in sensitivity to 5-FU or if there is a dosing regimen involving TGFB inhibitors and 5-FU that requires changes in the microbiome for an optimal 5-FU response. Thus, our data provide insight into how to potentially influence the microbiome to improve drug response.

TGFB signaling-deficient mice do not develop tumors in a germ-free environment,¹⁴ providing additional evidence for the importance of the microbiome in CRC. In *Smad3*^{-/-} mice, *Helicobacter* infection induces CRC.⁵¹ The mechanism(s) connecting impaired TGFB signaling, CRC risk, and the microbiome remains unknown. Gene alterations that impair TGFB signaling occur in 40% of CRC and reduced TGFB signaling activity is associated with many cancers, including CRC.⁹ Another potential mechanism for the impairment of TGFB signaling in the context of CRC is through CEACAM5. CEACAM5, along with CEACAM1 and 6, are present in the gastrointestinal epithelial cells and function as microbial sensors.^{37, 38, 52} CEACAMs with different specificity to different pathogens can be induced by pathogens and can regulate immune response directly or indirectly through TGFB signaling.^{16, 32, 47, 53} Previously, we found that CEACAM5 inhibits canonical TGFB signaling through SMAD3.^{4, 10} Here, by investigating the relationship between *CEACAM* genes expression levels and CRC patient survival, we found that, although high expression of several CEACAM genes correlated with reduced survival, high *CEACAM5* expression had the strongest relationship with poor survival time. Furthermore, increased CEACAM5 and reduced abundance of TGFB signaling components occur in adenoma, suggesting that these events are early contributors to CRC development. We identified 2 new CEACAM5 mutations that were predicted to have altered binding with TGFBR1 and experimentally showed that these mutants had enhanced tumorigenic activity and, like wild-type CEACAM5, inhibited TGFB signaling. Our data provide a potential mechanism to connect the changes in the gut microbial populations observed in the TGFB signaling-deficient mice with CRC, through CEACAM5-mediated inhibition of TGFB signaling, leading to altered epithelial or stem cell behavior and immune system function, as well as direct effects on the microbial population through its activity as a microbe receptor. Increased CEACAM5 abundance may alter the gut microbiome through its function as a microbial receptor and impairs TGFB signaling to further influence immune responses, alters the microbiome, and promotes tumorigenesis. Future studies are needed to evaluate how the mutations in CEACAM5 influence the microbe-binding function of this protein and how the presence of microbes affects the function of CEACAM5 and its mutants as an inhibitor of TGFB signaling.

The altered gut microbiome, the inhibition of TGFB signaling by CEACAM5, and the observation that TGFB signaling-deficient cells are sensitive to chemotherapeutic drugs may offer new therapeutic opportunities to inhibit CRC by co-targeting CEACAM5 and the microbiome. Indeed, CEACAM5 is not only an FDA-approved CRC biomarker, but anti-cancer vaccines to immunize against CEACAM5 have been tested in advanced CRC but not for prevention.⁴¹ Additionally, drugs targeting the B3 domain of CEACAM5 may represent a possible therapeutic strategy for CRC treatment. Finding enhanced sensitivity to 5-FU in the context of reduced TGFB signaling suggests that subpopulations of CRC patients, such as MSS CRC patients with *SMAD4* loss (18q deletion)⁵⁴ or MSI CRC patients with inactivating mutations in *TGFBR2*,⁵⁵ may benefit from DNA crosslinking chemotherapy, although our data indicate that such therapy may need to be combined with treatments to maintain a healthy microbiome composition. However, some MSI tumors are non-responsive to 5-FU and using genomic data is key to defining the appropriate patient populations. We plan to explore genetic alterations in TGFB signaling members in specific CRC subgroups that are chemoresistant or chemosensitive, evaluate if CRC-relevant *CEACAM* genes are regulated by TGFB signaling to define the mechanism for the inverse relationship between *CEACAM* gene expression and TGFB pathway activity, and determine the contribution of altered immune responses in the context of increased CEACAM5 abundance, and establish the molecular sequence of events related to CEACAM5 abundance changes, TGFB pathway activity impairment, and microbiome alterations in the initiation and progression of CRC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Nancy R. Gough, Department of Surgery, The George Washington University, for critical reading and editing of the manuscript; Wilma Jogunoori for assistance with the project; Kazufumi Ohshiro for sharing ideas. We thank the sequencing service provided by Castle Raley in the Genomic Core at the George Washington University.

Editing assistance provided by Nancy R. Gough, Department of Surgery, The George Washington University

Grant support/Funding: This work was supported by NIH grants R01 AA023146 (L. Mishra), NIH R01 CA236591 (L. Mishra), NIH U01 CA230690 (L. Mishra), VA Merit I01BX003732 (L. Mishra) and GW CTR (L. Mishra)

Abbreviations

CEACAM	carcinoembryonic antigen-related cell adhesion molecules
CRC	colorectal cancer
dsDNA	doublestranded DNA
5-FU	5-fluorouracil
GST	glutathione-S-transferase
HA-CEA	hemagglutinin-tagged-CEACAM5

mRNA	messenger RNA
mRNAsi	mRNA-based stemness index
MSI	microsatellite instability
PDB	Protein Data Bank
TCGA	The Cancer Genome Atlas
TCGA-COAD	The Cancer Genome Atlas-Colorectal Adenocarcinoma
TGFB	transforming growth factor b
TGFBR	transforming growth factor b receptor

REFERENCES

1. Siegel RL, Fedewa SA, Anderson WF, et al. Colorectal Cancer Incidence Patterns in the United States, 1974–2013. *J Natl Cancer Inst* 2017;109.
2. Liu PH, Wu K, Ng K, et al. Association of Obesity With Risk of Early-Onset Colorectal Cancer Among Women. *JAMA Oncol* 2019;5:37–44. [PubMed: 30326010]
3. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;487:330–7. [PubMed: 22810696]
4. Chen J, Raju GS, Jogunoori W, et al. Mutational Profiles Reveal an Aberrant TGF-beta-CEA Regulated Pathway in Colon Adenomas. *PLoS One* 2016;11:e0153933. [PubMed: 27100181]
5. Tang Y, Katuri V, Srinivasan R, et al. Transforming growth factor-beta suppresses nonmetastatic colon cancer through Smad4 and adaptor protein ELF at an early stage of tumorigenesis. *Cancer Res* 2005;65:4228–37. [PubMed: 15899814]
6. Massague J TGFbeta signalling in context. *Nat Rev Mol Cell Biol* 2012;13:616–30. [PubMed: 2292590]
7. Akhurst RJ. Targeting TGF-beta Signaling for Therapeutic Gain. *Cold Spring Harb Perspect Biol* 2017;9.
8. Morikawa M, Derynck R, Miyazono K. TGF-beta and the TGF-beta Family: Context-Dependent Roles in Cell and Tissue Physiology. *Cold Spring Harb Perspect Biol* 2016;8.
9. Korkut A, Zaidi S, Kanchi RS, et al. A Pan-Cancer Analysis Reveals High-Frequency Genetic Alterations in Mediators of Signaling by the TGF-beta Superfamily. *Cell Syst* 2018;7:422–437 e7. [PubMed: 30268436]
10. Li Y, Cao H, Jiao Z, et al. Carcinoembryonic antigen interacts with TGF- β receptor and inhibits TGF- β signaling in colorectal cancers. *Cancer Res* 2010;70:8159–68. [PubMed: 20889724]
11. Okugawa Y, Grady WM, Goel A. Epigenetic Alterations in Colorectal Cancer: Emerging Biomarkers. *Gastroenterology* 2015;149:1204–1225 e12. [PubMed: 26216839]
12. Staudacher JJ, Bauer J, Jana A, et al. Activin signaling is an essential component of the TGF-beta induced pro-metastatic phenotype in colorectal cancer. *Sci Rep* 2017;7:5569. [PubMed: 28717230]
13. Malta TM, Sokolov A, Gentles AJ, et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* 2018;173:338–354 e15. [PubMed: 29625051]
14. Engle SJ, Ormsby I, Pawlowski S, et al. Elimination of colon cancer in germ-free transforming growth factor beta 1-deficient mice. *Cancer Res* 2002;62:6362–6. [PubMed: 12438215]
15. Beauchemin N, Arabzadeh A. Carcinoembryonic antigen-related cell adhesion molecules (CEACAMs) in cancer progression and metastasis. *Cancer Metastasis Rev* 2013;32:643–71. [PubMed: 23903773]

16. Tchoupa AK, Schuhmacher T, Hauck CR. Signaling by epithelial members of the CEACAM family - mucosal docking sites for pathogenic bacteria. *Cell Commun Signal* 2014;12:27. [PubMed: 24735478]
17. Islam EA, Anipindi VC, Francis I, et al. Specific Binding to Differentially Expressed Human Carcinoembryonic Antigen-Related Cell Adhesion Molecules Determines the Outcome of *Neisseria gonorrhoeae* Infections along the Female Reproductive Tract. *Infect Immun* 2018;86.
18. Muenzner P, Rohde M, Kneitz S, et al. CEACAM engagement by human pathogens enhances cell adhesion and counteracts bacteria-induced detachment of epithelial cells. *J Cell Biol* 2005;170:825–36. [PubMed: 16115956]
19. Roda G, Jianyu X, Park MS, et al. Characterizing CEACAM5 interaction with CD8alpha and CD1d in intestinal homeostasis. *Mucosal Immunol* 2014;7:615–24. [PubMed: 24104458]
20. Khairnar V, Duhan V, Patil AM, et al. CEACAM1 promotes CD8(+) T cell responses and improves control of a chronic viral infection. *Nat Commun* 2018;9:2561. [PubMed: 29967450]
21. Kuespert K, Pils S, Hauck CR. CEACAMs: their role in physiology and pathophysiology. *Curr Opin Cell Biol* 2006;18:565–71. [PubMed: 16919437]
22. Martinez-Medina M, Denizot J, Dreux N, et al. Western diet induces dysbiosis with increased *E. coli* in CEABAC10 mice, alters host barrier function favouring AIEC colonisation. *Gut* 2014;63:116–24. [PubMed: 23598352]
23. Barnich N, Carvalho FA, Glasser AL, et al. CEACAM6 acts as a receptor for adherent-invasive *E. coli*, supporting ileal mucosa colonization in Crohn disease. *J Clin Invest* 2007;117:1566–74. [PubMed: 17525800]
24. Shamsaddini A, Pan Y, Johnson WE, et al. Census-based rapid and accurate metagenome taxonomic profiling. *BMC Genomics* 2014;15:918. [PubMed: 25336203]
25. Santana-Quintero L, Dingerdissen H, Thierry-Mieg J, et al. HIVE-hexagon: high-performance, parallelized sequence alignment for next-generation sequencing data analysis. *PLoS One* 2014;9:e99033. [PubMed: 24918764]
26. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 2016;44:D733–45. [PubMed: 26553804]
27. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550. [PubMed: 25516281]
28. Huse M, Chen YG, Massague J, et al. Crystal structure of the cytoplasmic domain of the type I TGF beta receptor in complex with FKBP12. *Cell* 1999;96:425–36. [PubMed: 10025408]
29. Boehm MK, Perkins SJ. Structural models for carcinoembryonic antigen and its complex with the single-chain Fv antibody molecule MFE23. *FEBS Lett* 2000;475:11–6. [PubMed: 10854848]
30. Yu T, Guo F, Yu Y, et al. *Fusobacterium nucleatum* Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. *Cell* 2017;170:548–563 e16. [PubMed: 28753429]
31. Gopalakrishnan V, Spencer CN, Nezi L, et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* 2018;359:97–103. [PubMed: 29097493]
32. Gray-Owen SD, Blumberg RS. CEACAM1: contact-dependent control of immunity. *Nat Rev Immunol* 2006;6:433–46. [PubMed: 16724098]
33. Dejea CM, Fathi P, Craig JM, et al. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* 2018;359:592–597. [PubMed: 29420293]
34. He Z, Gharaibeh RZ, Newsome RC, et al. *Campylobacter jejuni* promotes colorectal tumorigenesis through the action of cytolethal distending toxin. *Gut* 2019;68:289–300. [PubMed: 30377189]
35. Mizrahi DJ, Halpern EJ. *Clostridium septicum* aortitis and colon carcinoma. *J Cardiovasc Comput Tomogr* 2016;10:258–60. [PubMed: 26792077]
36. de Almeida CV, Taddei A, Amedei A. The controversial role of *Enterococcus faecalis* in colorectal cancer. *Therap Adv Gastroenterol* 2018;11:1756284818783606.
37. Yoshida N, Emoto T, Yamashita T, et al. *Bacteroides vulgatus* and *Bacteroides dorei* Reduce Gut Microbial Lipopolysaccharide Production and Inhibit Atherosclerosis. *Circulation* 2018;138:2486–2498. [PubMed: 30571343]

38. Wang K, Liao M, Zhou N, et al. Parabacteroides distasonis Alleviates Obesity and Metabolic Dysfunctions via Production of Succinate and Secondary Bile Acids. *Cell Rep* 2019;26:222–235 5. [PubMed: 30605678]
39. Yuan L, Zhang S, Li H, et al. The influence of gut microbiota dysbiosis to the efficacy of 5-Fluorouracil treatment on colorectal cancer. *Biomed Pharmacother* 2018;108:184–193. [PubMed: 30219675]
40. Chen J, Shukla V, Farci P, et al. Loss of the transforming growth factor-beta effector beta2-Spectrin promotes genomic instability. *Hepatology* 2017;65:678–693. [PubMed: 28114741]
41. Dotan E, Cohen SJ, Starodub AN, et al. Phase I/II Trial of Labetuzumab Govitecan (Anti-CEACAM5/SN-38 Antibody-Drug Conjugate) in Patients With Refractory or Relapsing Metastatic Colorectal Cancer. *J Clin Oncol* 2017;35:3338–3346. [PubMed: 28817371]
42. Blumenthal RD, Hansen HJ, Goldenberg DM. Inhibition of adhesion, invasion, and metastasis by antibodies targeting CEACAM6 (NCA-90) and CEACAM5 (Carcinoembryonic Antigen). *Cancer Res* 2005;65:8809–17. [PubMed: 16204051]
43. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* 2015;347:78–81. [PubMed: 25554788]
44. Takayama T, Katsuki S, Takahashi Y, et al. Aberrant crypt foci of the colon as precursors of adenoma and cancer. *N Engl J Med* 1998;339:1277–84. [PubMed: 9791143]
45. Tang LY, Heller M, Meng Z, et al. Transforming Growth Factor-beta (TGF-beta) Directly Activates the JAK1-STAT3 Axis to Induce Hepatic Fibrosis in Coordination with the SMAD Pathway. *J Biol Chem* 2017;292:4302–4312. [PubMed: 28154170]
46. Mitra A, Yan J, Xia X, et al. IL6-mediated inflammatory loop reprograms normal to epithelial-mesenchymal transition(+) metastatic cancer stem cells in preneoplastic liver of transforming growth factor beta-deficient beta2-spectrin(+/-) mice. *Hepatology* 2017;65:1222–1236. [PubMed: 27863449]
47. Martin-Gallausiaux C, Beguet-Crespel F, Marinelli L, et al. Butyrate produced by gut commensal bacteria activates TGF-beta1 expression through the transcription factor SP1 in human intestinal epithelial cells. *Sci Rep* 2018;8:9742. [PubMed: 29950699]
48. Yang L, Pang Y, Moses HL. TGF-beta and immune cells: an important regulatory axis in the tumor microenvironment and progression. *Trends Immunol* 2010;31:220–7. [PubMed: 20538542]
49. Nakayama H, Kinouchi T, Kataoka K, et al. Intestinal anaerobic bacteria hydrolyse sorivudine, producing the high blood concentration of 5-(E)-(2-bromovinyl)uracil that increases the level and toxicity of 5-fluorouracil. *Pharmacogenetics* 1997;7:35–43. [PubMed: 9110360]
50. Chen J, Zaidi S, Rao S, et al. Analysis of Genomes and Transcriptomes of Hepatocellular Carcinomas Identifies Mutations and Gene Expression Changes in the Transforming Growth Factor-beta Pathway. *Gastroenterology* 2018;154:195–210. [PubMed: 28918914]
51. Maggio-Price L, Treuting P, Zeng W, et al. Helicobacter infection is required for inflammation and colon cancer in SMAD3-deficient mice. *Cancer Res* 2006;66:828–38. [PubMed: 16424015]
52. P OC, de Wouters T, Giri R, et al. The gut bacterium and pathobiont Bacteroides vulgatus activates NF-kappaB in a human gut epithelial cell line in a strain and growth phase dependent manner. *Anaerobe* 2017;47:209–217. [PubMed: 28583864]
53. Huang YH, Zhu C, Kondo Y, et al. CEACAM1 regulates TIM-3-mediated tolerance and exhaustion. *Nature* 2015;517:386–90. [PubMed: 25363763]
54. Isaksson-Mettavainio M, Palmqvist R, Dahlin AM, et al. High SMAD4 levels appear in microsatellite instability and hypermethylated colon cancers, and indicate a better prognosis. *Int J Cancer* 2012;131:779–88. [PubMed: 21964812]
55. Markowitz S, Wang J, Myeroff L, et al. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* 1995;268:1336–8. [PubMed: 7761852]

Figure 1.

Altered gut microbiota species in TGFB deficient mice. A) Representative colon tumor in a TGFB signaling-deficient mouse. Tumor (left), pathology images showing H&E staining (middle and right). B) Shotgun metagenomics analyses were performed in fecal samples collected from wild type (WT) mice and TGFB deficient mice. C-D) Increased microbiome species (C) and decreased microbiome species (D) were observed in TGFB signaling-deficient mice (n=8) compared with WT mice (n=8). Data summarized from 8 mice from

each group were plotted as relative abundance (Y-axis) of bacteria species (X-axis), the significance is performed using an unpaired *t-test*, * $p < 0.05$.

Figure 2.

CEACAMs and TGFB signaling pathway in human CRC. A) Kaplan–Meier survival plots of patients stratified according to expression of the indicated CEACAM family members. The p-values were calculated by log-rank test. B) OncoPrints of genomic alterations of the indicated genes from CRC data of cBioportal. C) Expression of *CEACAM* genes in COAD cases sorted according to TGFB signaling activity as indicated by SMAD3 (TGFB signaling mediator) and SERPINE1 (TGFB target gene) expression. The TGFB pathway activity score threshold is set to top 10% as indicated by the vertical line. D) Representative

immunohistochemical data for human colon adenoma samples stained for CEACAM1, CEACAM6, TGFBR2, SMAD4, and SPTBN1.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3.

TGFB signaling and stemness in CRC. A) Relationship between TGFB signaling activity (TGFB response) and stemness in the CRC cohort from TCGA. Samples were stratified based on the Stemness status, defined by ranking the samples by their RNA-based Stemness Index (mRNAsi) and dividing the samples in the top, intermediate, and bottom thirds. B) Stemness ranking of gastrointestinal cancers from TCGA. Cancers were ordered from low to high stemness index. COAD, colorectal adenocarcinoma; ESCA, esophageal carcinoma; LIHC, liver hepatocellular carcinoma; PAAD, pancreatic adenocarcinoma; READ, rectal

adenocarcinoma; STAD, stomach adenocarcinoma. C) Stemness ranking for COAD molecular subtypes. CIN, chromosomal instability; HM-SNV, hypermutated-single-nucleotide variant predominant; GS, genomically stable; HM-indel, hypermutated-insertion deletion mutation. D) Model showing the proposed mechanism of CEACAM inhibition of the TGFB pathway in regulating host-immune response, stem cell homeostasis, and tumorigenesis.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 4.

Structure of CEACAM5 hotspot mutants in the B3 domain. A) Diagram of CEACAM5 point mutations in CRC from cBioportal with the 2 newly identified mutations indicated. B) Predicted 3-D model of CEACAM5 B3 domain. C) Ramachandran plot for predicted model generated using Discovery studio visualizer. D) Overlay of the predicted structures of the B3 domains of each mutant with the wild-type (WT). E) Predicted changes in solvent accessibility (SAS) of B3 domain WT and mutant at position 640 (top) and WT and mutant at position 643 (bottom). Arrow shows the region of change. F) Predicted changes in

electrostatic potential of B3 domain wildtype and mutant at position 640 (top) and wild-type and mutant at position 643 (bottom). Arrow shows the region of change in electrostatic potential. G-I) Predicted interactions between WT CEACAM5 or mutant CEACAM5 B3 domains (L640I and A643T) and TGFBR1. Arrows indicate the positions of the mutated residues L640I and A643T.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 5.

Functional analysis of CEACAM5 hotspot mutations. A) Endogenous CEACAM5 (CEA) abundance in colon cancer cell lines. Data are representative of 3 experiments. B) Effect of overexpression of wild-type CEACAM5 and the indicated mutants on cell proliferation. Representative data from 1 of the 3 independent experiments is shown. Statistical significance was determined by t-test, $n=3$; *, P -value < 0.05 ; **, P -value < 0.01 . C) Effect of overexpression of wild-type CEACAM5 and the indicated mutants on colony formation. Cells transfected with empty vector serve as the control (ctrl). Arrow shows colonies. D)

Quantification of colony numbers in C. E) Effect of knockdown of CEACAM5 (CEA-KD) on sphere formation by FET cells. Cells transfected with empty vector serve as controls (ctrl). F) Inhibition of TGFB-stimulated SMAD3 activation by CEACAM5 mutants. SMAD3 phosphorylation was assessed by Western blot from HCT116 cells overexpressing wild-type CEACAM5 (WT), the indicated mutants, or empty vector under control conditions or after 30-minute exposure to TGFB (200 pM). A representative blot from 3 independent experiments is shown.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript