



Published in final edited form as:

Cancer Prev Res (Phila). 2015 July ; 8(7): 614–619. doi:10.1158/1940-6207.CAPR-14-0444.

TFF2-CXCR4 Axis is Associated with *BRAF* V600E Colon Cancer

Manish K. Gala¹, Thomas Austin¹, Shuji Ogino^{2,4,5}, and Andrew T. Chan^{1,3,*}

¹Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School

²Department of Epidemiology, Harvard School of Public Health, Boston, MA

³Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

⁴Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA

⁵Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

Abstract

Oncogene-induced senescence (OIS), a tumor suppressive mechanism that is induced by the replicative and metabolic stress of oncogene activation, is a key barrier in the development of *BRAF* V600E colon cancer. Inhibition of this mechanism has been observed through epigenetic changes observed in sporadic serrated polyps, as well as through the germline mutations associated with those who develop serrated polyposis. We hypothesize upregulated autocrine factors exist that are specific to the serrated pathway and also promote bypass of oncogene-induced senescence. To identify such autocrine factors, we integrate analyses of microarrays of sessile serrated polyps and two large colon cancer cohorts, the Cancer Genome Atlas (TCGA) (n = 153) and French national Cartes d'Identité des Tumeurs (CIT) program (n = 462), with enhanced gene annotation through natural language processing techniques of the existing medical corpus. We reproducibly associate higher expression of the ligand-receptor axis of TFF2 and CXCR4 with *BRAF* V600E mutant colon cancer ($P = 3.0 \times 10^{-3}$ and 0.077, respectively for TCGA; $P = 3.0 \times 10^{-8}$ and 5.1×10^{-7} for CIT). Given well-described oncogenic roles of TFF2 and CXCR4 in colon cancer, and availability of CXCR4 inhibitors for other clinical indications, this ligand-receptor axis may represent an actionable target for prevention and treatment of this molecular subtype of colorectal cancer.

Keywords

BRAF; TFF2; CXCR4; senescence; big data

*Corresponding author: Andrew Chan, M.D., M.P.H., Gastrointestinal Unit, Massachusetts General Hospital GRJ-825C, Boston, MA 02114. achan@partners.org; Phone: 617 724 0283; Fax: 617 726 3673.

No other conflict of interest exists.

Disclosures: A.T.C. previously served as a consultant for Bayer Healthcare, Millennium Pharmaceuticals, Pfizer Inc, and Pozen Inc. M.K.G is a co-founder and has equity in New Amsterdam Genomics, Inc. This study was not funded by Bayer Healthcare, Millennium Pharmaceuticals, Pfizer Inc, Pozen Inc, or New Amsterdam Genomics Inc.

Introduction

Sessile serrated adenomas (SSAs) represent an alternative pathway to colon carcinogenesis, representing only 2–9% of polyps found during colonoscopy, but accounting for approximately 15% of colon cancers.(1, 2) Previously, we demonstrated that germline loss-of-functions mutations in pathways of oncogene-induced senescence (OIS) predispose individuals to develop multiple SSAs, which typically harbor activating somatic mutations in *BRAF*.(3) In addition, this OIS barrier to carcinogenesis has been observed in sporadic serrated polyps.(4) Given the overall complexity of OIS signaling, we hypothesize that such lesions also secrete autocrine factors to promote bypass of this important tumor-suppressive mechanism during neoplastic development and progression. Targeting such factors, if concurrently upregulated in colorectal tumors, may represent a complementary approach to existing chemotherapeutic strategies, as well as provide new opportunities for cancer prevention.

To identify novel disease-specific and therapeutic-relevant targets, recent emphasis has been placed on “big data” approaches given the ever increasing number of publicly available genomic and transcriptome datasets for colon cancer. However, purely agnostic, discovery-based approaches to analyze such extensive datasets through either genome-wide association studies (GWAS) or unbiased transcriptomic signatures have yielded limited success in finding novel precision medicine targets for colon cancer prevention and treatment that are applicable to reasonable subsets of patients.(5) Limitations from these approach are inherent in both their study design and appeal; discarding well-established mechanisms of pathogenesis in search of a wider array of targets, the necessity of a large Bonferroni correction for multiple hypothesis testing limits the number of confident hits.(6) However, when such strategies are necessarily aggregated to the pathway level, they become heavily reliant on curated annotation libraries, which are often incomplete in the context of the existing medical corpus. For negative regulators of senescence, the associated GO annotation library (GO:2000773) contains only 9 unique genes despite the larger number of genes already implicated in this process.

To overcome these obstacles, we employ alternative “big data” approaches to facilitate more comprehensive hypothesis-directed analytics across large datasets. We independently analyze and integrate the results of several publicly-available microarray datasets spanning serrated neoplastic progression from polyps to cancers with natural language processing techniques to interrogate the entire medical corpus for relevant senescence mechanisms. These approaches rely on the aggregate knowledge of the entire scientific community. We demonstrate the strengths of this approach by identifying a previously unrecognized ligand-receptor axis upregulated in *BRAF* V600E colon cancer, for which an already existing FDA-approved therapy may be repurposed.

Materials and Methods

Microarray and Protein Array Analyses

All publicly available microarray datasets for human sessile serrated polyps/adenomas were imported directly from the NCBI Gene Expression Omnibus (GSE12514, GSE43841, and

GSE45270) and analyzed in the Nextbio cloud platform (Illumina Inc., Santa Clara, CA).(7, 8) Prospectively, the conventionally applied threshold of genes that had a 2-fold overexpression in SSAs over normal mucosa (if the latter not present, then compared against tubular adenomas) and a $p < 0.05$ in 2/3 datasets were selected for further analysis. This filter identified 104 genes. Level 3 gene expression data (\log_2 normalized expression from Agilent 244K Custom Gene Expression 4502A-07), stage, *BRAF* mutation status, and mismatch repair (MMR) status for the Cancer Genome Atlas (TCGA) colon cancer cohort were obtained from the TCGA Data Portal (National Cancer Institute, Bethesda, MA) and Cbioportal (Memorial Sloan Kettering Cancer Center, New York, NY) accessed on February 5, 2014.(9) \log_2 normalized gene expression from the Affymetrix U133 2.0 plus array (Affymetrix, Santa Clara, CA), stage, MMR status, and *BRAF* V600E mutation of the French national Cartes d'Identité des Tumeurs (CIT) program were accessed from GSE39582. Since Affymetrix U133 2.0 arrays have multiple CXCR4 probesets, we utilized the probe 217028 which was predicted *in silico* to most accurately represent CXCR4 expression (by the Jetset algorithm) over combined calls from all the probesets.(10)

Level 3 protein reverse phase protein array data were available for a subset of TCGA colon cancer cases used in the original publication. (MD Anderson Cancer Center, Houston, TX). Phospho-ERK1/2 \log_2 normalized levels were available under values listed under probe MAPK_pT202_Y204-R-V.

Search Algorithms and Natural Language Processing

Full gene names were derived from the ExPASy GPSDB Gene/Protein Synonyms finder from the gene symbol hits obtained from the microarray data. Only “preferred” gene names and symbols designated by ExPASy were used for automated literature search queries. Paralog gene identification was performed by matching gene symbols, searching for an exact match after truncation of the numbers at the end. The intersection of full gene names from paralogs was stored for subsequent search.

To identify genes discovered during microarray analysis relevant for senescence, we created a python script that processed all titles and abstracts of the PubMed indexed articles as of March 1, 2014 using the NCBI E-Utilities function (National Center for Biotechnology Information, Bethesda, MD). Relevant hits included articles that were searched with the gene symbol or gene stem (if paralogs also overexpressed) encoded as text words, along with the term “senescence” as a text word appearing in title or abstract. From these resultant hits, search terms were scanned to occur within two sentences of each other (proximity search) in the title or abstract by using the “.” character as a determinant of sentence boundaries.

Next, to find genes that inhibit senescence, we parsed sentences containing the query term and “senescence” and determined the smallest word distance (with preference occurring in the same sentence), and word order between them. For the case in which the gene preceded “senescence” in the word order, up to 5 words occurring in the same sentence proximal to the query term were examined to ensure no inhibition was discovered by searching for word stems (and all associated synonyms retrieved using the Big Huge Thesaurus API) of the following words: knockdown, knockout, silence, delete, inhibit, inactivate, downregulate,

suppress.(11) An ch B, Goldblum JR, Bennett AE, Downs“tumor suppressor,” which was disregarded if found. Next, these similar word stems were examined in the 5 words preceding the term “senescence.” Cases with inhibition of the query term and no inhibition of “senescence,” and with no inhibition of the query term and inhibition of “senescence” were scored as positive hits.

In the case in which “senescence” preceded the query term we limited searches in which both phrases occurred within the same sentence, and analyzed words suggesting inhibition before each key word as described. We noted the location of the word “by” if it occurred in between terms to signify high likelihood of passive voice, and its location in relative to inhibitory words. In the scenario in which there was an inhibitory word prior to “senescence” or “by”, and no inhibitory term after “by” or before the query term, this was scored as a positive hit. Also, if there was no inhibitory word before “senescence” or “by”, and an inhibitory term after “by” and before the query term, this was also deemed a positive hit.

All positive hits determined by natural language processing were subsequently validated by manual curation. To further validate the success of algorithm, all initial hits of gene symbols intersecting with senescence were also manually evaluated by an independent reviewer (T.A.) which confirmed that the algorithm had not missed any other autocrine/paracrine effectors secreted.

Statistical Analyses

Linear regression analysis was performed using R 2.15.2 and the “lm” command with stage, BRAF mutation status (WT vs V600E), and MMR status - microsatellite stable (MSS) or microsatellite-high (MSI-H). Stage information was condensed into one of four possible values (I, II, III, or IV) for the TCGA cohort. CIT cohort had stages 0, I, II, III, and IV. For the TCGA cohort, MSI status was grouped into MSS (MSS or MSI-L) or MSI-H. P values for linear regression reported represent $\Pr(>|t|)$ values, where $P < 0.05$ was considered statistically significant. Two-tailed t-test was used to analyze the role of *BRAF* V600E mutational status on TFF2 expression in the CIT cohort after multivariate adaptive regression splines regression (R called “earth”) demonstrated that *BRAF* mutational status was the only relevant variable after the forward and backward pass. After a satisfying the Shapiro-Wilk test for normality, a Pearson’s correlation coefficient and associated P value were calculated with normalized TFF2 mRNA expression and phospho-ERK1/2 protein expression from the TCGA cohort.

Results

To perform this integrative analysis, we initially analyzed three independent microarray studies of sessile serrated polyps in comparison with either normal mucosa, or if the latter was unavailable, with tubular adenomas to search for protein-coding genes that were at least 2- fold overexpressed at a $P < 0.05$ reproducible in 2/3 datasets (Supplementary Table 1). Prospectively, we created computational rules to implicate genes involved in senescence inhibition from free text of abstracts indexed within Pubmed. Of the resulting hits from the serrated polyp microarray data, we discovered that the query terms “TFF1”, “trefoil factor”

(the stem encompassing paralog genes TFF1 and TFF2), “AGR2”, and “ME1” were linked to observations of senescence inhibition and tumor growth, with TFF1 explicitly linked to OIS (Table 1).(12)

Given that TFF1 and TFF2 are already established to be secreted effectors for epithelial restitution of the gastrointestinal tract, we next assessed if overexpression of TFF1 and TFF2 extends beyond the polyp stage to colon cancers derived from the serrated pathway. We analyzed two independent, large data sets of colon cancer gene expression that are linked to well-annotated molecular and clinical data (*BRAF* mutation status, mismatch repair status, and stage) from the initially published Cancer Genome Atlas Project and the French national Cartes d’Identité des Tumeurs (CIT) program (Table 2).(13, 14)

In the TCGA cohort, we applied multivariate linear regression incorporating the clinical categorical variables described. *BRAF* V600E tumors had significantly higher \log_2 normalized TFF2 expression in the TCGA cohort ($P = 3.0 \times 10^{-3}$). We next attempted to confirm these findings in the independent and larger CIT cohort. Given that the \log_2 normalized expression levels failed to satisfy conditions for linear regression, we used a multivariate adaptive regression splines technique. However, since model construction using forward and backward variable selection only left *BRAF* mutation status as the relevant variable, we instead performed a two-tailed t-test. This analysis confirmed the strong association between *BRAF* V600E and TFF2 expression (3.0×10^{-8}) in the CIT Cohort (Figure 1). Multivariate linear regression of \log_2 normalized TFF1 expression in the TCGA cohort was not significantly associated with *BRAF* V600E ($P = 0.13$), but demonstrated a stronger relationship in the in the CIT cohort ($P = 3.8 \times 10^{-3}$). TFF1 expression was not significantly associated with MSI status or stage in either the TCGA or CIT cohort.

Since increased TFF2 expression has been linked to increased ERK1/2 activation in other gastrointestinal cell types, as well as evidence that pro-motility effects of TFF2 are dependent on active ERK signaling, we next assessed if we could detect a positive relationship in the between TFF2 mRNA levels and phospho-ERK1/2 protein expression in colonic tissues within TCGA protein array data.(15, 16) Because there is already a well-described association between *BRAF* V600E on ERK1/2 signaling, we focused this analysis on *BRAF* WT samples with available protein array data (93 samples) (Supplementary Figure 1). We did observe a statistically significant, positive correlation between TFF2 mRNA expression and phospho-ERK1/2 levels (Pearson correlation coefficient = 0.22, two-tailed $P=0.039$).

Given the unambiguous upregulation of TFF2 in colon cancers derived from the serrated pathway, we next examined the relative expression of its putative receptor, CXCR4, which has been implicated in cancer recurrence and metastasis. In the TCGA cohort, multivariate regression demonstrated that higher CXCR4 expression was strongly associated with MSI-H status ($P=2.3 \times 10^{-4}$); an association with *BRAF* V600E mutation status that approached statistical significance was also found ($P=0.077$). In subgroup analyses restricted to MSS cancers, *BRAF* V600E was significantly associated with higher levels of CXCR4 ($P=0.043$). We again confirmed these associations in the larger CIT cohort (Figure 1). We also observed that CXCR4 expression was significantly associated with MSI-H status ($P=4.4 \times 10^{-5}$) and

BRAF V600E (5.1×10^{-7}), with no substantial collinearity observed (variance inflation factor test) between these two clinical variables in the regression model. Thus, in two independent cohorts, *CXCR4* expression in colon cancers is associated with the *BRAF* V600E mutation.

Discussion

In the context of a tumor's MMR status, the presence of the *BRAF* V600E mutation has been associated with higher mortality and worse outcomes.(17) To identify novel therapeutic strategies for treating this molecular subtype, we searched for autocrine factors that would assist in the bypass of OIS, a key tumor-suppressive mechanism that is inactivated in the course of serrated tumorigenesis. We employed multiple, independent gene expression datasets of polyps and cancers with enhanced annotation strategies based on natural language processing techniques to identify and reproducibly associate higher levels of TFF2 and *CXCR4* with *BRAF* V600E colon cancers. Furthermore, we demonstrate a modest, but statistically significant, positive correlation, between TFF2 mRNA levels and phospho-ERK1/2 status in *BRAF* wild-type colon cancers. In the context of the many numerous upstream signaling pathways that influence ERK1/2 signaling, this detectable correlation is supportive of the positive relationship between TFF2 and ERK signaling observed *in vitro* in other types of gastrointestinal cancers.(16)

Trefoil factors such as TFF1 and TFF2 are well-described signaling molecules that orchestrate intestinal epithelial reconstitution after injury – inducing epithelial proliferation and migration, while reducing local inflammation. Moreover, excess of these factors have predominantly been associated with protumorigenic effects in the colon. Previous studies have highlighted higher mRNA and protein levels of TFF2 in serrated polyps, and this study extends these observations to serrated cancers with *BRAF* V600E mutations.(7, 8, 18). Trefoil factors have been shown to promote colon cancer cell survival, anchorage-independent growth, and invasion.(19) Recently, TFF1 has been demonstrated to block early-stage OIS in a variety of cell types.(12) Blockade of its paralog, TFF2, in highly expressing colon cancers has also been demonstrated to induce apoptosis *in vitro*.(20) In addition to sporadic lesions, mRNA and protein expression of trefoil factors are also strongly upregulated in the tumors of those with serrated polyposis syndromes.(18) Taken together with our findings, these data support a general role for TFF1 and TFF2 as tumor promoters.

However, the function of TFF1 and TFF2 in carcinogenesis may also be context-dependent. Consistent with its well-described immunomodulatory role in tissue repair, an important exception to an oncogenic effect of TFF2 is its potential tumor suppressive role within an inflammatory milieu central to pathogenesis of certain cancers of the stomach (*H. pylori* induced intestinal-type) and colon (colitis-associated cancer). (21, 22) In this setting, global TFF2 deficiency appears to enhance the chronic inflammatory responses which predispose the transformation of gastrointestinal mucosa to neoplasia.

After demonstrating that TFF2 is reproducibly associated with serrated neoplasia from polyp to cancer, we subsequently demonstrated the novel association of *BRAF* V600E mutant

colon cancers with higher expression of CXCR4, the putative receptor for TFF2. Higher mRNA levels of CXCR4 in colon cancers have been previously validated at the protein level by immunohistochemistry, and have been linked to higher rates of recurrence, distant metastases, and poor survival. (23) CXCR4 activation has been described to activate EGFR and ERK signaling, pathways involved in drug resistance to the BRAF V600E inhibitor vemurafenib.(24–26)

Given the present clinical availability of a CXCR4 inhibitor (Plerixafor) for hematopoietic stem cell harvesting, as well as ongoing clinical trials of newer agents for indications such as myelokathexis, these results may have near-term clinical application. *In vitro* blockade of CXCR4 by Plerixafor in highly-expressing colon cancer cell lines has been recently demonstrated to chemosensitize cells to existing therapeutics and also independently reduce proliferation and migration.(27) CXCR4 inhibition can also inhibit the oncogenic effects of high levels of TFF2.(16) Notably, conditional knockout mouse models of either TFF2 or CXCR4 demonstrate normal colonic development.(28, 29) Also, TFF2 protein expression is relatively absent in normal colonic mucosa.(30) Based on this evidence, the biological effects CXCR4 inhibitors may be relatively selective for neoplastic tissue, limiting potential toxicity in normal tissues, a critical component of successful clinical translation.

In this study, we also link higher levels of TFF2, previously shown to be elevated in sessile serrated polyps, and now CXCR4 with the already clinically utilized molecular marker, *BRAF* V600E, in colorectal cancers. Based on these findings, this autocrine signaling axis with higher expression levels of both ligand and receptor is a promising adjuvant therapeutic target for stage II-IV *BRAF* V600E mutated colorectal cancers, as well as a candidate for cancer prevention for those with serrated polyposis who are at high-risk of developing such tumors (Figure 2). While TFF2 is not the sole ligand for CXCR4, the oncogenic effects of both proteins make CXCR4 an attractive therapeutic target for this molecular subtype. Experiments and clinical trials that repurpose CXCR4 inhibitors that have successfully completed phase I studies or are FDA approved should be considered for molecularly-tailored treatment and prevention of the serrated pathway to colorectal cancer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial Support: M.K. Gala was supported by the American College of Gastroenterology (Junior Faculty Career Development Award) and the National Institutes of Health (K23 DK103119). A. T. Chan was supported by the National Institutes of Health (K24 DK098311, R01 CA137178, R01 CA176272)

References

1. Lash RH, Genta RM, Schuler CM. Sessile serrated adenomas: prevalence of dysplasia and carcinoma in 2139 patients. *Journal of clinical pathology*. 2010; 63:681–6. [PubMed: 20547691]
2. Buda A, De Bona M, Dotti I, Piselli P, Zabeo E, Barbazza R, et al. Prevalence of different subtypes of serrated polyps and risk of synchronous advanced colorectal neoplasia in average-risk population undergoing first-time colonoscopy. *Clinical and translational gastroenterology*. 2012; 3:e6. [PubMed: 23238028]

3. Gala MK, Mizukami Y, Le LP, Moriichi K, Austin T, Yamamoto M, et al. Germline mutations in oncogene-induced senescence pathways are associated with multiple sessile serrated adenomas. *Gastroenterology*. 2014; 146:520–9. [PubMed: 24512911]
4. Kriegl L, Neumann J, Vieth M, Greten FR, Reu S, Jung A, et al. Up and downregulation of p16(Ink4a) expression in BRAF-mutated polyps/adenomas indicates a senescence barrier in the serrated route to colon cancer. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc.* 2011; 24:1015–22.
5. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. *Gastroenterology*. 2013; 144:799–807. e24. [PubMed: 23266556]
6. Williams SM, Haines JL. Correcting away the hidden heritability. *Annals of human genetics*. 2011; 75:348–50. [PubMed: 21488852]
7. Caruso M, Moore J, Goodall GJ, Thomas M, Phillis S, Tyskin A, et al. Over-expression of cathepsin E and trefoil factor 1 in sessile serrated adenomas of the colorectum identified by gene expression analysis. *Virchows Arch*. 2009; 454:291–302. [PubMed: 19172291]
8. Gonzalo DH, Lai KK, Shadrach B, Goldblum JR, Bennett AE, Downs-Kelly E, et al. Gene expression profiling of serrated polyps identifies annexin A10 as a marker of a sessile serrated adenoma/polyp. *The Journal of pathology*. 2013; 230:420–9. [PubMed: 23595865]
9. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling*. 2013; 6:pl1. [PubMed: 23550210]
10. Li Q, Birkbak NJ, Györfy B, Szallasi Z, Eklund AC. Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics*. 2011; 12:474. [PubMed: 22172014]
11. Watson, J. Big Huge Thesaurus. 2014. [cited 2014; Available from: <https://words.bighugelabs.com/api.php>]
12. Radloff DR, Wakeman TP, Feng J, Schilling S, Seto E, Wang XF. Trefoil factor 1 acts to suppress senescence induced by oncogene activation during the cellular transformation process. *Proc Natl Acad Sci U S A*. 2011; 108:6591–6. [PubMed: 21451135]
13. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012; 487:330–7. [PubMed: 22810696]
14. Marisa L, de Reynies A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS medicine*. 2013; 10:e1001453. [PubMed: 23700391]
15. Graness A, Chwieralski CE, Reinhold D, Thim L, Hoffmann W. Protein kinase C and ERK activation are required for TFF-peptide-stimulated bronchial epithelial cell migration and tumor necrosis factor-alpha-induced interleukin-6 (IL-6) and IL-8 secretion. *The Journal of biological chemistry*. 2002; 277:18440–6. [PubMed: 11884401]
16. Dubeykovskaya Z, Dubeykovskiy A, Solal-Cohen J, Wang TC. Secreted trefoil factor 2 activates the CXCR4 receptor in epithelial and lymphocytic cancer cell lines. *The Journal of biological chemistry*. 2009; 284:3650–62. [PubMed: 19064997]
17. Lochhead P, Kuchiba A, Imamura Y, Liao X, Yamauchi M, Nishihara R, et al. Microsatellite instability and BRAF mutation testing in colorectal cancer prognostication. *Journal of the National Cancer Institute*. 2013; 105:1151–6. [PubMed: 23878352]
18. Delker DA, McGettigan BM, Kanth P, Pop S, Neklason DW, Bronner MP, et al. RNA sequencing of sessile serrated colon polyps identifies differentially expressed genes and immunohistochemical markers. *PloS one*. 2014; 9:e88367. [PubMed: 24533081]
19. Rodrigues S, Rodrigue CM, Attoub S, Flejou JF, Bruyneel E, Bracke M, et al. Induction of the adenoma-carcinoma progression and Cdc25A-B phosphatases by the trefoil factor TFF1 in human colon epithelial cells. *Oncogene*. 2006; 25:6628–36. [PubMed: 16715141]
20. Siu LS, Romanska H, Abel PD, Baus-Loncar M, Kayademir T, Stamp GW, et al. TFF2 (trefoil family factor2) inhibits apoptosis in breast and colorectal cancer cell lines. *Peptides*. 2004; 25:855–63. [PubMed: 15177882]

21. Dubeykovskaya Z, Dubeykovskiy A, Worthley DL, Westphalen CB, Asfaha S, Kitur K, et al. Mo2079 Tff2 Is a Novel Tumor Suppressor That Inhibits Expansion of Gr1+ CD11b+ Myeloid-Derived Suppressor Cells and Blocks Colon Carcinogenesis. *Gastroenterology*. 2012; 143:e25.
22. Peterson AJ, Menheniott TR, O'Connor L, Walduck AK, Fox JG, Kawakami K, et al. Helicobacter pylori infection promotes methylation and silencing of trefoil factor 2, leading to gastric tumor development in mice and humans. *Gastroenterology*. 2010; 139:2005–17. [PubMed: 20801119]
23. Kim J, Takeuchi H, Lam ST, Turner RR, Wang HJ, Kuo C, et al. Chemokine receptor CXCR4 expression in colorectal cancer patients increases the risk for recurrence and for poor survival. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2005; 23:2744–53. [PubMed: 15837989]
24. Corcoran RB, Ebi H, Turke AB, Coffee EM, Nishino M, Cogdill AP, et al. EGFR-mediated re-activation of MAPK signaling contributes to insensitivity of BRAF mutant colorectal cancers to RAF inhibition with vemurafenib. *Cancer discovery*. 2012; 2:227–35. [PubMed: 22448344]
25. Zhao M, Discipio RG, Wimmer AG, Schraufstatter IU. Regulation of CXCR4-mediated nuclear translocation of extracellular signal-related kinases 1 and 2. *Molecular pharmacology*. 2006; 69:66–75. [PubMed: 16210428]
26. Kosriwong K, Menheniott TR, Giraud AS, Jearanaikoon P, Sripa B, Limpaboon T. Trefoil factors: tumor progression markers and mitogens via EGFR/MAPK activation in cholangiocarcinoma. *World journal of gastroenterology : WJG*. 2011; 17:1631–41. [PubMed: 21472131]
27. Heckmann D, Maier P, Laufs S, Wenz F, Zeller WJ, Fruehauf S, et al. CXCR4 Expression and Treatment with SDF-1alpha or Plerixafor Modulate Proliferation and Chemosensitivity of Colon Cancer Cells. *Translational oncology*. 2013; 6:124–32. [PubMed: 23544165]
28. Farrell JJ, Taupin D, Koh TJ, Chen D, Zhao CM, Podolsky DK, et al. TFF2/SP-deficient mice show decreased gastric proliferation, increased acid secretion, and increased susceptibility to NSAID injury. *J Clin Invest*. 2002; 109:193–204. [PubMed: 11805131]
29. Zimmerman NP, Vongsa RA, Faherty SL, Salzman NH, Dwinell MB. Targeted intestinal epithelial deletion of the chemokine receptor CXCR4 reveals important roles for extracellular-regulated kinase-1/2 in restitution. *Laboratory investigation; a journal of technical methods and pathology*. 2011; 91:1040–55.
30. Madsen J, Nielsen O, Tornoe I, Thim L, Holmskov U. Tissue localization of human trefoil factors 1, 2, and 3. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*. 2007; 55:505–13. [PubMed: 17242463]

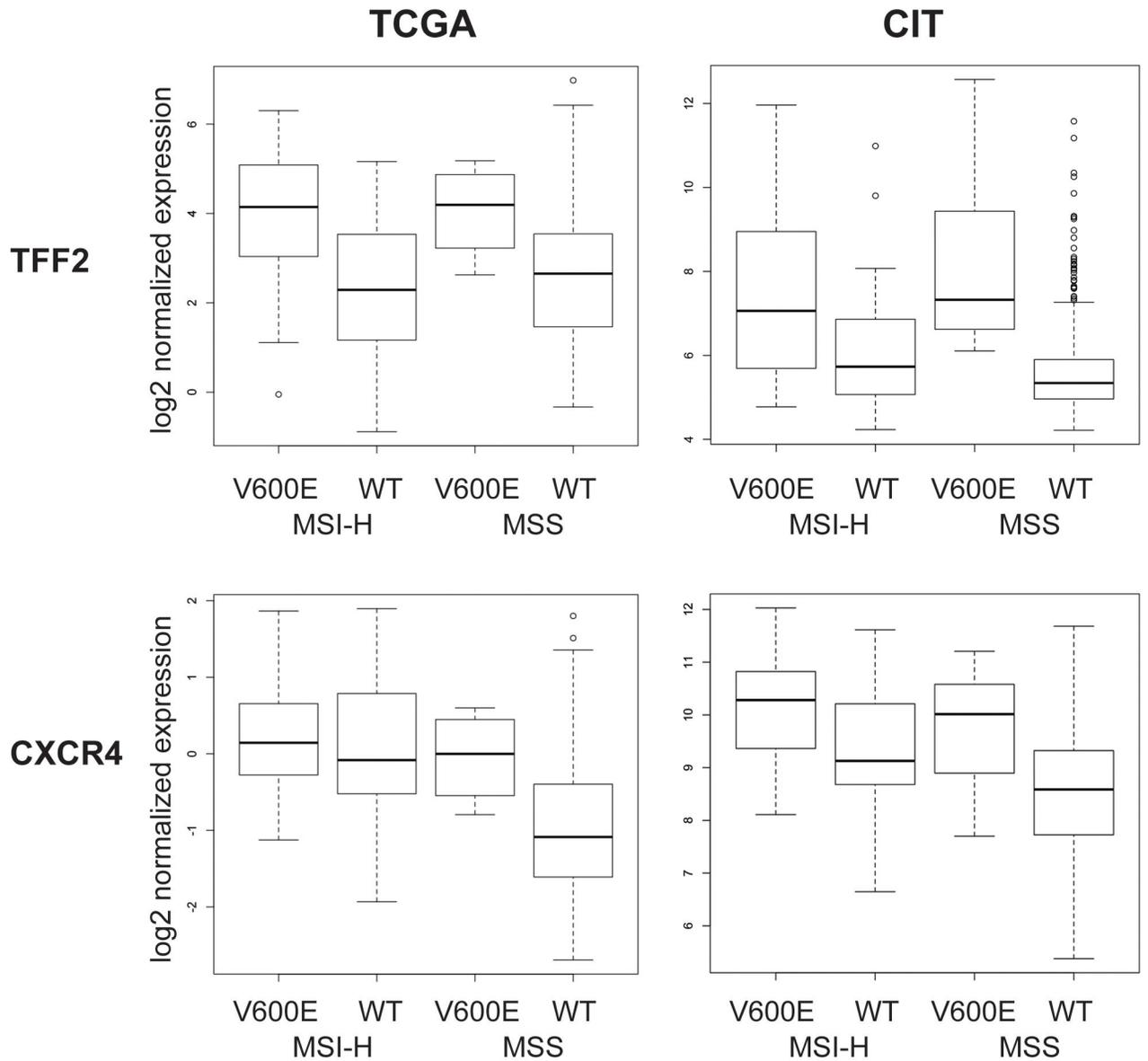


Figure 1.

Log₂ normalized gene expression of TFF2 and CXCR4 in the TCGA Colon Cancer and CIT cohorts. In multivariate regression, *BRAF* V600E mutation was significantly associated with higher levels of TFF2 in TCGA ($P = 3.0 \times 10^{-3}$), which was also validated by t-test in the CIT cohort ($P = 3.0 \times 10^{-8}$). *BRAF* V600E mutation was also associated in multivariate regression models with higher CXCR4 expression which approached statistical significance in TCGA ($P=0.077$). This association was statistically significant in the larger CIT cohort ($P = 5.1 \times 10^{-7}$).

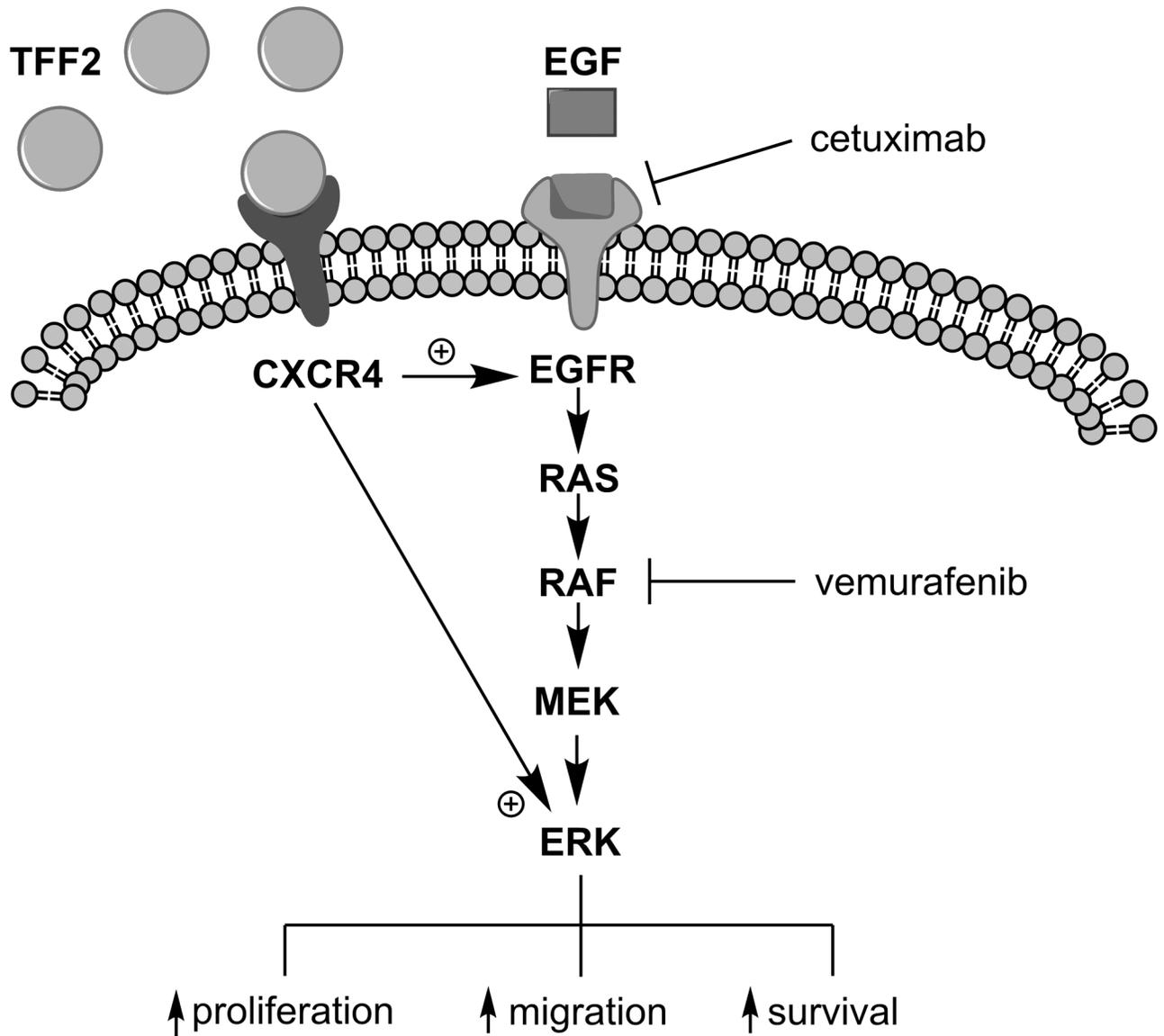


Figure 2.

Interaction of the TFF2-CXCR4 axis with the active RAS-RAF-MEK-ERK pathway in the serrated pathway of colorectal cancer. Presently, cetuximab is an EGFR inhibitor approved for colorectal cancer. Clinical trials for the use of vemurafenib for *BRAF* V600E are ongoing. ERK refers to extracellular signal-regulated kinases.

Table 1
Identification of Inhibitors of Senescence Overexpressed in Sessile Serrated Adenomas

Gene	Microarray Fold Change			Secreted	Natural Language Processing	
	GSE12514	GSE43841	GSE45270		PMID	Parsed Phrase
TFF1	6.3	N/A	5.4	Yes	21451135	Trefoil factor 1 acts to <u>suppress senescence</u>
TFF2	2.7	40.2	18.4	Yes		(Paralog of TFF1)
AGR2	2.4	1.9	2.5	No	22467239	Knockdown of AGR2 induces cellular senescence
ME1	2.2	3.2	5.0	No	23334421	<u>Downregulation of ME1</u> and <u>ME2</u> also modulates the outcome of p53 activation, leading to strong induction of senescence

104 Genes were found at least 2-fold reproducibly overexpressed in SSAs. Of these genes, natural language processing techniques of the existing medical corpus identified three genes (TFF1, AGR2, and ME1) to inhibit senescence. One paralog (TFF2) was also overexpressed. Bold text reflects query terms, where underlined word designates the negative modifier. PMID refers to PubMed ID number.

Table 2

Characteristics of the Cancer Genome Atlas (TCGA) and Cartes d'Identité des Tumeurs (CIT) Cohorts

	The Cancer Genome Atlas (%)	Cartes d'Identité des Tumeurs (%)
Total number of cases	153	462
<i>BRAF</i> Mutation Status		
<i>BRAF</i> V600E	17 (11.1)	45 (9.7)
<i>BRAF</i> WT	136 (88.9)	417 (90.3)
Microsatellite Status		
MSS	125 (81.7)	390 (84.4)
MSI-H	28 (18.3)	72 (15.6)
Stage		
Stage 0	0 (0)	4 (0.9)
Stage I	28 (18.3)	32 (6.9)
Stage II	63 (41.2)	204 (44.2)
Stage III	39 (25.5)	164 (35.5)
Stage IV	23 (15.0)	58 (12.6)

WT: Wildtype; MSS: microsatellite stable; MSI-H: microsatellite instability high

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript