Tumor and Stem Cell Biology

A Systematic Analysis of Oncogenic Gene Fusions in Primary Colon Cancer

Wigard P. Kloosterman[1], Robert R.J. Coebergh van den Braak[2], Mark Pieterse[1], Markus J. van Roosmalen[1], Anieta M. Sieuwerts[3,4], Christina Stangl[1], Ronne Brunekreef[1], Zarina S. Lalmahomed[2], Salo Ooft[5], Anne van Galen[3], Marcel Smid[3], Armel Lefebvre[1], Fried Zwartkruis[6], John W.M. Martens[3,4], John A. Foekens[3], Katharina Biermann[7], Marco J. Koudijs[1], Jan N.M. Ijzermans[2], and Emile E. Voest[5]

Abstract

Genomic rearrangements that give rise to oncogenic gene fusions can offer actionable targets for cancer therapy. Here we present a systematic analysis of oncogenic gene fusions among a clinically well-characterized, prospectively collected set of 278 primary colon cancers spanning diverse tumor stages and clinical outcomes. Gene fusions and somatic genetic variations were identified in fresh frozen clinical specimens by Illumina RNA-sequencing, the STAR fusion gene detection pipeline, and GATK RNA-seq variant calling. We considered gene fusions to be pathogenically relevant when recurrent, producing divergent gene expression (outlier analysis), or as functionally important (e.g., kinase fusions). Overall, 2.5% of all specimens were defined as harboring a relevant gene fusion (kinase fusions 1.8%). Novel configurations of *BRAF, NTRK3*, and *RET* gene fusions resulting from chromosomal translocations were identified. An R-spondin fusion was found in only one tumor (0.35%), much less than an earlier reported frequency of 10% in colorectal cancers. We also found a novel fusion involving USP9X-ERAS formed by chromothripsis and leading to high expression of ERAS, a constitutively active RAS protein normally expressed only in embryonic stem cells. This USP9X–ERAS fusion appeared highly oncogenic on the basis of its ability to activate AKT signaling. Oncogenic fusions were identified only in lymph node–negative tumors that lacked BRAF or KRAS mutations. In summary, we identified several novel oncogenic gene fusions in colorectal cancer that may drive malignant development and offer new targets for personalized therapy. *Cancer Res; 77(14); 3814–22. ©2017 AACR.*

Introduction

Colorectal cancer is the third most common malignant disease in men and second in women with an estimated yearly incidence of 1.35 million new cases associated with 694,000 annual deaths (1, 2). Within colorectal cancer, colon cancer and rectal cancer are considered two separate disease entities that are treated differently (3). This article focuses on primary nonmetastatic colon cancer (stage I to III) to avoid the identification of genetic aberrations that are a result of neoadjuvant treatment (e.g., in case of rectal cancer).

The classical driver mutations of colon cancer have been studied extensively and consist of constitutive activation of the WNT pathway by mutations in the tumor suppressor *APC*, inactivation of *TP53* and activation of RAS/MAPK pathways through mutation of *RAS* family members (4). These key primary drivers are sufficient for the transformation of primary colon stem cells into genomically instable adenocarcinomas (5, 6). Besides these classical driver genes, large-scale genomic screening has revealed hypermutated and nonhypermutated colorectal cancer, which contain a different repertoire of mutated genes (7). Nonhypermutated colorectal cancers mostly carry mutations in the classical driver genes *APC, TP53, KRAS, PIK3CA*, and *SMAD4* and form the majority of colorectal tumors. Hypermutated colorectal cancers often harbor genetic changes in DNA mismatch repair genes along with mutations in *BRAF, APC, TGFBR2*, and *ACVR2A* (7).

Genomic instability is a frequent hallmark of colorectal cancer, particularly of nonhypermutated tumors. Profiling of genomic copy number aberrations has revealed numerous recurrent changes, located at known fragile sites (*FHIT, WWOX*) or targeting tumor suppressors (*APC, PTEN, SMAD4*; ref. 7). Genomic instability can lead to the formation of fusion genes (8–10). Fusion genes have attracted significant attention because they were identified as potential cancer-specific targets for treatment.

[1]Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, CX Utrecht, the Netherlands. [2]Department of Surgery, Erasmus MC University Medical Center, Rotterdam, the Netherlands. [3]Erasmus MC Cancer Institute, Erasmus MC University Medical Center, Department of Medical Oncology and Cancer Genomics Netherlands, Rotterdam, the Netherlands. [4]Cancer Genomics Center Netherlands, Utrecht, the Netherlands. [5]Division of Molecular Oncology, Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam, the Netherlands. [6]Department of Molecular Cancer Research, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, the Netherlands. [7]Department of Pathology, Erasmus MC University Medical Center, Rotterdam, the Netherlands.

AACR

Several fusion genes (e.g., *BCR-ABL, EML4-ALK*) are clinically used to select patients for treatment (9). One of the most prevalent fusion genes described in colorectal cancer involve the R-spondin family members *RSPO2* and *RSPO3* (11). R-spondin fusions can activate WNT signaling and are mutually exclusive with mutations in *APC*. Recent work showed that inhibition of RSPO3 fusions impairs tumor growth (12). Other recurrent fusions in colorectal cancer contain the *TCF7L1* and *TCF7L2* genes, encoding TCF3 and TCF4 transcription factors, although their relevance for colorectal cancer development is currently unknown (7, 13, 14). Finally, a variety of kinase fusions have been observed in colorectal cancer, such as those involving BRAF or receptor tyrosine kinases (10, 11, 15, 16). Despite the growing support for a role of gene fusions in colorectal cancer development and their potential therapeutic value, small sample sizes, differences in experimental approaches, and the low frequency of fusions have resulted in conflicting results regarding their prevalence and relevance.

We report a comprehensive and unbiased screening for gene fusions in a unique, clinically well-defined, and prospectively collected cohort of 278 primary stage I to III colon cancers. We found that 2.5% of colon cancers in our dataset contained an oncogenic gene fusion and we identified novel fusions, including an *USP9X-ERAS* fusion with strong oncogenic activity *in vitro*.

Materials and Methods

Sample collection

Patients were selected from the MATCH-study, a prospective multicenter cohort study from 2007 onwards including adult patients undergoing curative surgery in one of seven hospitals in the Rotterdam region, the Netherlands (institutional review board number MEC-2007-088). All patients gave written informed consent for the storage and use of tissue samples for research purposes, and the collection of clinical data. The study has been conducted in accordance with the guidelines of the Declaration of Helsinki.

Only samples with at least 40% invasive tumor cells were included in the final analysis, with the number of samples per RNA sequencing run depending upon this percentage.

DNA and RNA was isolated from 30-µm sections taken from the frozen tumor tissue obtained at primary surgery. Only samples with an RNA integrity value of at least 7.0 were included in the final analysis.

RNA-sequencing

Total RNA (500 ng) from tumor samples was used as input for the Illumina TruSeq stranded RNA-seq protocol. Libraries were pooled and sequenced on Illumina HiSeq2500 or NextSeq instruments. We used the STAR fusion gene detection pipeline (version STAR-2.4.1) for analysis of RNA-seq data (17).

A list of filtered junctions was annotated by adding fusion gene counts, donor gene counts, acceptor gene counts, overlap with protein domains and calculation of expression z-scores for the donor and acceptor genes relative to samples without the fusion.

GATK RNA-seq variant calling best practices were used for somatic variant calling in RNA-seq data.

Mate-pair sequencing and analysis

Mate-pair library preparation was done using the Illumina Nextera Mate Pair library kit. Libraries were sequenced on NextSeq using 2*75 bp configuration. Discordant read pairs were detected from BAM files using a custom analysis pipeline as described previously (18). FREEC was used to detect copy number variations (19).

Whole exome sequencing

DNA was sequenced by GATC Biotech using Illumina's HiSeq protocol (paired-end 100 bp, captured regions according to SureSelect v5).

Raw sequence data for both normal and tumor DNA were mapped to the human reference genome GRCh37 using BWA (v 0.7.5a-r405; ref. 20). BAM files were used for variant calling using GATK (v3.3.0) and Mutect (v1.1.6), followed by custom filtering steps (21, 22).

Fusion gene expression and Western blotting

HEK293T cells or NIH-3T3-A14 cells were transfected with 2 µg of pBABE constructs containing fusion genes by the calcium phosphate method. NIH-3T3-A14 cells were obtained from Burgering and colleagues in 2014 (23). HEK293T cells were obtained from ATCC in the late 1980s. Both cell lines have only been tested and authenticated on the basis of their morphologic appearance. The cell lines were cultured for up to ten passages after thawing before use in experiments. Mycoplasma testing was done every three months using MycoAlert (Lonza). Cells were cotransfected with constructs encoding either MYC-tagged ERK1 or GFP-tagged AKT. Western blotting was done with rabbit polyclonal antibodies directed against phospho-AKT (Ser473; D9E, Cell Signaling Technology), phospho-p44/42 MAPK (ERK1/2, Thr202/Tyr204, Cell Signaling Technology), c-MYC (910E, Santa Cruz Biotechnology), and mouse monoclonal anti-α-Tubulin (CP06, Calbiochem). Detection was done with fluorescently labeled secondary antibodies (goat-anti-rabbit IgG (H+L) 800 CW and donkey-anti-mouse (680 RD) from LI-COR Biosciences).

Data access

The sequencing data described in this study can be accessed through the European Genome Phenome Archive under accession number EGAS00001002197.

A detailed description of materials and methods can be found in the Supplementary Materials and Methods.

Results

Transcriptome sequencing of 278 primary colon tumors

To identify rearranged transcripts we explored the transcriptomic profile of colon cancers, making use of a large prospectively collected cohort of primary tumor samples from patients with stage I to III colon cancer (Rotterdam MATCH study). Primary tumor samples were selected from our database, based on clinical and technical criteria (Supplementary Fig. S1). The study cohort included stage I ($n = 66$), stage II ($n = 115$) and stage III ($n = 97$) pathologically confirmed adenocarcinomas (Table 1). Detailed clinical description including follow-up time, adjuvant therapy, disease outcome, tumor stage, lymph node status, histologic data, and patient details was collected (Supplementary Table S1).

Tumor samples with at least 40% invasive tumor cells (based on histologic examination) were sectioned and tissue slices were consecutively used for RNA and DNA isolation. Total RNA was extracted from tissue slices and used for the preparation of RNA-seq libraries subsequent to removal of abundant noncoding mRNAs (ribominus). Libraries were sequenced at a mean depth of

Kloosterman et al.

Table 1. Patient and tumor characteristics

| Characteristics	All patients N = 278 (%)	Lymph node–negative tumors N = 181 (%)	Lymph node–positive tumors N = 97 (%)	No fusion gene N = 271 (%)	Fusion gene N = 7 (%)
Gender					
Female	132 (47.5%)	92 (50.8%)	40 (41.2%)	127 (46.9%)	5 (71.4%)
Male	146 (52.5%)	89 (49.2%)	57 (58.8%)	144 (53.1%)	2 (28.6%)
Age					
Median (IQR)	68.2 (62.4–75.2)	70.2 (63.2–76.9)	70.0 (61.4–70.8)	68.1 (62.4–75.2)	71.1 (67.4–76.1)
Tumor stage					
Stage I	66 (23.7%)	66 (36.5%)	—	65 (24.0%)	1 (14.3%)
Stage II	115 (41.4%)	115 (63.5%)	—	109 (40.2%)	6 (85.7%)
Stage III	97 (34.9%)	—	97 (100%)	97 (35.8%)	0 (0%)
T status					
T2	79 (28.4%)	66 (36.5%)	13 (13.4%)	78 (28.8%)	1 (14.3%)
T3	194 (69.8%)	110 (60.8%)	84 (86.6%)	190 (70.1%)	4 (57.1%)
T4	5 (1.8%)	5 (2.8%)	—	3 (1.1%)	2 (28.6%)
Nodal status					
N0[a]	148 (53.2%)	148 (81.8%)	—	143 (52.8%)	5 (71.4%)
N0[b]	33 (11.9%)	33 (18.2%)	—	31 (11.4%)	0 (0%)
N1	64 (23.0%)	—	63 (64.9%)	64 (23.6%)	0 (0%)
N2	33 (11.9%)	—	34 (35.1%)	33 (12.2%)	2 (28.6%)
N0	181 (65.1%)	181 (100%)	—	174 (64.2%)	7 (100%)
N+	97 (34.9%)	—	97 (100%)	97 (35.8%)	0 (0%)
Tumor grade					
Good	23 (8.3%)	16 (8.8%)	7 (7.2%)	23 (8.5%)	0 (0%)
Moderate	220 (79.1%)	152 (84.0%)	68 (70.1%)	213 (78.6%)	7 (100%)
Poor	24 (8.6%)	10 (5.5%)	14 (14.4%)	24 (8.9%)	0 (0%)
Unknown	11 (4.0%)	3 (1.7%)	8 (8.2%)	11 (4.1%)	0 (0%)
Adjuvant therapy					
No	182 (65.5%)	182 (100%)	1 (1.0%)	175 (64.6%)	7 (100%)
Yes	96 (34.5%)	—	95 (99.0%)	96 (35.4%)	0 (0%)
MSI status					
MSS	217 (78.1%)	134 (74.0%)	83 (85.6%)	215 (79.3%)	2 (28.6%)
MSI	61 (21.9%)	45 (24.9%)	14 (14.4%)	56 (20.7%)	5 (71.4%)
Location					
Left	133 (47.8%)	84 (46.4%)	49 (50.5%)	130 (48.0%)	3 (42.9%)
Right	145 (52.2%)	97 (53.6%)	48 (49.5%)	141 (52.0%)	4 (57.1%)

Abbreviation: MSI, microsatellite instability.
[a]Total lymph node yield ≥ 10.
[b]Total lymph node yield < 10.

27M paired reads per library in either 2*100 bp or 2*75 bp configuration (Supplementary Table S1). Next, the data were mapped to the human reference genome (GRCh37) to identify discordantly mapping reads indicating potential somatic fusion genes using STAR software and a custom annotation pipeline (17). The pipeline parameters were set to achieve maximal sensitivity, leading to the prediction of 3 million raw potential fusion gene calls. These calls were subsequently filtered through a series of rational filtering steps (Fig. 1), including read coverage, removal of paralogous gene sets, filtering against control data, prediction of in-frame fusion and recurrence among tumor samples (16). Using these specific filtering criteria, we obtained a dataset of 75 fusion genes, which were subjected to experimental verification by RT-PCR (Supplementary Fig. S2; Supplementary Table S2). A total of 22 out of 75 tested fusion genes were validated in the correct tumor specimen and were absent in the corresponding control tissue. We observed two cases where a *TFG-GPR128* fusion was present in both tumor and corresponding normal colon tissue. This fusion has previously been described in renal cell cancer and was later shown to be caused by a germline structural genomic variation (24, 25). For a *DLG1-BRAF* fusion, we also observed a weak RT-PCR product in the corresponding control tissue, which is likely a result of contamination of the control tissue with tumor cells. For 35 predicted fusion genes, we did not observe an RT-PCR product, indicating that these are either false positive calls or that
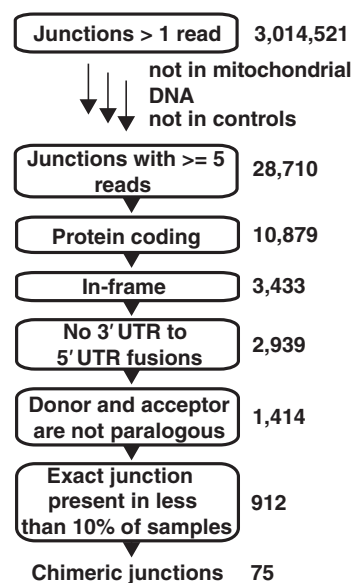
Figure 1.
Overview of fusion gene detection and filtering approach.
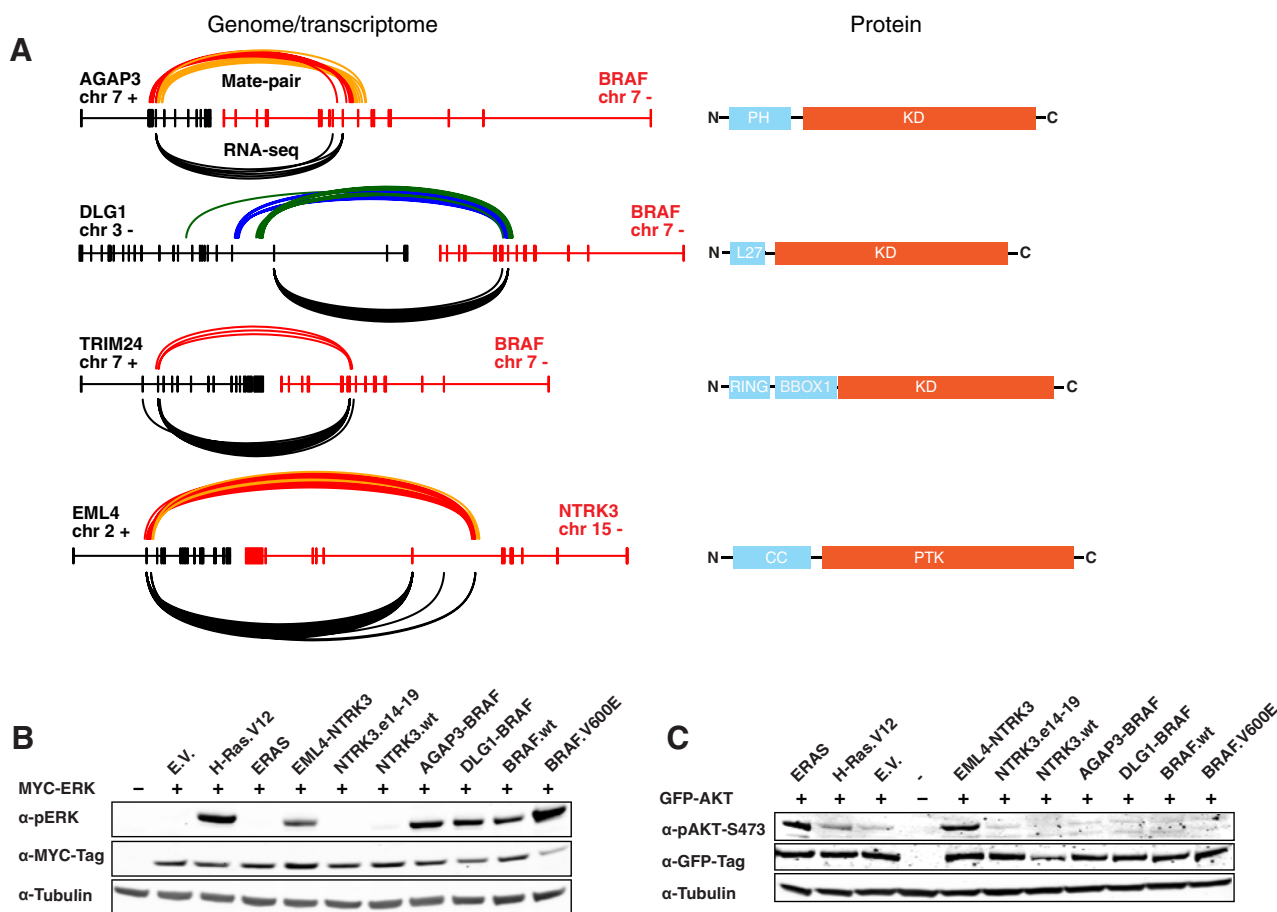
Figure 2.
Structure and characterization of fusion genes. **A,** Exon and protein structure of the *TRIM24-BRAF, AGAP3-BRAF, DLG1-BRAF*, and *EML4-NTRK3* fusions. On top of the exonic structures, we plotted arcs indicating breakpoint junction sequence reads detected by mate-pair sequencing of tumor genomic DNA. Coloring of the arcs indicates orientation of the breakpoint junction as indicated by the respective mate-pair reads: red, head-to-head inverted; yellow, tail-to-tail inverted; blue, tail-to-head; green, head-to-tail. Below the exonic structures, chimeric RNA-seq reads are plotted (black arcs) indicating which exon–exon connections were observed from the sequence data. KD, kinase domain; L27, L27 protein interaction module; PH, pleckstrin homology; RING, zinc finger domain ring type; BBOX1, B-box-type zinc finger domain; PTK, protein tyrosine kinase domain; CC, coiled-coil domain. **B,** Western blot results depicting the effects of fusion genes overexpression in HEK293 cells on ERK1 phosphorylation. **C,** Western blot results displaying the effects of fusion gene overexpression on AKT phosphorylation in NIH-3T3-A14 cells.

the fusion gene RT-PCR detection assay was suboptimal. The remainder of fusion genes was not specific for the tumor sample. To assess whether our filtering strategy indeed enriched for true positive fusion genes, we also subjected a series of 70 predicted fusions that did not pass our filtering steps to experimental verification by RT-PCR. Out of these 70 predicted fusions none could be confirmed (data not shown).

BRAF fusions are recurrent and present at low frequency in colon cancer

We next searched among the validated fusion transcripts for those, which were recurrent and in-frame. We identified three unique fusions (1.1% of 278 samples) involving the *BRAF* onco-gene (*AGAP3-BRAF, TRIM24-BRAF, DLG1-BRAF*, Fig. 2A; Supplementary Fig. S3). *BRAF* fusions have been described in a variety of cancer types (10, 16). The structure of the *TRIM24-BRAF* fusion was identical to those reported previously, with exon 3 of *TRIM24* connected to exon 10 of *BRAF* (15). The *AGAP3-BRAF* fusion

contained a junction between exon 8 of *AGAP3* and exon 9 of *BRAF*, which is different from the exon 9-exon 9 configuration previously described (15). The *DLG1-BRAF* fusion, containing a junction between exon 5 of *DLG1* and exon 9 of *BRAF*, is novel and extends the broad spectrum of known *BRAF* fusions in cancer.

We sequenced the genomes of the tumor samples with *BRAF* fusion genes using large-insert mate-pair sequencing (insert size 2.5 kb) to detect somatic structural variations that could account for *BRAF* fusion formation. In two cases (*AGAP3-BRAF* and *TRIM24-BRAF*) the fusion was caused by an inversion event, while the novel *DLG1-BRAF* fusion resulted from a reciprocal translocation between chromosomes 3 and 7 (Fig. 2A).

All three fusion genes contained the entire C-terminal kinase domain of BRAF by fusion of exon 9 or 10 to their respective donor genes. We hypothesized that disconnection of the BRAF kinase domain from its N-terminal autoinhibitory domain leads to constitutive activation (26). To assess whether our *BRAF* fusions can activate signaling pathways, we cloned the *DLG1-BRAF* and

AGAP3-BRAF fusion genes and expressed them in HEK293 cells. Subsequent analysis for activation of ERK/MAPK signaling was performed by coexpression of ERK1. Protein analysis revealed a strong effect of BRAF fusion proteins on ERK1 phosphorylation, underscoring their role as oncogenes in colon cancer (Fig. 2B). Although the effect of BRAF fusions on ERK1 phosphorylation appeared stronger than for the native BRAF protein, the effect was less strong than for BRAF carrying the activating V600E mutation.

Identification of *NTRK3* and *RET* kinase fusion genes

To further evaluate the relevance of the remaining fusion genes that were verified by RT-PCR, we reasoned that fusions that lead to upregulation of the acceptor gene may be of particular importance (27). Therefore, we analyzed the expression of our entire set of fusions and compared the expression of the donor and acceptor genes to all other tumor samples without such a fusion (outlier analysis, Supplementary Table S2).

An *EML4-NTRK3* fusion was among the top hits that resulted from this analysis with an expression Z score of 17.96 for the *NTRK3* gene. This fusion was formed through a reciprocal translocation that joined the 5′ part of *EML4* (ending with exon 2) with the 3′ exons of *NTRK3* (starting with exon 14) in a lymph node negative adenocarcinoma (Fig. 2A; Supplementary Fig. S4A). An additional *NTRK3* fusion (*ETV6-NTRK3*) has been reported previously in colon cancer and an *EML4-NTRK3* fusion has been observed in glioma (11, 28). By examining the expression of the individual exons across *ETV6-NTRK3*, we noticed that the fusion also leads to an increased expression of the exons encoding the tyrosine kinase domain, which is retained in the fusion transcript (Supplementary Fig. S4B). On the basis of these results, we analyzed the exonic expression in 732 RNA-sequencing datasets of colorectal cancer samples from The Cancer Genome Atlas (TCGA) and observed a similar increase in expression of the kinase encoding exons in two datasets derived from colon adenocarcinomas, suggesting the presence of *NTRK3* fusion genes (Supplementary Fig. S4C; ref. 7).

Neurotrophin tyrosine kinase (NTRK) 1 and 3 are receptor kinases that are frequently activated by gene fusion in a variety of cancers (10). The tyrosine kinase domain is always maintained in the chimeric proteins and fused to an oligomerization domain provided by the N-terminal fusion partner. To assess the molecular effects of the *EML4-NTRK3* fusion gene reported here, we expressed it in HEK293 cells together with *ERK1* and found that the *EML4-NTRK3* fusion activates MAPK/ERK signaling by phosphorylation of ERK1 (Fig. 2B). A truncated version of the fusion gene was not active, suggesting that the EML4 coiled-coil domain (CCD) is supposed to promote receptor activation by dimerization, similar as for *EML4-ALK* fusions found in lung cancer (29). We also tested the same fusion construct in the context of A14 cells cotransfected with AKT. Following serum starvation, we observed phosphorylation of AKT exceeding the levels of AKT phosphorylation by KRAS V12A under the same conditions (Fig. 2C). Altogether, we conclude that the *EML4-NTRK3* fusion affects oncogenic signaling pathways and that *NTRK3* fusion genes are recurrent but low-frequent in colorectal cancer.

Another top candidate with a high expression Z-score involved an in-frame fusion with exon 1–9 of the integral endoplasmic reticulum membrane protein Ribosome-binding protein 1 (RRBP1) fused to exons 12–20 of the *RET* gene, harboring the complete N-terminal kinase domain (Supplementary Fig. S5A; Supplementary Fig. S5B). The N-terminal part of RRBP1 contains the ribosome receptor lysine/proline domain as well as a coiled-coil domain (CCD). Previously reported fusions of *RET* to CCDs of 5′ partners have been shown to initiate ligand-independent activation of the kinase domain, suggesting a similar mechanism in this fusion (30, 31). The *RET* gene is a known target for gene fusions in hereditary and sporadic papillary thyroid cancers and lung adenocarcinoma, and RET fusions have recently also been described in advanced colorectal cancer (32–34).

In addition, we observed a fusion involving the kinase gene *PSKH2*, which is highly expressed in the sample with the fusion, but not at all in other tumor samples (Supplementary Table S2). However, this fusion was not pursued further because we observed several different splice variants with only partial open reading frames.

ERAS activation through gene fusion in colon cancer

One particularly interesting novel fusion gene that resulted from our outlier expression analysis contained the *ERAS* gene (Fig. 3A and B). *ERAS* is a single-exon RAS-family member that is expressed only in embryonic stem cells (35). The ERAS protein is constitutively active and leads to enhanced PI3K signaling and cellular transformation. Elevated *ERAS* expression has been described in some gastric cancer samples and a role in tumorigenesis has been implied (36). We observed that *ERAS* was highly expressed in one tumor sample in our dataset and no detectable expression was observed in the other tumor samples (Fig. 3B). The high expression was driven by the fusion of *ERAS* with *USP9X*, a highly expressed housekeeping gene (Fig. 3A; Supplementary Fig. S6A). As opposed to canonical fusion genes, which often involve formation of a novel chimeric protein sequence, the *USP9X-ERAS* fusion was formed by fusion of 5′UTR sequences, which leads to an exchange of the *ERAS* promoter with the *USP9X* promoter and not the formation of a novel chimeric protein sequence.

To gain insight in the formation of the *USP9X-ERAS* fusion, we analyzed structural variations using mate-pair sequencing. This revealed that the fusion gene was caused at the genomic level by a highly local chromothripsis event on chromosome X spanning solely the region covered by *USP9X* and *ERAS* (Fig. 3C). The chromothripsis involved at least 18 genomic breakpoint junctions and led to multiple copy number changes. We cloned the *USP9X-ERAS* fusion gene and expressed it in NIH-3T3 A14 cells. Analysis of phosphorylated AKT showed that the *USPX9-ERAS* fusion can activate AKT signaling (Fig. 2C). To get further support for a potential role of *ERAS* expression in cancer development, we assessed 521 RNA-seq datasets from colon cancer from the TCGA consortium (Supplementary Table S3). This revealed several colon cancer datasets showing detectable mRNA expression levels of *ERAS* (Supplementary Fig. S6B), albeit not as high as the sample with the *USP9X-ERAS* fusion described here. We also assessed stomach cancer RNA-seq datasets from TCGA and observed one sample with high expression of *ERAS* (Supplementary Fig. S6C). Altogether, our data suggest that induction of *ERAS* expression could be an alternative mode of promoting oncogenesis through the AKT pathway in colon cancer.

Low frequency of known R-spondin fusions

Previous work reported a number of different fusion genes in colon cancer most prominently those that involve genes that interact with the WNT signaling pathway (11, 13). Fusions involving the R-spondin genes *RSPO2* and *RSPO3* have been reported in up to 10% of microsatellite stable (MSS) colorectal
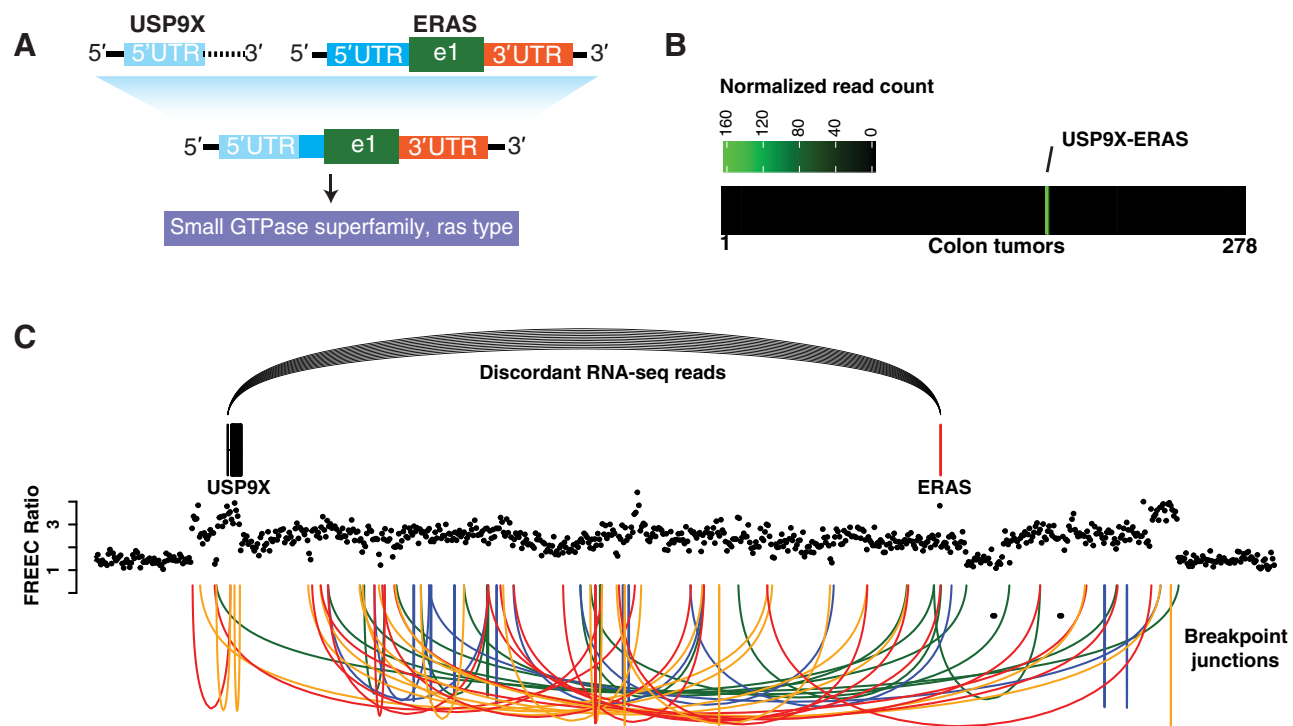
Figure 3.
Genomic origin, structure, and expression of a novel *USP9X-ERAS* fusion gene. **A,** Schematic drawing indicating the transcript structure of the *USP9X-ERAS* fusion. The fusion was caused by a breakpoint junction in the 5′UTR of *USP9X* and *ERAS*, resulting in control of *ERAS* by the *USP9X* promoter. **B,** *ERAS* expression levels across the entire cohort of 278 colon tumors included in this study. **C,** RNA-seq and mate-pair sequencing data across chromosomal regions involving the *ERAS* and *USP9X* genes. Individual chimeric RNA-seq reads are depicted as black arcs. Genomic breakpoint junctions (not individual sequence reads) are shown as colored arcs (red, head-to-head inverted; yellow, tail-to-tail inverted; blue, tail-to-head; green, head-to-tail). The genomic copy number profile is displayed using black dots, which each represent the copy number of a genomic interval as determined based on analysis of mate-pair data using FREEC.

cancers in one study and appear mutually exclusive with mutations in *APC* (11). To achieve maximal sensitivity for picking up gene fusions, we evaluated our raw fusion gene calls for the presence of both types of RSPO fusion genes, but could only detect one *EIF3E-RSPO2* fusion in an MSS sample (Supplementary Fig. S7A). To verify the sensitivity of our pipeline for picking up *RSPO2* and *RSPO3* fusion genes, we reanalyzed the raw RNA-seq FASTQ files as published recently using our STAR-based pipeline (11). Our bioinformatics pipeline could detect all seven published fusions. In addition, we measured normalized read depth across the *RSPO2* and *RSPO3* genes, revealing a strong upregulation of expression for samples with the corresponding R-spondin fusion in the published tumor samples (Supplementary Fig. S7B). We only observed elevated *RSPO2* expression for the one tumor sample in our cohort that showed the presence of an *EIF3E-RSPO2* fusion (Supplementary Fig. S7C), further supporting the low frequency of RSPO fusion genes in our dataset. We conclude that R-spondin fusions may not be as frequently present as previously indicated or that sampling bias, selection bias or treatment regime may explain the observed discrepancies.

Oncogenic fusions are mutually exclusive with activating mutations in *KRAS*, *BRAF*, and *NRAS* and restricted to stage I and II tumors.

We used the GATK-RNAseq mutation calling pipeline to detect indels and single-nucleotide changes in the RNA-seq data from all 278 tumor samples (stage I–III). The analysis was focused on

cancer genes that are of major relevance for colon cancer development, including *BRAF*, *KRAS*, *HRAS*, *NRAS*, *SMAD4*, *TP53*, *APC*, and *PTEN*. Passed variant calls in *BRAF*, *KRAS*, *HRAS*, and *NRAS* were overlapped with known hotspot mutations from the COSMIC database (37). For mutations in tumor suppressor genes, we filtered the variants against existing databases of germline variants to enrich for somatic variants. To estimate the reliability of the RNA-based variant calls, we compared them against paired tumor–normal exome sequencing data that were generated for a subset of 44 samples. For *BRAF*, *KRAS*, *HRAS*, and *NRAS*, all variants identified in the RNA-seq data were also found in the exome data and false negatives were not observed. On the basis of the RNA-seq variant calls, we identified 27 *BRAF* V600E mutations in the entire cohort of 278 tumors, which is in line with the estimated frequency (10%) of this mutation type in colorectal cancer (7). Our analysis of fusion genes showed that activation of *BRAF* may additionally be caused by fusion gene formation in an additional 1.1% of colon cancers (Fig. 4).

Besides mutations in *BRAF*, we also found 103 tumors with a hotspot mutation in *KRAS* ($n = 99$, 36%) and *NRAS* ($n = 4$, 1.4%). In line with previous observations in other cancer types, we observed that the presence of MAPK/ERK and PI3K/AKT activating hotspot mutations in *BRAF*, *KRAS*, and *NRAS* are mutually exclusive with the presence of oncogenic fusion genes in colon cancer ($P = 0.018$; refs. 15, 16). Finally, we noted that all oncogenic fusions (including *EIF3E-RSPO2*) were found in samples with lymph node–negative stage I and II tumors ($P = 0.047$)
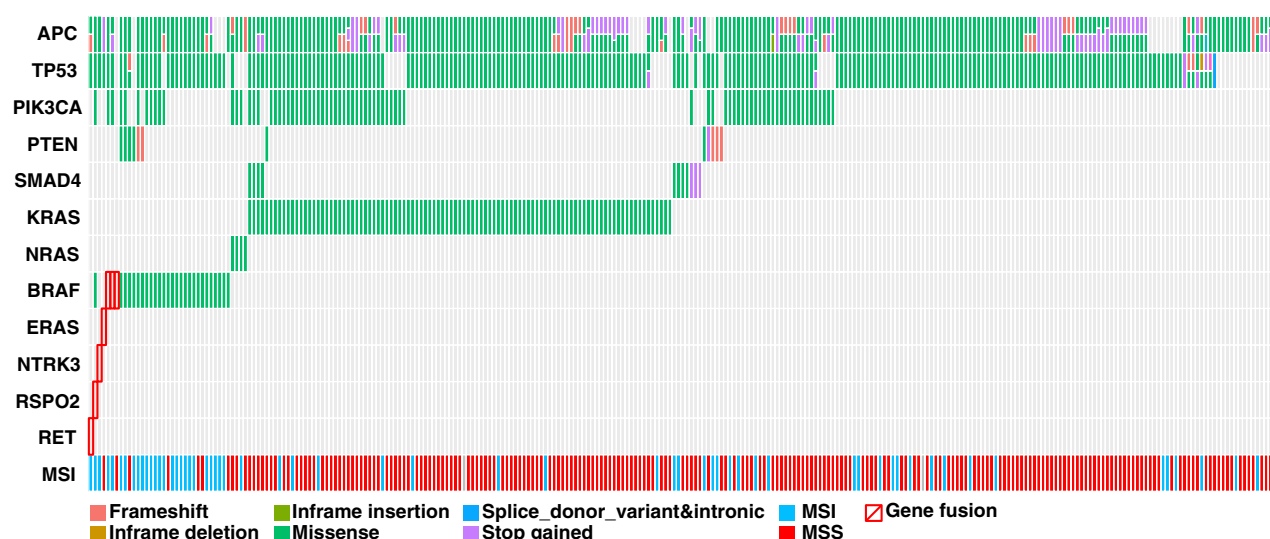
Figure 4.
Schematic overview of mutation status, microsatellite instability (MSI) status, and presence of fusion genes in colon tumors. Mutations were called for a selected set of known colon cancer genes based on RNA-seq data and intersection with COSMIC mutations. Mutation types are indicated with different colors according to their predicted effect.

and none of the samples showed a relapse in subsequent years (median follow up 50.9 months). However, the latter results should be interpreted with caution due to the small numbers.

Discussion

Our comprehensive analysis of RNA sequencing data from 278 well-characterized stage I to III colon cancers yielded a number of known and novel fusion genes, which may have clinical implications. In the era of personalized medicine, tumors are increasingly molecularly profiled, leading to better identification of patients for specific treatments (38). For colorectal cancer, small gene panels including *BRAF*, *KRAS*, and *NRAS* are most often used since mutations in these genes are of clinical relevance (39). Our analyses show that beyond these single gene tests, fusion genes may also be important.

Three of the fusion genes identified in our cohort involve the *BRAF* oncogene, which has previously been found in 4 (0.2%) colon cancer samples out of 2,154 colorectal cancer samples (15). Here we show that *BRAF* fusions occur in 1.1% of stage I–III colon cancers. Two of the *BRAF* fusions (*AGAP3-BRAF* and *TRIM24-BRAF*) consist of know fusion configurations, while the *DLG1-BRAF* fusion is novel (15). The *BRAF* fusions activate oncogenic signaling pathways in cells lines, indicating that they form genuine oncogenes in colon cancer, in addition to known oncogenic mutations in *BRAF* and *KRAS*. Although *BRAF* fusions are relatively rare, they may be highly relevant drug targets for the individual patient, similar as mutations in *BRAF* (40–42).

An expression outlier analysis involving samples with and without fusion genes, revealed *EML4-NTRK3*, *RRBP1-RET*, and *USPX9-ERAS* fusion genes. The *EML4-NTRK3* fusion gene has not been described in colon cancer, but was reported in a single case of glioma (28). However separately, both the *EML4* and *NTRK3* gene have been described as part of gene fusions in various types of cancers (10). *EML4* has mainly been described in conjunction

with the *ALK* kinase gene in non–small cell lung cancer occurring in five different variants (43). All of these contain a CCD, which is responsible for the dimerization and constitutive activation of its acceptor gene product. This is consistent with our findings that the *EML4-NTRK3* fusion induces ERK1 and AKT phosphorylation, while expression of a truncated version of *NTRK3* or the entire *NTRK3* gene did not reveal such activity.

RET fusions have been described in up to one-third of papillary thyroid cancers, in 2% of lung adenocarcinoma and recently in 0.2% of 3,117 advanced colorectal tumors (32–34). Tumors carrying a *RET* fusion in that colorectal cancer cohort were pan-negative for known driver mutations such as *KRAS*, *BRAF*, *PIK3CA*, and *EGFR*, which was also true for the tumor carrying the RRBP1-RET fusion in our cohort. RET kinase inhibitors might form a promising treatment for colorectal cancers containing oncogenic *RET* fusions (34).

An entirely novel fusion gene described in this work, comprises the *USP9X* and *ERAS* genes. Although this fusion has only been found in a single colon cancer sample in our study, its high expression and *in vitro* activity demonstrate that expression of ERAS has strong oncogenic capacity. We observed ERAS expression in colon cancer RNA-seq datasets from TCGA, suggesting that ERAS expression could be a recurrent oncogenic mechanism in colon cancer, similarly as has been proposed for stomach cancer (36).

R-spondin fusions were described as a recurrent genomic aberration in colon cancer patients by Seshagiri and colleagues, whom identified seven R-spondin fusions in a cohort of 74 colon cancer patients (9.5%; ref. 11). In their cohort, tumors with an R-spondin fusion did not contain a loss of function mutation in *APC* or copy loss, except for one tumor, which contained a single *APC* allele. Five out of seven R-spondin fusions occurred in a tumor with a *KRAS* mutation (13.5% of all *KRAS*-mutant tumors) and two in a tumor carrying a *BRAF*-mutation (40% of all *BRAF*-mutant tumors). In our cohort of 278 patients we observed only one R-spondin fusion, which was present in a tumor sample

carrying a *BRAF* mutation (3.7% of all *BRAF*-mutated tumors). However, the percentage of *KRAS*-mutated tumors differed substantially between the cohort of Seshagiri and our cohort (*KRAS* 50% vs. 35.6% $P = 0.024$ and *BRAF* 6.8% vs. 9.7% $P = 0.43$, respectively). The presence of R-spondin fusions in a subset of colorectal adenomas (traditional serrated adenoma) with frequent *KRAS* mutations has been recently shown (44). These data suggest that differences between tumor cohorts may explain the differences in the total number of identified R-spondin fusions.

Our findings are in line with new insights that broader and systematic use of genetic profiling including DNA and RNA sequencing is needed to maximize identification of patients that could potentially benefit from targeted treatment (45). Sharing of datasets including clinical characteristics and treatment outcome, such as our dataset, may help to overcome sample size limitations of individual studies and improve insight into the clinical merit of specific infrequent genetic aberrations and fusion genes (46). We found that oncogenic fusion genes were present in lymph node–negative tumors, although this finding needs to be substantiated in larger studies. Most of the previous studies reporting fusion genes in colorectal cancer did not include clinical or histopathologic characteristics, especially not stage.

In conclusion, we have created a large and comprehensive catalog of fusion genes in a unique clinically well-defined prospectively collected cohort of stage I to III primary colorectal cancers and identified several known and novel fusion genes with biological activity and possible prognostic value. We anticipate that incorporating *in vitro* platforms such as (tumor)organoids may facilitate testing of fusion genes for functional relevance, differences in oncogenic capacity and response to antitumor drugs (5).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: W.P. Kloosterman, Z.S. Lalmahomed, J.W.M. Martens, J.A. Foekens, M.J. Koudijs, E.E. Voest
Development of methodology: W.P. Kloosterman, R. Brunekreef, M.J. Koudijs, E.E. Voest

References

1. GLOBOCAN. Lyon, France: IARC; 2016. Available from: http://globocan.iarc.fr/Default.aspx.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87–108.
3. Tamas K, Walenkamp AM, de Vries EG, van Vugt MA, Beets-Tan RG, van Etten B, et al. Rectal and colon cancer: not just a different anatomic site. Cancer Treat Rev 2015;41:671–9.
4. Fearon ER. Molecular genetics of colorectal cancer. Annu Rev Pathol 2011;6:479–507.
5. Drost J, van Jaarsveld RH, Ponsioen B, Zimberlin C, van Boxtel R, Buijs A, et al. Sequential cancer mutations in cultured human intestinal stem cells. Nature 2015;521:43–7.
6. Matano M, Date S, Shimokawa M, Takano A, Fujii M, Ohta Y, et al. Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids. Nat Med 2015;21:256–62.
7. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330–7.
8. Kumar-Sinha C, Kalyana-Sundaram S, Chinnaiyan AM. Landscape of gene fusions in epithelial cancers: seq and ye shall find. Genome Med 2015;7:129.
9. Mertens F, Johansson B, Fioretos T, Mitelman F. The emerging complexity of gene fusions in cancer. Nat Rev Cancer 2015;15:371–81.
10. Stransky N, Cerami E, Schalm S, Kim JL, Lengauer C. The landscape of kinase fusions in cancer. Nat Commun 2014;5:4846.
11. Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy CB, et al. Recurrent R-spondin fusions in colon cancer. Nature 2012;488:660–4.
12. Storm EE, Durinck S, de Sousa e Melo F, Tremayne J, Kljavin N, Tan C, et al. Targeting PTPRK-RSPO3 colon tumours promotes differentiation and loss of stem-cell function. Nature 2016;529:97–100.
13. Bass AJ, Lawrence MS, Brace LE, Ramos AH, Drier Y, Cibulskis K, et al. Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. Nat Genet 2011;43:964–8.
14. Nome T, Hoff AM, Bakken AC, Rognum TO, Nesbakken A, Skotheim RI. High frequency of fusion transcripts involving TCF7L2 in colorectal cancer: novel fusion partner and splice variants. PLoS One 2014;9:e91264.
15. Ross JS, Wang K, Chmielecki J, Gay L, Johnson A, Chudnovsky J, et al. The distribution of BRAF gene fusions in solid tumors and response to targeted therapy. Int J Cancer 2016;138:881–90.

16. Yoshihara K, Wang Q, Torres-Garcia W, Zheng S, Vegesna R, Kim H, et al. The landscape and therapeutic relevance of cancer-associated transcript fusions. Oncogene 2015;34:4845–54.

17. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21.

18. Kloosterman WP, Tavakoli-Yaraki M, van Roosmalen MJ, van Binsbergen E, Renkens I, Duran K, et al. Constitutional chromothripsis rearrangements involve clustered double-stranded DNA breaks and nonhomologous repair mechanisms. Cell Rep 2012;1:648–55.

19. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 2012;28:423–5.

20. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60.

21. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303.

22. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31:213–9.

23. Burgering BM, Medema RH, Maassen JA, van de Wetering ML, van der Eb AJ, McCormick F, et al. Insulin stimulation of gene expression mediated by p21ras activation. EMBO J 1991;10:1103–9.

24. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature 2013;499:43–9.

25. Chase A, Ernst T, Fiebig A, Collins A, Grand F, Erben P, et al. TFG, a target of chromosome translocations in lymphoma and soft tissue tumors, fuses to GPR128 in healthy individuals. Haematologica 2010;95:20–6.

26. Ciampi R, Knauf JA, Kerler R, Gandhi M, Zhu Z, Nikiforova MN, et al. Oncogenic AKAP9-BRAF fusion is a novel mechanism of MAPK pathway activation in thyroid cancer. J Clin Invest 2005;115:94–101.

27. Giacomini CP, Sun S, Varma S, Shain AH, Giacomini MM, Balagtas J, et al. Breakpoint analysis of transcriptional and genomic profiles uncovers novel gene fusions spanning multiple human cancer types. PLoS Genet 2013;9:e1003464.

28. Klijn C, Durinck S, Stawiski EW, Haverty PM, Jiang Z, Liu H, et al. A comprehensive transcriptional portrait of human cancer cell lines. Nat Biotechnol 2015;33:306–12.

29. Choi YL, Takeuchi K, Soda M, Inamura K, Togashi Y, Hatano S, et al. Identification of novel isoforms of the EML4-ALK transforming gene in non-small cell lung cancer. Cancer Res 2008;68:4971–6.

30. Mulligan LM. RET revisited: expanding the oncogenic portfolio. Nat Rev Cancer 2014;14:173–86.

31. Takahashi M, Ritz J, Cooper GM. Activation of a novel human transforming gene, ret, by DNA rearrangement. Cell 1985;42:581–8.

32. Ju YS, Lee WC, Shin JY, Lee S, Bleazard T, Won JK, et al. A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. Genome Res 2012;22:436–45.

33. Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, et al. RET, ROS1 and ALK fusions in lung cancer. Nat Med 2012;18:378–81.

34. Le Rolle AF, Klempner SJ, Garrett CR, Seery T, Sanford EM, Balasubramanian S, et al. Identification and characterization of RET fusions in advanced colorectal cancer. Oncotarget 2015;6:28929–37.

35. Takahashi K, Mitsui K, Yamanaka S. Role of ERas in promoting tumour-like properties in mouse embryonic stem cells. Nature 2003;423:541–5.

36. Kubota E, Kataoka H, Aoyama M, Mizoshita T, Mori Y, Shimura T, et al. Role of ES cell-expressed Ras (ERas) in tumorigenicity of gastric cancer. Am J Pathol 2010;177:955–63.

37. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res 2015;43:D805–11.

38. Moriarity A, O'Sullivan J, Kennedy J, Mehigan B, McCormick P. Current targeted therapies in the treatment of advanced colorectal cancer: a review. Ther Adv Med Oncol 2016;8:276–93.

39. Han SW, Kim HP, Shin JY, Jeong EG, Lee WC, Lee KH, et al. Targeted sequencing of cancer-related genes in colorectal cancer using next-generation sequencing. PLoS One 2013;8:e64271.

40. Kopetz S, Desai J, Chan E, Hecht JR, O'Dwyer PJ, Maru D, et al. Phase II pilot study of vemurafenib in patients with metastatic BRAF-mutated colorectal cancer. J Clin Oncol 2015;33:4032–8.

41. Prahallad A, Sun C, Huang S, Di Nicolantonio F, Salazar R, Zecchin D, et al. Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. Nature 2012;483:100–3.

42. Sievert AJ, Lang SS, Boucher KL, Madsen PJ, Slaunwhite E, Choudhari N, et al. Paradoxical activation and RAF inhibitor resistance of BRAF protein kinase fusions characterizing pediatric astrocytomas. Proc Natl Acad Sci U S A 2013;110:5957–62.

43. Takeuchi K, Choi YL, Soda M, Inamura K, Togashi Y, Hatano S, et al. Multiplex reverse transcription-PCR screening for EML4-ALK fusion transcripts. Clin Cancer Res 2008;14:6618–24.

44. Sekine S, Yamashita S, Tanabe T, Hashimoto T, Yoshida H, Taniguchi H, et al. Frequent PTPRK-RSPO3 fusions and RNF43 mutations in colorectal traditional serrated adenoma. J Pathol 2016;239:133–8.

45. Voest EE, Bernards R. DNA-guided precision medicine for cancer: a case of irrational exuberance? Cancer Discov 2016;6:130–2.

46. Siu LL, Lawler M, Haussler D, Knoppers BM, Lewin J, Vis DJ, et al. Facilitating a culture of responsible and effective sharing of cancer genome data. Nat Med 2016;22:464–71.

Cancer Research

AAGR American Association for Cancer Research

A Systematic Analysis of Oncogenic Gene Fusions in Primary Colon Cancer

Wigard P. Kloosterman, Robert R.J. Coebergh van den Braak, Mark Pieterse, et al.

Updated version	Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-16-3563
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2017/05/16/0008-5472.CAN-16-3563.DC1

Cited articles	This article cites 45 articles, 8 of which you can access for free at: http://cancerres.aacrjournals.org/content/77/14/3814.full#ref-list-1

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/77/14/3814. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.