Accepted Article

Germline deletions in the tumour suppressor gene *FOCAD* are associated with polyposis and colorectal cancer development

Robbert D.A. Weren[1], Ramprasath Venkatachalam[1], Jean-Baptiste Cazier[2], Henner F. Farin[3], C. Marleen Kets[1], Richarda M. de Voer[1], Lilian Vreede[1], Eugène T.P. Verwiel[1], Monique van Asseldonk[4], Eveline J. Kamping[1], Lambertus A. Kiemeney[5], Kornelia Neveling[1], Katja K.H. Aben[5,6], Luis Carvajal-Carmona[7], Iris D. Nagtegaal[4], Hans K. Schackert[8], Hans Clevers[3], Marc van de Wetering[3], Ian P. Tomlinson[2], Marjolijn J.L. Ligtenberg[1,4], Nicoline Hoogerbrugge[1], Ad Geurts van Kessel[1] and Roland P. Kuiper[1].

[1]Radboud university medical center, Department of Human Genetics, Nijmegen, The Netherlands; [2]Wellcome Trust Centre for Human Genetics University of Oxford, Oxford, United Kingdom; [3]Hubrecht Institute, University Medical Center Utrecht, Utrecht, The Netherlands; [4]Radboud university medical center, Department of Pathology, Nijmegen, The Netherlands; [5]Radboud university medical center, Radboud Institute for Health Sciences, Nijmegen, The Netherlands; [6]Netherlands Comprehensive Cancer Organization, Utrecht, The Netherlands; [7]Genome Center and Department of Biochemistry and Molecular Medicine, School of Medicine, University of California, Davis, USA, [8]Department of Surgical Research, Universitätsklinikum Carl Gustav Carus, Technische Universität Dresden, Germany.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/path.4520

This article is protected by copyright. All rights reserved

Accepted Article

Correspondence:

Roland P Kuiper, PhD

☏ 0031 24 36 14107

✉ Roland.Kuiper@radboudumc.nl

Department of Human Genetics

Radboud university medical centre

PO Box 9101

6500 HB Nijmegen

The Netherlands

Conflict of Interest statement

The authors declare no competing financial interests.

This article is protected by copyright. All rights reserved

Accepted Article

Abstract

Heritable genetic variants can significantly affect the life-time risk of developing cancer, including polyposis and colorectal cancer (CRC). Variants in genes currently known to be associated with a high risk for polyposis or CRC, however, explain only a limited number of hereditary cases. The identification of additional genetic causes is, therefore, crucial to improve CRC prevention, detection and treatment. We have performed genome-wide and targeted DNA copy number profiling and resequencing in early-onset and familial polyposis/CRC patients, and show that deletions affecting the open reading frame of the tumour suppressor gene *FOCAD* are recurrent and significantly enriched in CRC patients compared to unaffected controls. All patients carrying *FOCAD* deletions exhibited a personal or family history of polyposis. RNA *in-situ* hybridization revealed *FOCAD* expression in epithelial cells in the colonic crypt, the site of tumour initiation, as well as in colonic tumours and organoids. Our data suggest that monoallelic germline deletions in the tumour suppressor gene *FOCAD* underlie moderate genetic predisposition to the development of polyposis and CRC.

Keywords

Polyposis and colorectal cancer; cancer predisposition; FOCAD; copy number variation; gene expression

This article is protected by copyright. All rights reserved

Accepted Article

Introduction

Adenomatous polyposis, the development of numerous polyps in the colon and rectum, is strongly associated with the prevalence of heritable genomic variants. These polyps have been shown to act as precursors of in-situ carcinomas. Therefore, individuals who develop adenomatous polyposis are considered to be at an increased risk to develop colorectal cancer (CRC) [1], the second most frequent cause of cancer-related death in the Western world. In the past, several genes have been associated with a high risk for polyposis, including *APC*, *MUTYH*, *AXIN2*, *SMAD4*, *BMPR1A*, *STK11*, *POLD1* and *POLE*[2]. Although germline mutations in these genes underlie the majority of polyposis cases, approximately 20% of the cases remain unexplained[3,4]. The identification of additional heritable genomic variants will be instrumental for increasing our understanding of the molecular mechanisms underlying polyposis and CRC initiation. This, in turn, will lead to an improved clinical management of individuals and families at risk, including surgical removal of polyps at regular intervals during surveillances[5].

It is generally accepted that aberrant proliferation of epithelial cells in colonic crypts represents an initiating step in the development of polyposis and CRC, and it has convincingly been shown that so-called crypt base columnar (CBC) and +4 cells, both considered to be crypt stem cells, possess cancer initiating potential[6-8]. Due to this potential, normal intestinal proliferation of these epithelial cells requires strict regulation. Loss of this strict regulation may underlie the development of multiple polyps in the colon, as illustrated by Familial Adenomatous Polyposis (FAP), Juvenile Polyposis Syndrome (JPS) and Hereditary Mixed Polyposis Syndrome (HMPS)[9]. FAP is caused by loss of functional expression of APC, a negative

This article is protected by copyright. All rights reserved

Accepted Article

regulator of β-catenin, which results in increased activation of the WNT signaling pathway and transcriptional activation of proliferation-enhancing genes, including *MYC* and *CCND1*[10,11,12]. JPS and HMPS are both caused by decreased activation of the TGF-β signaling pathway, due to loss of functional expression of the cytoplasmic mediator gene *SMAD4* and the serine-threonine kinase type I receptor gene *BMPR1A* [13,14]. Deregulation of both the WNT and the TGF-β signaling pathways is known to be associated with aberrant proliferation of epithelial cells in the colonic crypt. The proliferation of these cells is regulated by genes expressed in stem cell progenitor cells[15]. These latter genes may, therefore, act as potential polyposis or CRC susceptibility genes.

In search for novel CRC susceptibility genes in unexplained CRC families, we previously screened a cohort of 41 early-onset CRC subjects with a clear positive family history of CRC for the presence of rare DNA copy number variants (CNVs). This screening effort revealed several germline CNVs in genes that are considered to be candidates for CRC susceptibility, such as *PTPRJ* and *GREM1*[16,17]. *PTPRJ* was previously identified as a CRC susceptibility gene in mice[18] and loss of heterozygosity (LOH) of *PTPRJ* has frequently been observed in early stages of colorectal cancer development[19]. Genome-wide association studies (GWAS) have revealed that *GREM1*-related germline variants are associated with CRC susceptibility[20,21]. In addition, it has been found that a 40-kb germline duplication upstream of the *GREM1* locus is associated with an increased expression of *GREM1.* This duplication was found to be recurrent in hereditary mixed polyposis patients of Ashkenazi descent[22]. These examples clearly illustrate the power of CNV screening in the identification of novel heritable genomic variants affecting the risk to develop polyposis/CRC.

This article is protected by copyright. All rights reserved

Accepted Article

Here, we show that the tumour suppressor[23] gene *FOCAD* (encoding Focadhesin; previously known as *KIAA1797*), located on 9p21.3, is recurrently affected by CNVs in early-onset/familial CRC index patients with a personal or family history of polyposis. Previous studies have suggested a relatively low overall expression of *FOCAD* in the colon compared to other tissues[23]. We show that *FOCAD* is abundantly expressed in epithelial cells within the colonic crypt and that, as such, this gene may play a role in the development of polyposis and/or CRC. The variable numbers of *FOCAD* expressing cells in colonic organoids and tumours from different patients suggest a role of this gene in at least a subset of colonic tumours. Our findings indicate that intragenic deletions in *FOCAD* are a novel risk factor for polyposis and CRC development.

This article is protected by copyright. All rights reserved

Accepted Article

Materials and Methods

Patient material

Our initial discovery cohort, encompassing 41 patients with microsatellite stable (MSS) early-onset or familial CRC, has been described previously[16]. For a subsequent targeted screening of the *FOCAD* locus, we used an independent validation cohort of 1,232 patients diagnosed with early-onset and/or familial CRC from the Radboud university medical center, Nijmegen, the Netherlands (*n*=89), from the Universitätsklinikum Carl Gustav Carus, Dresden, Germany (*n*=159), and from the Wellcome Trust Centre for Human Genetics, University of Oxford, United Kingdom (*n*=984). An additional unrelated validation cohort of 38 polyposis patients from the Netherlands, who were diagnosed with at least 10 polyps and (i) developed serrated adenomas or (ii) developed CRC or (iii) had a positive familial history of polyposis, was also included. In order to exclude common copy number polymorphisms, we compared the patient-derived data with CNVs reported in the Database of Genomic Variants (http://projects.tcag.ca/variation)[24], our in-house database of copy number variants obtained from healthy individuals in the Netherlands (*n*=1,604), another control cohort encompassing 1,880 individuals from the Nijmegen Biomedical Study[25] and our in-house database of genomic variants, Genome Diagnostics, department of Human Genetics, Nijmegen (*n*=9,000). To determine the frequency of (highly conserved) single nucleotide variants (SNVs) in *FOCAD* in subjects without a known history of polyposis/CRC, exome data from the Exome Variant Server (EVS)(*n*=6,500)[26] and our in-house exome sequencing database (MDI) (*n*=2,096) were retrieved. An overview of all cohorts, including the selection criteria and genomic screening techniques used, is provided in

This article is protected by copyright. All rights reserved

Accepted Article

supplementary Table S1. All patient and control samples were obtained after informed consent.

Multiplex ligation-dependent probe amplification (MLPA) and genomic qPCR

MLPA probes were designed for the *FOCAD*/*miR-491* locus according to guidelines provided by MRC-Holland (Amsterdam, The Netherlands) and, subsequently, MLPA assays were performed and analyzed as described before[16]. Genomic qPCR was performed on a 7500 Fast Real-Time PCR system (Applied Biosystems, Forster City, CA, USA) as described before[16] using SYBR Green-based quantification according to the manufacturer's protocol (Bio-Rad, Veenendaal, the Netherlands). Both MLPA and genomic qPCR primers are available upon request.

Real-time quantitative RT-PCR

Real-time quantitative reverse-transcriptase PCR (RT-PCR) was performed as described previously[27]. Briefly, cDNA was prepared from 1-2 µg of RNA through RT-PCR with Oligo (dT) and random primers using a RNA LA PCR kit (AMV; Takara Bio Inc, Shiga, Japan). Real-time quantification was performed on a 7500 Fast Real-Time PCR system (Applied Biosystems, Forster City, CA, USA) using SYBR Green-based quantification (Applied Biosystems). All experiments were performed at least in duplicate and data were normalized using the housekeeping gene *HPRT* (primer sequences available upon request).

Targeted re-sequencing

This article is protected by copyright. All rights reserved

Accepted Article

Amplification of the 43 coding exons of *FOCAD* of a selected number of samples of our extended cohort from the Netherlands and Germany (*n*=117 out of 248 samples) as well as polyposis patients from the Netherlands (*n*=33 out of 38 samples) (see above) was performed using an Access Array IFC system (Fluidigm, primer sequences available upon request) and, subsequently, the amplified fragments were used for library preparation and massive parallel sequencing. Sequencing of the 117 selected cohort samples and the 33 polyposis patient samples was performed using a 454 GS FLX sequencer (Roche) with Titanium series reagents and an IonTorrent semiconductor sequencer (Life Technologies), respectively. Data analyses were performed using the Seqnext (JSI medical systems GmbH) and NextGENe software packages (Softgenetics), respectively. The average depths of coverage per amplicon are provided in the supplementary data (Figure S1 and S2). In-silico pathogenicity predictions were performed using an in-house data analysis pipeline[28], PolyPhen 2.0 (http://genetics.bwh.harvard.edu/pph2/), SIFT, Align GVGD and phyloP (Alamut version 2.1; Interactive Biosoftware, Rouen, France). Nonsynonymous SNVs present in dbSNP, in our in-house exome database and/or in EVS were considered non-damaging polymorphisms and, therefore, excluded from our analyses. Sanger sequencing was performed to confirm novel variants which fulfilled our a priori quality settings (Table S2).

Colonic organoids

Human colonic organoids and tumouroids were obtained and cultured as described previously[29]. Briefly, colonic samples were collected during endoscopies and, after crypt/adenoma isolation, cultured using previously described conditions[30].

This article is protected by copyright. All rights reserved

Accepted Article

RNA *in-situ* hybridization

RNA *in-situ* hybridization analyses were performed as described previously[31]. Briefly, digoxigenin-labeled RNA probes were generated using IMAGE clone 1204456 as template (Source BioScience). Both healthy and neoplastic colonic tissues were fixed, embedded in paraffin and pretreated prior to hybridization. After hybridization, the detection of signals was performed using an alkaline phosphatase coupled anti-digoxigenin antibody.

Statistics

A one-sided χ^2 test with Yates correction was applied to determine statistical significance of enrichment of *FOCAD* deletions in CRC patients compared to the control groups. Statistical significance of the RT-PCR results was determined using a two-tailed *t*-test assuming equal variances. For both tests, the predetermined level of significance was 0.05.

This article is protected by copyright. All rights reserved

Accepted Article

Results

FOCAD deletions are recurrent in familial and early-onset CRC patients

In a previous microarray-based CNV screen of familial and early-onset microsatellite stable (MSS) colorectal cancer (CRC) patients (*n*=41), we identified an intragenic deletion affecting the *FOCAD* gene locus, which encodes the potential tumour suppressor Focadhesin and miRNA *miR-491*, in one of the index patients (patient A)[16]. In order to assess whether an association of *FOCAD* deletions with CRC development could be confirmed, we screened an additional cohort of familial and early-onset CRC patients (*n*=1,232) as well as healthy controls (*n*=1,880), using targeted MLPA analysis. This screen revealed two additional *FOCAD* deletions in the CRC patient cohort (patients B and C), but none in the healthy control cohort (Fig. 1). In addition, no *FOCAD* deletions were found to be reported in our in-house database of copy number variants obtained from healthy individuals in the Netherlands (*n*=1,604), whereas only a single intragenic *FOCAD* deletion was identified in the samples run by our array diagnostics pipeline at the department of Human Genetics, Nijmegen (*n*=9,000). Therefore, the enrichment of *FOCAD* deletions in CRC patients (2/1,232) compared to the control group (1/12,400) is significant (*P*=0.0067). Together, these results indicate that deletions in *FOCAD* are rare but recurrent in familial and early-onset CRC patients.

The open reading frame is affected in all patients with *FOCAD* deletions

To determine the genetic boundaries of the *FOCAD* deletions, we mapped the

This article is protected by copyright. All rights reserved

Accepted Article

deletion breakpoints in each of the three patients, using a combined MLPA and genomic qPCR-based approach. All three deletions were found to be different in size and location, encompassing exons 4 to 23 in patient A, exons 2 to 14 in patient B and exons 7 to 20 in patient C (Fig. 2). Therefore, the *miR-491* locus, located within intron 4 of the *FOCAD* gene, was only affected in patients A and B. All three deletions do affect the *FOCAD* gene: two deletions include exon 4, containing the ATG start site (patients A and B), whereas the deletion in patient C results in a frameshift in the coding sequence and, consequently, a premature translational stop. As a consequence, all identified deletions encompass several exons and disturb the open reading frame (ORF) of *FOCAD*, strongly suggesting a loss-of-function scenario.

Truncating second-hit mutations are not detected in polyps and tumours of *FOCAD* deletion carriers

Based on Knudson’s two-hit paradigm, we questioned whether loss of the remaining wild-type allele might be a common event in the development of polyps and tumours in *FOCAD* deletion carriers. Therefore, Sanger sequencing was performed to reveal the presence of truncating somatic second-hit mutations in available tumour tissues of two index patients with *FOCAD* deletions. Truncating somatic mutations were identified in none out of three independent adenomas derived from patient C, nor in a tumour sample derived from patient B. These findings suggest haplo-insufficiency, rather than a classical two-hit tumour suppressor gene scenario, to be the most likely mechanism underlying CRC susceptibility in these patients.

This article is protected by copyright. All rights reserved

Accepted Article

Pathogenic *FOCAD* single nucleotide variants are not enriched in polyposis/CRC patients

Since all identified deletions result in a loss-of-function of *FOCAD*, we reasoned that pathogenic single nucleotide variants (SNVs) affecting the function of its encoded protein (Focadhesin) might also be enriched in early-onset or familial CRC patients as compared to healthy controls. No nonsense, frameshift or splice site mutations were identified in a selected cohort of 117 early-onset/familial CRC patients and 33 polyposis patients. We did, however, identify three previously unreported missense variants: p.Y759F, p.T1313A and p.S1660F (for all variants, see Table S3), one of which (p.S1660F) was predicted to be deleterious (Table S4). But, in order to draw firm conclusions about its pathogenicity additional (functional) evidence has to be obtained. Moreover, missense variants with similar in-silico characteristics have also been reported in healthy controls (Table S5).

Germline deletions of *FOCAD* are associated with polyposis

To reveal whether *FOCAD* deletion-positive patients share any phenotypic characteristics, the clinical data of our patients and their family members were collected and compared (Table 1). Patient A developed over 20 polyps and a rectal microsatellite stable carcinoma at the age of 33. He had four relatives (second and third degree) with CRC, but both parents did not develop CRC (Figure S3). Co-segregation analysis revealed that the same deletion was present in the germline of his mother, who was not affected by polyposis or CRC at time of last contact (age 66). Patient B was diagnosed with CRC at 62 years of age, in addition, was found to have four relatives (first, second and third degree) diagnosed with CRC. Patient B developed one traditional adenoma, her sister developed three adenomas and two

This article is protected by copyright. All rights reserved

Accepted Article

hyperplastic polyps and her son was diagnosed with two hyperplastic polyps (Figure S3). Patient C also had a family history positive for CRC, i.e., three relatives (first and second degree) were affected (Figure S3). He was diagnosed with adenomatous polyps at 64 years of age and had a well-documented history of constitutive polyp development for at least seven years. Therefore, in addition to familial CRC, these *FOCAD* deletion carriers appear to share a personal or family history of polyposis.

FOCAD is expressed in epithelial cells of the colonic crypt

Based on information available in public databases and published data[23], *FOCAD* appears to be ubiquitously expressed in almost all tissues, with highest levels in brain and relatively low levels in colonic tissues and localization of the encoded protein Focadhesin in the focal adhesion complex has been shown [23]. To address the question whether *FOCAD* may execute a biologically relevant role within the colon, we determined its expression level and pattern in normal colonic tissue. Using real-time quantitative RT-PCR, we indeed confirmed that the overall expression level of *FOCAD* in colonic tissue is relatively low (Fig. 3a). Since normal colonic tissue contains several cell types not directly associated with colonic tumourigenesis (e.g. myofibroblasts, endothelial cells and nonpericryptal fibroblasts), we set out to determine the expression of *FOCAD* in human colonic organoids. Human colonic organoids originate from crypt base columnar (CBC) stem cells and form *in vitro* crypt-villus-like structures without a mesenchymal niche and, therefore, only contain epithelial cells known to be involved in CRC development. We found that the expression level of *FOCAD* is higher in organoids compared to normal colonic samples (~12 fold on average, normalized on HPRT) (Fig. 3b), which

This article is protected by copyright. All rights reserved

Accepted Article

indicates that *FOCAD* is primarily expressed in epithelial cells within the colon. By subsequently employing RNA *in-situ* hybridization (ISH), we again found that *FOCAD* is expressed in the epithelial cells within the colon (Fig. 3c). Our findings show that *FOCAD* is expressed in colonic epithelial cells which may be involved in tumour formation and, therefore, *FOCAD* may act as a tumour suppressor in polyposis and CRC.

Variable expression of *FOCAD* in tumours and tumouroids

To explore the expression pattern of *FOCAD* in CRC samples, we initially compared the expression levels of this gene in 12 matched normal and primary tumour tissues using quantitative RT-PCR. This analysis revealed variable, but overall increased levels of *FOCAD* expression in the tumours (Fig. 4a). However, when expression levels were compared between organoids and patient-matched tumouroids, which are the tumourigenic equivalents of human colonic organoids originating from adenocarcinoma stem cells[30], no differences were observed. In fact, some tumouroids even showed significantly reduced expression of *FOCAD* compared to their matched organoids (Fig. 4b). Next, we applied RNA *in-situ* hybridization to multiple primary colonic tumour tissues and revealed that the number of *FOCAD* expressing cells differs between different tumours (Fig. 4c). This difference in *FOCAD* expressing cells may explain the observed differences in *FOCAD* expression levels in tumour samples, and suggests an enrichment of a specific epithelial cell sub-type in some tumours. The exact nature of these cells, however, still needs to be defined. Taken together, our results show that cells expressing *FOCAD* are present in most tumours and that, in accordance with the absence of second-hit mutations, complete loss of expression of *FOCAD* in colonic

This article is protected by copyright. All rights reserved

tumours is not likely to be a common event in CRC development.

Accepted Article

Discussion

Our data show that rare intragenic deletions in the *FOCAD* gene recurrently occur in familial and early-onset colorectal cancer (CRC) patients and that these deletions are significantly enriched in patient cohorts compared to unaffected control cohorts (P=0.0067). Furthermore, we noticed that germline *FOCAD* deletions may be associated with a polyposis phenotype, since multiple polyps were observed in all affected individuals or their family members. This observation is in agreement with a recent report in which three individuals with attenuated polyposis were described carrying a deletion or truncating mutation in *FOCAD*[32] (Figure 2). In addition, a constitutional monoallelic deletion in the *FOCAD* gene has recently been reported in an early-onset breast cancer patient[33], and recurrent deletions and somatic point mutations in *FOCAD* have been observed in sporadic cases of other cancer types[34,35,36,37] (Figures 2 and S4). These somatic and constitutional deletions in *FOCAD* substantiate its putative role as a novel cancer (susceptibility) gene.

Polyposis/CRC susceptibility factors can be divided into very rare variants with a high penetrance[38], intermediate to rare variants with a moderate penetrance[16] and common variants with a low penetrance[39]. Extensive co-segregation analyses within the families reported here could, unfortunately, not be performed due to a lack of material from the affected, often deceased, family members. However, since this

This article is protected by copyright. All rights reserved

Accepted Article

germline deletion was also found to be present in a non-affected control and in the non-affected mother of patient A, we conclude that germline *FOCAD* deletions are not fully penetrant and that additional germline variants may act as modifiers, as has for example been reported before for *APC* and *MLH1*[40,41]. Indeed, our patient with the earliest age of onset (Patient A, 33 years) was found to harbour a *de novo* pathogenic germline mutation in the exonuclease domain of *POLE*, p.Leu424Val, which may explain the early age of onset compared to the other polyposis/CRC patients with a germline deletion in *FOCAD*. Rare variants such as *FOCAD* deletions with moderate penetrances may, however, still account for a significant number of the unexplained hereditary polyposis/CRC cases, which can now easily be identified through the availability of efficient and robust detection methods[28,42,43].

To further assess whether *FOCAD* may play a role in polyposis and CRC development, we determined the expression pattern of this gene in healthy colonic tissues, including *in vitro* cultured organoids. Based on information available in public databases and published data[23], *FOCAD* appears to be ubiquitously expressed in almost all tissues, with highest levels in brain. Here, we show that *FOCAD* is expressed in epithelial cells within the colon. Since normal colonic tissue contains several non-epithelial cell types, the overall expression of *FOCAD* is relatively low. This observation is in line with the observed relatively high expression of *FOCAD* in organoids, which consist of only colonic epithelial cells. The high expression levels of *FOCAD* in epithelial cells within the colon is also in line with its potential role as novel polyposis/CRC susceptibility gene, since aberrant proliferation of epithelial cells in colonic crypts is considered to be an initiating step in the development of polyposis/CRC.

All encountered deletions affect the open reading frame (ORF) of *FOCAD* and

This article is protected by copyright. All rights reserved

Accepted Article

previously published data have shown that Focadhesin, the *FOCAD*-encoded protein, acts as a tumour suppressor [23]. The observed germline deletions strongly suggest a loss-of-function scenario, but somatic second-hit mutations could not be identified in the tumour and adenoma samples tested (from patient B and C respectively) and RNA-ISH on sporadic colonic tumours revealed that loss of *FOCAD* expression is not a common event in colonic tumourigenesis. These findings may point towards a haplo-insufficiency scenario. On the other hand, complete loss of *FOCAD* expression due to homozygous *FOCAD* deletions has been reported in glioblastomas [23] and breast cancers[36]. In addition, somatic second hit mutations in *FOCAD* were reported in an individual with attenuated polyposis[32]. Together, these data suggest that *FOCAD* is a tumour suppressor gene that can be subject to either a classical two-hit or a haplo-insufficiency scenario. In the epithelial cells of the colon loss of one allele of this gene may already underlie the development of polyposis and CRC.

The lack of enrichment of deleterious germline single nucleotide variants (SNV) in the *FOCAD* gene in our CRC cohorts is remarkable, since overall such variants are more frequent in CRC predisposing genes as compared to deletions[42,44]. Protein truncating *FOCAD* variants are very rare in the normal population, i.e., only two identical nonsense variants were identified in our in-house exome database (*n*=2,096) and truncating variants were only observed in a small percentage of subjects (<0.07%) listed in the ExAC database [45] (Figure S4). In contrast to our results, it has been reported that targeted sequencing of 192 polyposis/CRC patients revealed potentially truncating germline SNVs in two attenuated polyposis patients[32]. We assume that our strictly selected cohort was too small and/or heterogeneous to reveal a significant enrichment of truncating variants

This article is protected by copyright. All rights reserved

Accepted Article

in the *FOCAD* gene.

Previously published data have shown that Focadhesin serves as a novel component of the focal adhesion complex[23]. Although the exact function of Focadhesin remains to be established, components of the focal adhesion complex, such as the focal adhesion kinase (FAK), have already been linked to intestinal tumourigenesis[46]. Similar to the role of FAK in intestinal tumour development, Focadhesin does not only act as a novel interaction partner for the focal adhesion complex, but also exerts crucial functions in cell survival and proliferation, as illustrated by its negative effect on tumour growth[23]. Thus, like other components of the focal adhesion complex, a role of Focadhesin in the regulation of cell proliferation and, therefore, tumourigenesis is assumed.

In conclusion, we show that *FOCAD* germline deletions are recurrent and significantly enriched in patients with a positive (familial) history of polyposis/CRC. All identified deletions affect the ORF, suggesting a loss-of-function scenario. The enhanced expression of *FOCAD* in epithelial cells within colonic crypts suggests a regulatory role of this gene in the proliferation of potentially tumour-initiating colonic stem cells. We conclude that *FOCAD* may serve as a novel polyposis/CRC susceptibility gene.

Acknowledgements

We thank Rolph Pfundt for the analysis of CNVs encountered in diagnostic samples run by our diagnostics pipeline at the department of Human Genetics, Nijmegen. We thank Nienke Wieskamp for expert bioinformatic assistance. This

This article is protected by copyright. All rights reserved

work was supported by research grants from the Dutch Cancer Society (KWF, grant 2009-4335) and the Netherlands Organization for Scientific Research (NWO, grant 917-10-358).

Accepted Article

Author contributions

RDAW, RV, AGvK and RPK designed the study. SNP array and CNV analysis was performed by RV, JBC and ETPV. RDAW, RV, HFF, RMdV, LV, EJK and KN performed laboratory experiments and/or analyzed data. RV and LV performed second hit screenings, targeted re-sequencing was performed by RDAW, EJK and KN. Human colonic organoid and tumouroid culturing and expression analyses were performed by HFF and MvdW under supervision of HC. In situ hybridization was performed by HFF. CMK, HKS and NH were responsible for patient counseling and clinical data acquisition. Samples of the independent validation cohort were collected and provided by HKS, LCC and IPT. Samples from the Nijmegen Biomedical Study were provided by LAK and KKHA. Tumour histology was evaluated by IDN and DNA isolation was performed by MvA. MJLL, NH, AGvK and RPK supervised the work. RDAW, AGvK and RPK wrote the manuscript, with assistance and final approval from all coauthors.

This article is protected by copyright. All rights reserved

Accepted Article

List of Online Supporting Information

Table S1) Overview of selected cohorts in our study.

Table S2) Quality settings applied for variant calling using the SeqNext and NextGENe software package.

Table S3) Variants identified by targeted amplicon resequencing of *FOCAD* in early-onset CRC and polyposis.

Table S4) *In silico* prediction of three previously unreported missense variants in the *FOCAD* gene.

Table S5) *In silico* prediction scores of potential pathogenic missense variants called in the EVS database.

Figure S1. Average depth of coverage per *FOCAD* amplicon obtained with 454 targeted amplicon resequencing of 117 patients with early-onset CRC.

Figure S2. Average depth of coverage per *FOCAD* amplicon per barcode obtained with IonTorrent targeted amplicon resequencing of 33 polyposis patients.

Figure S3. Pedigrees of *FOCAD* deletion patients.

Figure S4. Germline and somatic variants in the *FOCAD* gene.

This article is protected by copyright. All rights reserved

Accepted Article

References

1) Liljegren A, Lindblom A, Rotstein S, *et al.* Prevalence and incidence of hyperplastic polyps and adenomas in familial colorectal cancer: correlation between the two types of colon polyps. *Gut* 2003;**52**:1140–1147.
2) Kilpivaara O, Aaltonen LA. Diagnostic cancer genome sequencing and the contribution of germline variants. *Science* 2013;**339**: 1559-62.
3) Hes FJ, Ruano D, Nieuwenhuis M, *et al.* Colorectal cancer risk variants on 11q23 and 15q13 are associated with unexplained adenomatous polyposis. *J Med Genet.* 2014;**51**: 55-60.
4) Mongin C, Coulet F, Lefevre JH, *et al.* Unexplained polyposis: a challenge for geneticists, pathologists and gastroenterologists. *Clin Genet* 2012;**81**: 38-46.
5) Wilding A, Ingham SL, Lalloo F, *et al.* Life expectancy in hereditary cancer predisposing diseases: an observational study. *J Med Genet* 2012;**49**:264-269.
6) Visvader JE. Cells of origin in cancer. Nature 2011;469;314-322.
7) Powell AE, Wang Y, Li Y, *et al.* The Pan-ErbB Negative Regulator Lrig1 Is an Intestinal Stem Cell Marker that Functions as a Tumor Suppressor. *Cell* 2012;**149**: 146–158.
8) Schepers AG, Snippert HJ, Stange DE, *et al.* Lineage Tracing Reveals Lgr5+ Stem Cell Activity in Mouse Intestinal Adenomas. *Science* 2012;**337**:730-735.
9) Medema JP, Vermeulen L. Microenvironmental regulation of stem cells in intestinal homeostasis and cancer. *Nature* 2011;**474**:318-326.
10) Clevers H., Nusse R. Wnt/b-Catenin Signaling and Disease. Cell 2012;149:1192-1205.
11) He TC, Sparks AB, Rago C *et al.* Identification of c-MYC as a target of the APC pathway. *Science* 1998;**281**:1509-1512.
12) Tetsu O, McCormick F. Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* 1999;**398**:422–426.

This article is protected by copyright. All rights reserved

Accepted Article

13) Howe JR, Roth S, Ringold JC, *et al.* Mutations in the SMAD4/DPC4 Gene in Juvenile Polyposis. *Science* 1998;**280**:1086-1088.

14) Howe JR, Bair JL, Sayed MG, *et al.* Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. *Nat Genet* 2001;**28**:184-187.

15) Medema JP, Vermeulen L. Microenvironmental regulation of stem cells in intestinal homeostasis and cancer. *Nature* 2011;**474**: 318-26.

16) Venkatachalam R, Verwiel ET, Kamping EJ, *et al.* Identification of candidate predisposing copy number variants in familial and early-onset colorectal cancer patients. *Int J Cancer* 2011;**129**:1635-1642.

17) Venkatachalam R, Ligtenberg MJ, Hoogerbrugge N, *et al.* Germline epigenetic silencing of the tumor suppressor gene PTPRJ in early-onset familial colorectal cancer. *Gastroenterology* 2010;**139**:2221-2224.

18) Ruivenkamp CA, van Wezel T, Zanon C, *et al.* Ptprj is a candidate for the mouse colon-cancer susceptibility locus Scc1 and is frequently deleted in human cancers. *Nat Genet* 2002;**31**:295-300.

19) Ruivenkamp C, Hermsen M, Postma C, *et al.* LOH of PTPRJ occurs early in colorectal cancer and is associated with chromosomal loss of 18q12–21. *Oncogene* 2003;**22**:3472–3474.

20) Jaeger E, Webb E, Howarth K, *et al.* Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat Genet* 2008;**40**:26-28.

21) Lewis A, Freeman-Mills L, de la Calle-Mustienes E, *et al.* A Polymorphic Enhancer near GREM1 Influences Bowel Cancer Risk through Differential CDX2 and TCF7L2 Binding. *Cell Report* 2014;**8**:983-990.

22) Jaeger E, Leedham S, Lewis A, *et al.* Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. *Nat Genet* 2012;**44**:699-705.

23) Brockschmidt A, Trost D, Peterziel H, *et al.* KIAA1797/FOCAD encodes a novel focal adhesion protein with tumour suppressor function in gliomas. *Brain* 2012;**135**:1027-1041.

24) Iafrate AJ, Feuk L, Rivera MN, *et al.* Detection of large-scale variation in the human genome. *Nat Genet* 2004;**36**:949-951.

This article is protected by copyright. All rights reserved

Accepted Article

25) Hoogendoorn EH, Hermus AR, de Vegt F, *et al.* Thyroid function and prevalence of anti-thyroperoxidase antibodies in a population with borderline sufficient iodine intake: influences of age and sex. *Clin Chem* 2006;**52**:104-111.

26) Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: http://evs.gs.washington.edu/EVS/) [01 (07, 2013)

27) de Voer RM, Geurts van Kessel A, Weren RD, *et al.* Germline Mutations in the Spindle Assembly Checkpoint Genes BUB1 and BUB3 Are Risk Factors for Colorectal Cancer. Gastroenterology 2013;**145**:544-547.

28) Gilissen C, Hoischen A, Brunner HG and Veltman JA. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* 2012;**20**:490–497

29) Sato T, Vries RG, Snippert HJ, *et al.* Single Lgr5 stem cells build crypt–villus structures in vitro without a mesenchymal niche. *Nature* 2009;**459**:262-265.

30) Sato T, Stange DE, Ferrante M, *et al.* Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* 2011;**141**:1762-1772.

31) Gregorieff A, Pinto D, Begthel H, *et al.* Expression pattern of Wnt signaling components in the adult intestine. *Gastroenterology* 2005;**129**:626–638.

32) Horpaopan S, Spier I, Zink AM, *et al.* Genome-wide CNV analysis in 221 unrelated patients and targeted high-throughput sequencing reveal novel causative candidate genes for colorectal adenomatous polyposis. *Int J Cancer* 2015; **136**: E578–E589.

33) Krepischi ACV, Achatz MIW, Santos EM, *et al.* Germline DNA copy number variation in familial and early-onset breast cancer. *Breast Cancer Res* 2012;**14**:R24.

34) Furuta K, Arao T, Sakai K, *et al.* Integrated analysis of whole genome exon array and array-comparative genomic hybridization in gastric and colorectal cancer cells. *Cancer Sci* 2012;**103**:221-227.

35) Wiech T, Nikolopoulos E, Weis R, *et al.* Genome-wide analysis of genetic alterations in Barrett's adenocarcinoma using single nucleotide polymorphism arrays. *Lab Invest* 2009;**89**:385-397.

This article is protected by copyright. All rights reserved

Accepted Article

36) Natrajan R, Mackay A, Lambros MB, *et al.* Whole-genome massively parallel sequencing analysis of BRCA1 mutant oestrogen receptor-negative and -positive breast cancers. *J Pathol* 2012;**227**:29–41.

37) Forbes SA, Beare D, Gunasekaran P, *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* 2014.

38) Strate LL, Syngal S. Hereditary colorectal cancer syndromes. *Cancer Causes Control* 2005;**16**:201–213.

39) Dunlop MG, Dobbins SE, Farrington SM, *et al.* Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat Genet* 2012;**44**:770-776

40) Kemp Z, Thirlwell C, Sieber O, *et al.* An update on the genetics of colorectal cancer. *Hum Mol Genet* 2004;**13**:177-185.

41) Crabtree MD, Fletcher C, Churchman M, *et al.* Analysis of candidate modifier loci for the severity of colonic familial adenomatous polyposis, with evidence for the importance of the N-acetyl transferases. *Gut* 2004;**53**:271-276.

42) Kuiper RP, Ligtenberg MJL, Hoogerbrugge N, Geurts van Kessel A. Germline copy number variation and cancer risk. *Curr Opin Genet Dev* 2010;**20**:282–289.

43) McCarthy MI, Abecasis GR, Cardon LR, *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008;**9**:356-369.

44) Haraksingh RR, Snyder MP. Impacts of Variation in the Human Genome on Gene Regulation. *J Mol Biol* 2013;**425**:3970-3977.

45) Exome Aggregation Consortium (ExAC), Cambridge, MA (URL: http://exac.broadinstitute.org) [30 (12, 2014) accessed].

46) Morton JP, Myant KB, Sansom OJ. A FAK-PI-3K-mTOR axis is required for Wnt-Myc driven intestinal regeneration and tumorigenesis. *Cell Cycle* 2011;**10**:173-175.

This article is protected by copyright. All rights reserved

Accepted Article

Table 1. Overview of clinical phenotypes of carriers of mono-allelic *FOCAD* deletions.

Patient	A	B	C
Gender	Male	Female	Male
Age of Onset	33y	62y	64y
CRC	Yes	Yes	No
Polyposis	>20 polyps	Son and sister with polyps	Constitutive polyp development
#Relatives with CRC	4	4	3
FOCAD exons deleted	4 to 23	2 to 14	7 to 20

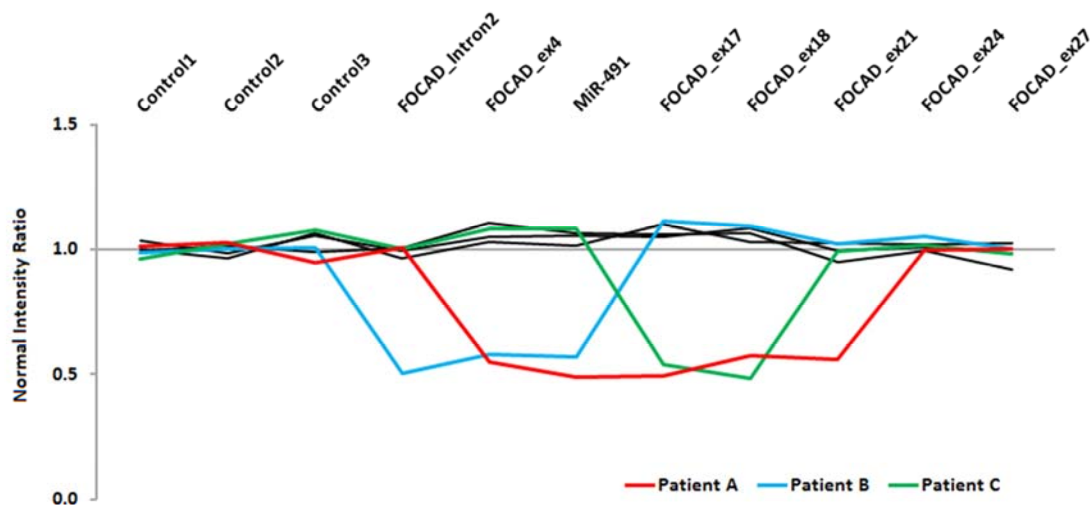

Figure 1. Confirmation and identification of genomic deletions in the *FOCAD* gene. Confirmation of the genomic deletion by MLPA in germline DNA of patient A (index patient) and identification of genomic *FOCAD* deletions by MLPA in the germline of patient B and C (extended cohort). Three controls were included in the MLPA assay. ex: exon.

This article is protected by copyright. All rights reserved

Accepted Article

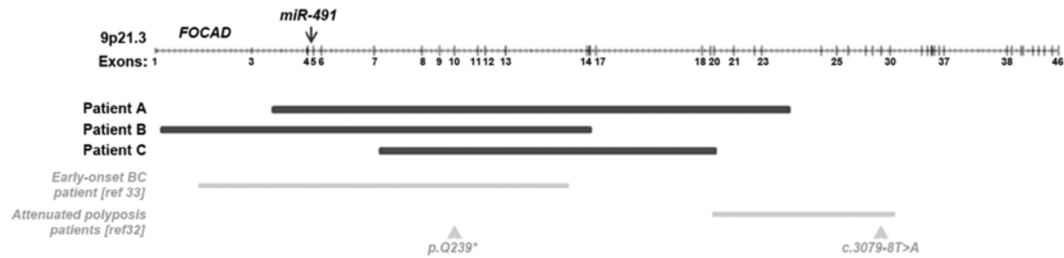

Figure 2. Diagram of germline *FOCAD* deletions, all affecting the ORF. The upper line schematically represents the *FOCAD* gene, its exons (1-46) and the location of the miR-491 locus therein (intron 4). Bars represent the deletions identified in each patient (A-C). Deletions (bars) and truncating mutations (arrow heads) depicted in grey have been encountered in an early-onset breast cancer (BC) and three attenuated polyposis patients, respectively (see references 32 and 33).

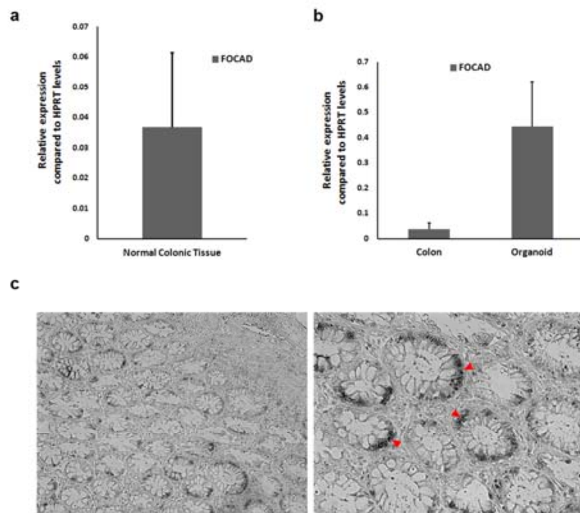

Figure 3. Expression of *FOCAD* in normal colonic tissue. A) Relative average expression of *FOCAD* in normal colonic tissue normalized to housekeeping gene *HPRT* (*n*=12). B) Expression levels of *FOCAD* in normal and organoid samples (average expression of 12 and 5 tissue samples, respectively). C) RNA in-situ hybridization reveals expression of *FOCAD* (red arrowheads) in epithelial cells within the colon. Left-Right: 10x and 20x magnification, respectively.

This article is protected by copyright. All rights reserved

Accepted Article

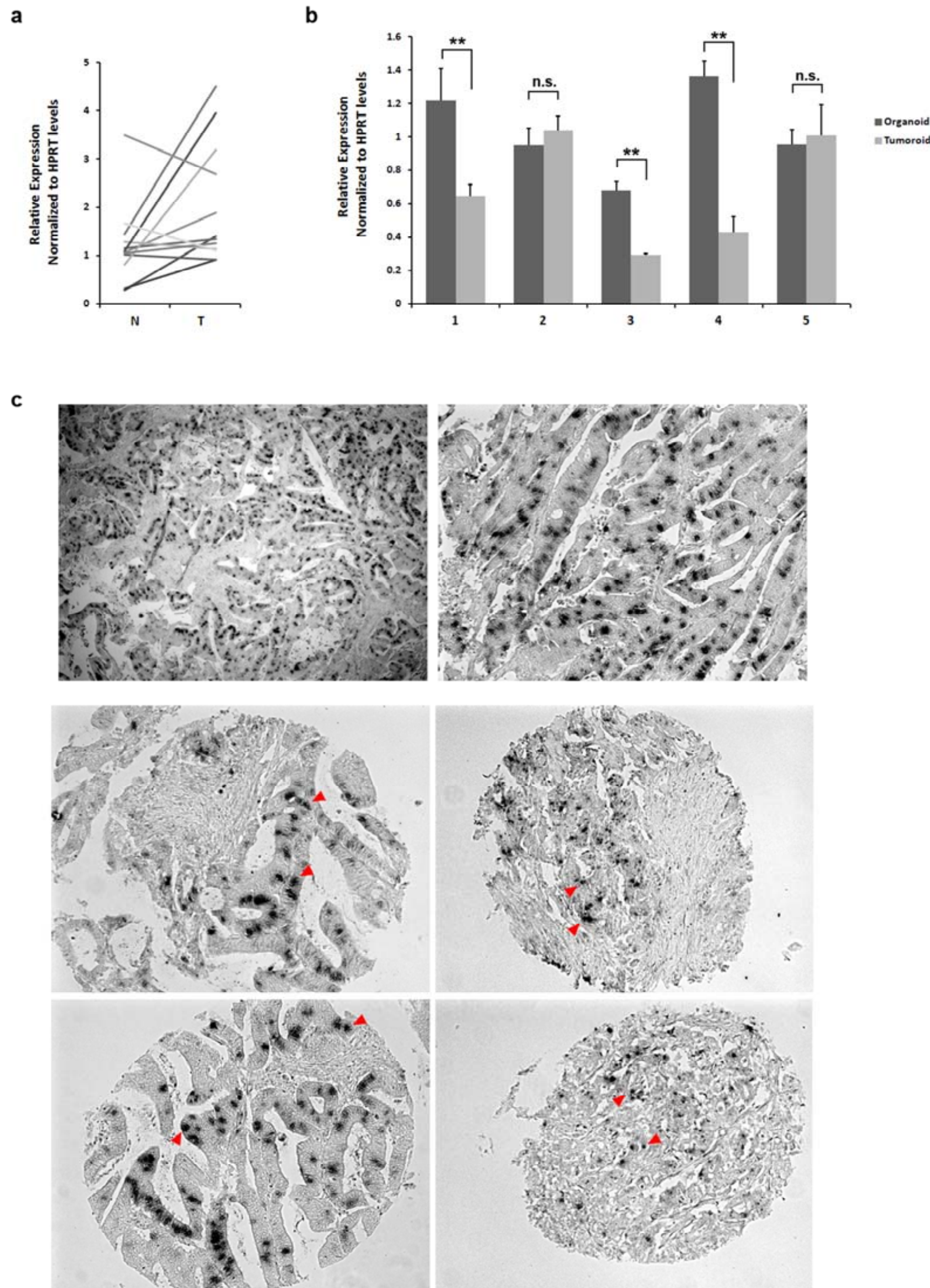

Figure 4. Expression of *FOCAD* in colonic carcinoma tissue. A) Expression levels of *FOCAD* in normal and matched tumour samples (*n*=12), normalized to housekeeping gene *HPRT*. B) Expression levels of *FOCAD* in organoid and patient-matched tumouroid (*n*=5). C) RNA in-situ hybridization for *FOCAD* in colonic tumour samples. Upper panels: large numbers of *FOCAD* expressing epithelial cells are observed in tumour sections (4x and 10x magnification). Lower Panels: Microarray tumour sections show that the amounts of *FOCAD* expressing cells (red arrowheads) vary between colon tumours from different patients (4 samples, 10x magnification). Tumours derived from germline *FOCAD* deletion carriers could not be included in this analysis. ** $P<0,01$

This article is protected by copyright. All rights reserved