의학박사 학위논문

조기발병대장암의 분자유전학적인 특징 연구; 융합유전자인 *FAM174A-WWC1*의 기능 연구

Molecular genetic analysis of early-onset colorectal cancer; functional characterization of a novel fusion gene, *FAM174A-WWC1* in cell lines

2020년 2월

서울대학교 대학원

의학과 외과학 전공

신 루 미

조기발병대장암의 분자유전학적인 특징 연구; 융합유전자인 *FAM174A−WWC1*의 기능 연구

지도 교수 정 승 용

이 논문을 의학박사 학위논문으로 제출함
2019년 10월

서울대학교 대학원
의학과 외과학 전공
신 루 미

신루미의 박사 학위논문을 인준함
2020년 1월

위 원 장 김 종 일 (인)

부위원장 정 승 용 (인)

위 원 구 자 록 (인)

위 원 장 진 영 (인)

위 원 유 병 철 (인)

Molecular genetic analysis of early-onset colorectal cancer; functional characterization of a novel fusion gene, *FAM174A-WWC1* in cell lines

by Rumi Shin, M.D.

(Directed by Seung-Yong Jeong, M.D., Ph.D.)

A Thesis Submitted to the Department of Surgery in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy in Medicine (Surgery) at Seoul National University College of Medicine, Seoul, Korea

January, 2020

Approved by thesis committee

Professor _____Chairman

Professor _____Vice Chairman

Professor _____

Professor _____

Professor _____

Abstract

Molecular genetic analysis of early-onset colorectal cancer; functional characterization of a novel fusion gene, *FAM174A-WWC1* in cell lines

Rumi Shin

Medicine (Surgery)

The Graduate School

Seoul National University

Introduction: Recent epidemiological studies suggest that there has been a significant increase of colorectal cancer (CRC) diagnosis in young adults, even though the overall incidence and mortality has decreased in

recent years. Although these early-onset CRCs (EOCRCs) are likely to suggest a hereditary predisposition, known familial CRC syndromes account for only 20% of EOCRC cases; the genetic aberrations remain unknown for the other 80% of cases. Therefore, we aimed to establish reproducible biological resources and contribute to expanding the mutational database for EOCRC.

Methods: Four cell lines were established from the original tumor tissues of individual CRC patients diagnosed prior to 30 years of age, and next-generation sequencing was used to identify the genetic features of EOCRC.

Results: Analysis of the mutation profile related to CRC carcinogenesis revealed no significant mutations except for the *TP53* splice mutation in the SNU-1460 cell line. Whole-exome sequencing data showed *PCLO* gene mutation in all EOCRC samples and four COlon ADenocarcinoma (COAD) cohort samples obtained from The Cancer Genome Atlas-Genomic Data Commons (TCGA-GDC).

In addition, we identified one novel fusion gene, *FAM174A-WWC1*, and analyzed its functional role. *FAM174A-WWC1* expression in normal fibroblasts caused alterations in cellular morphology and modulated the intercellular expression levels of E-cadherin and N-cadherin. Moreover, *FAM174A-WWC1* abrogated the membrane expression of Yes-associated

protein 1 (YAP1) and significantly increased the level of nuclear YAP1. Lastly, *FAM174A-WWC1* expression increased oncogenic capacity and invasiveness of normal fibroblasts. These findings suggest that this fusion gene may act as a driver mutation in EOCRC.

Conclusions: There were fewer mutations related to CRC carcinogenesis in 4 EOCRC cell lines. Mutations were common in the *PCLO* gene in our cell lines and EOCRC from TCGA-GDC. A novel fusion gene, *FAM174A-WWC1* increases the oncogenic and metastatic capacity of cells in vitro. This fusion gene may represent a robust molecular and therapeutic target in EOCRC.

Keywords: Early-onset colorectal cancer, Next generation sequencing, Fusion gene, *WWC1*

Student Number: 2013-30549

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Colorectal cancer (CRC) is the third most common cancer diagnosed worldwide and represents the second most common cause of all cancer-related deaths in Korea. The peak incidence of CRC is observed among those between 60 and 70 years old. However, up to 10% of all CRC occurs in individuals younger than 50 years of age (1). Recent epidemiological studies suggest that there has been a significant increase in CRC diagnosis among young adults, even though the overall incidence and mortality of CRC has decreased in recent years (2-4). This trend was not observed on the Surveillance, Epidemiology, and End Results (SEER) data of United States (5), but also in data from 20 European countries. In European data spanning 2004 to 2016, the CRC incidence increased 7.9% per year among subjects aged 20–29 years, 4.9% per year among those aged 30–39 years, and 1.6% per year among those aged 40–49 years (6). Notably, a different trend is seen in the CRC incidence of South Korea, which increased among those aged 25–29 years from 2010 to 2017, but decreased across all other age groups (7). The incidence of CRCs in young individuals is projected to increase by as much as 90% and 140%, respectively, by 2030 (8).

Early-onset CRC (EOCRC) is often characterized by more advanced stage, distal location (especially in the rectum), mucinous and poorly differentiated tumors with signet ring cells, and a poorer prognosis (9, 10). EOCRC is likely to represent a hereditary predisposition (4, 11, 12), but familial syndromes

account for only 20% of EOCRC. Among them, 10% of tumors have been found to have deficiencies in the DNA mismatch repair gene (dMMR) (13). Several studies have shown that 15%–21% of younger patients with CRC have tumors with dMMR (14-18). However, the genetic basis of dMMR differs markedly between EOCRC and later-onset CRC (LOCRC) and dMMR tumor has different implications in younger patients versus older patients (19). Although *MLH1* silencing is a major cause of microsatellite instable (MSI) tumors in LOCRC, MMR failure with aneuploidy accounts for up to 50% of microsatellite- and chromosome-instable tumors among EOCRC patients (8).

In the remaining 80% of sporadic EOCRC cases, several genetic etiologies have been reported (4). The carcinogenic pathway most frequently implicated in EOCRC is chromosomal instability (CIN) via the adenoma-carcinoma sequence (20). These tumors are microsatellite stable (MSS) and often have multiple alterations of chromosome number, chromosomal rearrangements, and/or gene amplification/deletion of oncogenes/oncosuppressors. Other pathways may be altered in EOCRC, including the WNT (*APC* and *CTNNB1* genes), *TP53, DCC, SMAD2/4,* and *KRAS* signaling pathways. In addition, microsatellite- and chromosome-stable (MACS) tumors, which account for up to 63% of EOCRC with MSS, are more common among early-onset MSS compared to late-onset MSS (21). A subset of EOCRC patients have a higher degree of hypomethylation at *long interspersed nucleotide element-1 (LINE-1)* sequences compared with those seen among LOCRC patients, and

hypomethylation at *LINE-1* has been linked with worse clinical outcomes (22).

A molecular analysis comparing MSS tumors between 39 cases of sporadic EOCRC (patients ≤ 45 years) and those from older patients identified 219 highly discriminatory genes between two groups (23). Due to these distinct clinicopathologic and genetic characteristics, previous studies have suggested that these EOCRC should be considered a unique molecular subgroup.

As evidence has accumulated regarding the molecular basis for CRC and new deep-sequencing methods have been developed, researchers have begun to study the potential involvement of fusion genes in CRC. Structural chromosome rearrangements, such as translocations, interstitial deletions, or chromosomal inversions, may form a chimeric gene from previously separate genes (24). This type of hybrid gene often generates a novel protein with functional capabilities that are distinct from those of its parental genes. Such fusion genes tend to be specifically found in neoplastic tissues; the efforts to unravel their diverse functionalities during carcinogenesis have provided critical clues regarding the disease mechanisms that are involved in tumorigenesis and have enabled some tumors to be subtly sub-classified (25). Since gene fusions are closely associated with specific tumor phenotypes derived from their unique chimeric domains, they are ideal targets for anti-cancer treatment and risk stratification (26). The fusion gene has also been associated with patient prognosis. For example, the overall survival rate of patients with metastatic CRC harboring *ALK, ROS*, and *NTRK* fusion genes was found to be lower than that of patients lacking these fusion genes (27).

Numerous studies have examined how the between-gene exchanges of coding or regulatory DNA sequences impact several types of cancer, including chronic myeloid leukemia (28, 29) and lung cancer (30, 31). Since 2014, there has been an exponential increase in the number of fusion genes known to be associated with various carcinomas, including CRC (Table 1) (25). Indeed, fusion genes may be found in up to 12.9% of metastatic CRCs and 8.8% in primary CRCs (32, 33). However, no previous study has examined the potential contribution of fusion genes to EOCRC.

In the present study, four cell lines were established from EOCRC patients and next-generation sequencing was used to identify their molecular and genetic features.

Table 1. Fusion genes found in colorectal cancer

Author, year	Patients	Fusion genes
Choi et al., 2018 (33)	147 CRC 47 matched normal tissue	*ANAAPC1-ZC3H8, APC-COMMD10, ATIC-UXS1, BOD1-WWC1, BOK-ING5, CHIC1-HDX, GTF3A-CDK8, LMNA-NTRK1, NAGLU-IKZF3, OSBPL10-GADL1, PTPRK-RSPO3, RAB1B-DPP3, RASA1-LOC644100, RIMS3-SCMH1, RNF121-FOLR2, STRN-ALK, TAF4-NDRG3, TMOD3-MAPK6, TMPRSS2-PDE9A, TPM3-NTRK1, TRIM24-BRAF, USP32-NSF, VTI1A-TCF7LR, ZYG11A-GPX*
Kloosterman et al., 2017 (34)	278 stage I to III colon cancer (Rotterdam MATCH study)	*TRIM24-BRAF, AGAP3-BRAF, DLG1-BRAF, EML4-NTRK3, RET-Coiled Coil Domain, USP9X-ERAS, EIF3E-RSPO3*

		LMNA/TPM3/SCYL3/ETV6-NTRK,
Pietrantonio et al., 2017 (27)	27 metastatic CRC	*CAD/EML4/CENPF/PRKAR1B/MAPRE3, STRN-ALK,*
		SLC34A2/GOPC/unknown-ROS1 fusions
Le Rolle et al., 2015 (35)	3117 metastatic CRC	*CCDC6-RET, NCOA4-RET*
Sehagiri et al., 2012 (36)	70 pairs of CRC	*EIF3E-RSPO2, PTPRK-RSPO3, PTPRK-RSPO3*

Abbreviation: CRC, colorectal cancer.

Materials and Methods

Establishment of cell lines and cell culture conditions

The analyzed cell lines were established using MSS CRC tissues of patients who had undergone surgery for CRC before 30 years of age. We excluded the tissues of patients who had been diagnosed with a hereditary CRC syndrome, such as Hereditary Non-Polyposis Colorectal Cancer (HNPCC) or Familial Adenomatous Polyposis (FAP), and those with a family history of CRC, a history of other malignancy, or an MSI tumor.

Cell lines were established from pathologically proven CRC, using the previously described protocol (37). Solid tumors were finely minced with scissors and dispersed into small aggregates by pipetting. Appropriate amounts of fine neoplastic tissue fragments were seeded to 25 cm^2 flasks. Most of the tumor cells were initially cultured in ACL-4 medium supplemented with 5% heat-inactivated fetal bovine serum (FBS; called AR5 medium). ACL-4 is a fully defined medium that was specifically formulated for the selective growth of human lung adenocarcinoma cells and has proven useful in the establishment of CRC and hepatocellular carcinoma cell lines. ADF5 medium, which comprised a 1:1 mixture of Dulbecco's modified Eagle's medium and Ham's F-12 supplemented with 5% heat-inactivated FBS, was also used for the initial culture of tumor cells. Cultures were maintained

in RPMI1640 supplemented with 10% heat-inactivated FBS (R10). Initial passages were performed when heavy tumor cell growth was observed, and subsequent passages were performed every 1 or 2 weeks. Adherent cells were recovered from sub-confluent cultures, trypsin treated, dispersed by pipetting, and used for passages. If stromal cell growth was noted in the initial cultures, differential trypsinization using a confiner was used to obtain a pure tumor cell population. After rinsing the medium with PBS, we repeated to add 1.5 mL of diluted trypsin (0.05%) in the culture medium, place it into incubator at 37℃ for 3 min, and harvest cancer cell. Cultures were maintained in humidified incubators at 37°C in an atmosphere of 5% CO2 and 95% air. The locations and stages of the original tumors used to generate the newly established EOCRC cell lines are listed in Table 2.

DNA fingerprinting analysis

DNA fingerprinting was performed to confirm that the cell line derived from the patient tissue coincides with the actual tissue of the patients. Total DNA was isolated from cell pellets using a QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. DNA was amplified using an AmpFlSTR Identifiler PCR amplification kit (Applied Biosystems, CA, USA) (38). A single round of PCR was used to amplify 15 short tandem repeat markers (CSF1PO, D2S1338, D3S1358, D5S818, D7S820, D8S1179,

D13S317, D16S539, D18S51, D19S433, D21S11, FGA, TH01, TPOX and vWA) and an amelogenin gender-determining marker at loci containing highly polymorphic microsatellite markers. Amplified products were analyzed using an ABI 3730 genetic analyzer (Applied Biosystems, MA, USA). Additionally, DNA was PCR amplified at loci containing the highly polymorphic microsatellite markers, D1S1586 and D3S1765. PCR products were denatured by 95% formamide and electrophoresed on a 7M urea polyacrylamide gel for 2 h at 60 W. Gels were dried and visualized by autoradiography. The short tandem repeat (STR) profiles of the established EOCRC cell lines are listed in Table 4.

Growth properties and morphology in vitro

Suspensions of 5 x 10^3 cells were seeded to 35 mm tissue culture dishes containing culture medium. Cells were counted in triplicate at 24 h intervals for 10 days, using the WST-1 assay. The doubling time of the cells was calculated from the growth phase. Mycoplasma contamination was tested by the 16S-rRNA-gene-based PCR amplification method, using an e-Myco Mycoplasma PCR detection kit (Intron Biotechnology, Gyenggi, Korea).

Nucleic acid isolation and complementary DNA synthesis

Genomic DNA was extracted from the cell lines using a QIAamp DNA Mini

Kit and RNA was extracted using the TRIzol reagent (Life Technologies, CA, USA) and an RNeasy Plus Mini Kit (Qiagen, Hilden, Germany) according to the manufacturers' protocols. For cDNA synthesis, a QuantiTect reverse transcription (RT) kit (Qiagen, Hilden, Germany) was used. One microgram of total RNA, 2 μL of gDNA Wipeout Buffer, and diethyl pyrocarbonate (DEPC) water up to 14 μL were combined and incubated at 42°C for 2 min. The mixture was mixed with Quantiscript RT buffer, RT primer mix, and Quantiscript reverse transcriptase, and incubated at 42°C for 45 min. Finally, the mixture was incubated at 95°C for 2 min and cooled to room temperature.

Whole-exome sequencing analysis

Total DNA was isolated from cell pellets using a QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's protocol. Exomic sequences were enriched in samples using the Illumina Exome Enrichment protocol (Illumina, CA, USA), and the captured libraries were sequenced using Illumina HiSeq 2000 Sequencers. The reads were mapped to the UCSC hg19 human genome assembly using Burrows-Wheeler Alignment (BWA version 0.5.8, http://bio-bwa.sourceforge.net/). Single nucleotide polymorphisms and small insertions/deletions were called with SAMtools (version 0.1.7, http://samtools.sourceforge.net/). The annotated variants were filtered against the variations reported in dbSNP132, the 1000 Genomes Project (November

2010 edition), and the NHLBI Exome Sequencing Project (http://evs.gs.washington.edu/EVS/) databases. The computational tools, SIFT (39) and PolyPhen-2 (40), were used to predict the impact of missense mutations, while MaxEntScan (41) and NNSplice (http://www.fruitfly.org/seq_tools/splice.html) were used to predict the effects of splice variants.

RNA-sequencing analysis

Total RNA was isolated from cell lysates using TRIzol (Qiagen, Hilden, Germany) and a Qiagen RNeasy Kit (Qiagen, Hilden, Germany). Sequencing libraries were prepared using an Illumina TruSeq Stranded Total RNA Library Prep Kit. Fifty-one million reads were obtained from the cell lysates. Base-calling and alignment were performed using the Tuxedo Suite, and the rejected reads were analyzed using FusionMap, ChimeraScan, and Defuse, with default parameters for RNA and alignment to GRCh37.72. The output was filtered to include in-frame fusions, with at least one rescued read and two unique seed reads and excluding known and recurrent artifacts. RNA isolated from the EOCRC cell lines was sequenced and identified RNA sequences spanning two different regions in a chromosome. The in-frame fusion with the highest read score spanned the *FAM174A* gene on chromosome 5q21.1 and WWC1 on chromosome 5q34. The exon Reads Per

Kilobase of transcript per Million mapped reads (RPKM) track displays the RPKM calculated for each exon, with the library size normalization factor considered as the number of reads aligned to exons alone. The fusion read spanning exon 2 of *FAM174A* and exon 10 of *WWC1* was also confirmed with Integrative Genomics Viewer (IGV) version 2.0 (Broad institute, CA, USA).

Verification of FAM174A-WWC1 in cancer cell lines

Based on previously described gene fusion isoforms involving the *FAM174A* and *WWC1* genes, we used Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primerblast) to design oligonucleotide primers to encompass the breakpoint junctions of *FAM174A-WWC1* (*FAM174A* Forward: GTC AGG ACG GTC AGG ATG AGA, *WWC1* Reverse: GCT TCC TCC AGT TCT CTC ACA A). PCR was performed using an Intron PCR Kit (Intron, Seoul, Korea) and the following cycling conditions: 94°C for 5 min; 35 cycles of 94°C for 30 sec, 55°C for 45 sec, 72°C for 1 min; and a final extension of 72°C for 10 min. RT-PCR performed using primers complimentary to sequences in exon 2 of *FAM174A* and exon 10 of *WWC1* yielded a product whose sequence suggested the existence of an in-frame junction.

Cloning of FAM174A-WWC1

After confirming the FAM174A-WWC1 fusion gene by RT-PCR, we designed cloning primers to amplify the complete coding sequence of the *FAM174A-WWC1* fusion gene (*FAM174A* Forward: CAC CAT GAA GGC CTC GCA GTG CTG C, *WWC1* Reverse: GAC GTC ATC TGC AGA GAG AGC TGG GAT). PCR was performed using Q5® Hot Start High-Fidelity Taq (NEB) and the following cycling conditions: 94°C for 5 min; 3 cycles of 94°C for 30 sec, 72°C for 3 min; and a final extension of 72°C for 10 min. The amplified *FAM174A-WWC1* sequence was cloned into the pENTR/SD/D/TOPO vector (Life Technologies, CA, USA) according to the manufacturer's recommendations. The fusion junction was verified by Sanger sequencing. The entry vector was then subcloned into the plenti6.2/V5-DEST lentiviral expression vector (Life Technologies, CA, USA) according to the manufacturer's instructions.

Immunocytochemistry

To identify the location and possible function of the novel fusion gene, *FAM174A-WWC1*, immunocytochemistry was performed. Cells were seeded on a chambered cover glass and grown to the desired confluency. The chambered cover glass was designed to be hydrophilic and was not treated with any extracellular matrix (ECM) component prior to seeding. After 72 h, the cells were washed three times with cold Dulbecco's phosphate-buffered

saline (DPBS) for 15 min, and then fixed and permeabilized with BD Cytofix/Cytoperm™ (BD bioscience, NJ, USA). Cells were washed with washing solution, blocked for 1 h with DPBS containing 2% FBS, washed with cold DPBS, and incubated with *WWC1* and V5 Tag antibodies diluted in 0.05% PBS.T for 1.5 h at room temperature. The cells were then washed with 0.05% PBS.T, incubated with Alexa 488 and Alexa 568 secondary antibodies diluted in 0.05% PBS.T for 1 h at room temperature, and incubated for 30 min with 1x DAPI and rhodamine-conjugated phalloidin diluted in distilled water. The cells were washed three times with DPBS and imaged under an LSM800 Confocal Laser Scanning Microscope (Zeiss, Berlin, Germany) with data analysis performed using the ZEN software (Zeiss, Berlin, Germany). The intensity of each channel was fixed to compare the target protein expression between the samples. The digital resolution, scan speed, and average number of images were set to 1024 x 1024, 40 sec per channel, and 8, respectively. The images were focused at the very bottom of each fixed cell, which enabled us to investigate the protruding regions of the cell.

2D and 3D migration assays

To identify the morphologic and functional changes exhibited by cells transduced with the *FAM174A-WWC1* fusion gene, a 2D gap-closure migration assay was performed using a CytoSelect™ 24-Well Wound Healing

Assay kit (Cell Biolabs, CA, USA) according to the manufacturer's protocol., NIH3T3 (a mouse embryonic fibroblast cell line) Control and NIH3T3 *FAM174A-WWC1* cells were diluted to 2 x 10^5 cells/mL in RPMI1640 containing 10% FBS and 1% penicillin streptomycin. The kit-provided inserts were placed into each well of the 24-well tissue culture plate, and 500 µl of the diluted cell solution was dispensed to each side of the insert to obtain a total cell solution of 1 mL per well. After 24 h of incubation at 37°C in an atmosphere of 5% CO2 and 95% air, the inserts were removed. Immediately, the cells within the wells, which had adhered with an insert-generated gap between the sides were washed twice with complete medium. The wound closure was recorded under a phase-contrast microscope at 24 h intervals for 5 days. The closed wound area was calculated using ImageJ version 1.8.0 (National Institutes of Health, MD, USA). The whole procedure was performed in triplicate, and the average closed wound area was analyzed using GraphPad Prism version 7.00 (GraphPad Software, CA, USA) for Windows.

The 3D migration assay was performed using AIM 3D cell culture chips (AIM BIOTECH, Ayer Rajah, Singapore) according to the manufacturer's protocol. Briefly, cells were diluted to 2 x 10^5 cells/mL in RPMI1640 without FBS. The cell solution was mixed at a 1:1 ratio with reduced growth factor basement membrane matrix, and 10 µL of the mixture was used to fill each

gel inlet channel. The apparatus was incubated for 30 min in a 37°C incubator, the media channels were hydrated with 120 μL of RPMI1640 mixed with 10% FBS and 1% penicillin streptomycin, and further incubation was performed for 24 h at 37°C. At 24 h intervals, visual inspection was performed and images were obtained under a phase-contrast microscope using a 4X objective. Images were analyzed using ImageJ version 1.8.0. The entire procedure was performed in triplicate, and the average closed wound area was analyzed using GraphPad Prism version 7.00 for Windows.

3D spheroid formation assay and 3D spheroid cell invasion assay

To examine whether the *FAM174-WWC1* fusion gene has an oncogenic effect, 3D spheroid formation and 3D spheroid cell invasion assays were performed. The 3D spheroid formation assay was performed using a Cultrex® 96-Well 3D Spheroid BME Cell Invasion Assay kit (Trevigen, MD, USA) according to the manufacturer's protocol. Briefly, the reduced growth factor basement membrane matrix was thawed on ice overnight in a 4ºC refrigerator. HEK293 cell (a human embryonic kidney cell line) aggregates were dissociated to a single-cell suspension using trypsin. Cells were diluted to 1 x 10^6 cells/mL in RPMI1640 supplemented with 10% FBS and 1% penicillin streptomycin. The cell solution was mixed with the reduced growth factor basement membrane matrix, and 50 μL of the mixture was dispensed to each well of the provided

3D culture-containing 96-well spheroid formation plate. The plate was centrifuged at 200 x g for 3 min at room temperature in a swinging-bucket rotor and incubated at 37ºC in a tissue culture incubator. After 24 h, 50 µL of the provided invasion matrix was added per well, on ice. The plate was centrifuged at 300 x g at 4ºC for 5 min in a swinging-bucket rotor to eliminate bubbles and position spheroids within the invasion matrix towards the middle of the well. The plate was transferred to a tissue culture incubator set at 37ºC for 1 h to promote gelation of the invasion matrix. Images were obtained at 72 h after transfer to invasion matrix using a phase-contrast microscope equipped with a 4X objective.

The 3D spheroid cell invasion assay was conducted by replacing the reduced growth factor basement membrane matrix with 10X Spheroid Formation ECM (provided) to artificially induce spheroid formation and substituting the 96-well spheroid formation plate for the 96 hang-in-drop plate to retain cell aggregation. The formed spheroids were transferred to a 96 U-bottom plate containing 50 µL/well of invasion matrix (provided). After 1 h, 100 µL of warm (37ºC) chemoattractant-containing cell culture medium was added. The spheroid in each well was photographed every 24 h for 5 days under a phase-contrast microscope using a 4X objective. Images were analyzed using ImageJ version 1.8.0. All procedures were performed in triplicate, and the intact spheroids and disseminated cells were analyzed by circularity

estimation using Plugin from GraphPad Prism version 7.00 for Windows.

Results

General characteristics of the cell lines

Clinicopathologic information regarding the tumors used to derive the new cell lines is provided in Table 2. On *in vitro* cultivation, three of the newly generated cell lines (NCC-375, SNU-1460, and SNU-1826) grew as monolayers of substrate-adherent cells, and one (SNU-2446) formed floating and adherent aggregates. Most of the tumor cells were polygonal in shape and exhibited round-to-oval nuclei with prominent single-to-double nucleoli (Fig. 1). Each cell line was passaged at least three times prior to being analyzed for the following. The population-doubling times ranged from 35 to 82 h (Table 3). All cell lines were confirmed to be free of bacterial and mycoplasma contamination (Fig. 2). The 15 studied tetranucleotide repeats and the gender-determining marker, amelogenin were heterogeneously distributed in each cell line, without cross-contamination (Table 4).

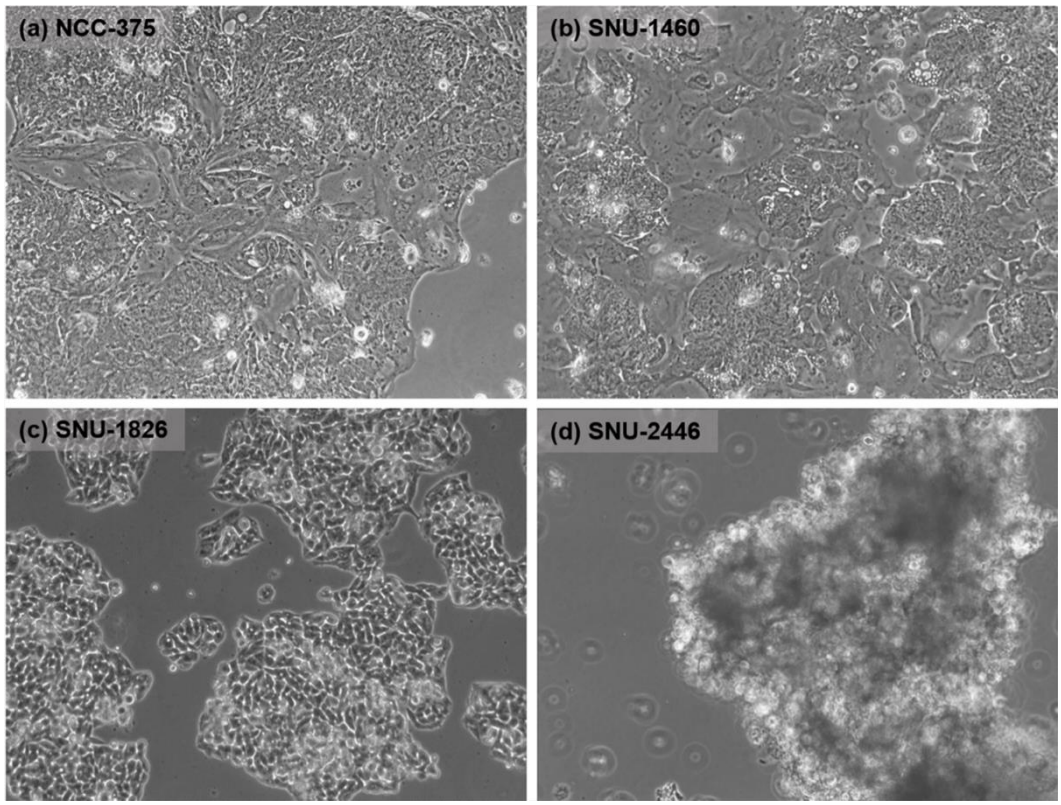


Figure 1. **Phase-contrast microscopic images of four established early-onset colorectal cancer cell lines, imaged at 70% confluence (x100).** On *in vitro* cultivation, most of the tumor cells were polygonal in shape and had round-to-oval nuclei with prominent single-to-double nucleoli. Three of the cell lines (NCC-375, SNU-1460 and SNU-1826) grew as monolayers of substrate-adherent cells, while one (SNU-2446) formed floating and adherent aggregates.

Table 2. Clinicopathologic characteristics of the patient sources of the four early-onset colorectal cancer cell lines

	Sex	Age	Family history	Degree of differentiation	Stage	Pathology	Location	Recurrence
NCC-375	F	24	None	M/D	T3N2M1	Adc	Rectal cancer with liver and lung metastasis	Recur
SNU-1460	M	25	None	M/D	T4aN0M0	Adc	Sigmoid colon cancer	NED
SNU-1826	M	15	GF (stomach ca.)	P/D	TxN1M0	Adc	Cecal cancer	NED
SNU-2446	F	21	None	P/D	T4bN2M1	Mucinous carcinoma	Descending colon cancer with peritoneal seeding	PD

Abbreviations: M/D, moderately differentiated; P/D, poorly differentiated; Adc, adenocarcinoma; GF, grandfather; NED, no evidence of disease; PD, progression of disease

Table 3. *In vitro* growth and morphology of the generated cell lines

	Growth pattern	Doubling time (h)	Cell morphology
NCC-375	Adherent	35	Fusiform
SNU-1460	Adherent	48	Polygonal
SNU-1826	Adherent	45	Polygonal
SNU-2446	Adherent/ floating	82	Round/ polygonal

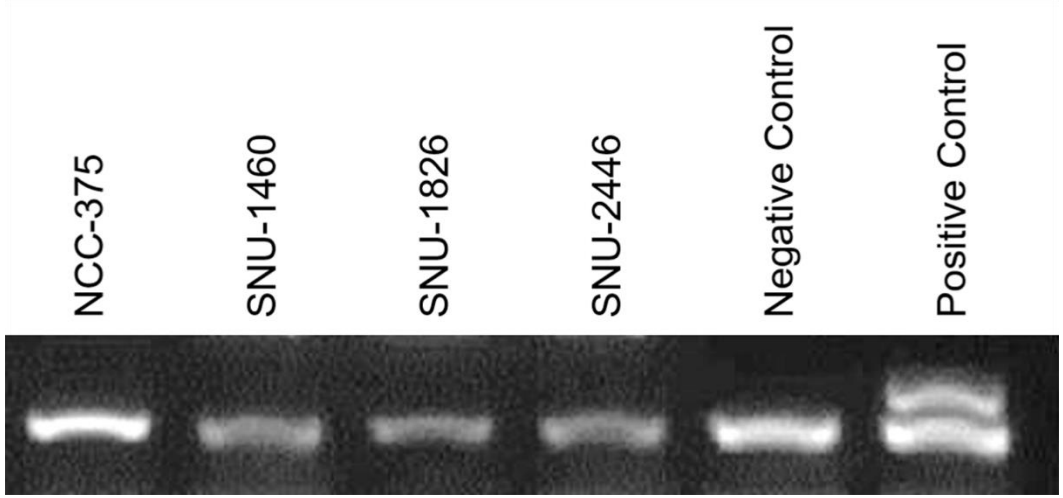


Figure 2. Mycoplasma test. The four newly established early-onset colorectal cancer cell lines were free of bacterial contamination.

Table 4. Short tandem repeat profiles of the four early-onset colorectal cancer cell lines

Cell line	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539
NCC-375	13,15	32.2,33.2	10,12	10,12	18	7	9,12	10
SNU-1460	13	30.2,31	8,11	12	15,17	6,7	9,12	10
SNU-1826	13	32,32.2	9,10	10	15	6	8,11	9,11
SNU-2446	12	29,34	9,10	10,12	15,17	9	11	13

Cell line	D2S1338	D19S433	Vwa	TPOX	D18S51	Amelogenin	D5S818	FGA
NCC-375	17,23	13,13.2	18	11	15	X	12,17	20,23
SNU-1460	23,24	13,14.2	14,18	8,11	13	X	11	21.23
SNU-1826	17	11.2,14	17,18	8	15	X	11	22
SNU-2446	18,20	13	14	9,11	19	X	12	24

Genomic analyses

Due to the uncharacteristically aggressive clinical course and lack of the prototypical *FAM174A-WWC1* fusion in the tissues from which the new cell lines were derived, the histological diagnosis was in question. To more comprehensively characterize the molecular basis of EOCRC, the four newly established EOCRC cell lines were subjected to whole-exome and transcriptomic analyses. Our analysis of mutations previously associated with the carcinogenic pathway of CRC identified only a single mutation in one of the cell lines: a splice mutation in *TP53* in SNU1460. We did not find any meaningful mutation in *APC, KRAS, BRAF, TNF, MAPK, VEGF, PIK3CA, or CTNNB1* (Table 5). In order to integrate the whole-exome sequencing data, we obtained from our EOCRC cell lines with a larger database, we extracted the top 100 mutated genes in COlon ADenocarcinoma (COAD) cohorts under age 35, using The Cancer Genome Atlas-Genomic Data Commons (TCGA-GDC) data portal. Genes found in both the TCGA-GDC results and the EOCRC patients were sorted, and mutations predicted to affect the protein structure (e.g., frameshift or nonsense mutations) were screened. The *PCLO* gene was mutated in all four of our new EOCRC cell lines and four COAD samples obtained from TCGA-GDC. Mutations in the *PCLO* gene have been associated with depressive disorders, but the relationship between the *PCLO* gene and EOCRC will require further analysis. The results of our mutational analysis are presented in Tables 6-9.

Table 5. Mutational profiles of the new EOCRC cell lines at genes previously associated with CRC

	NCC-375	SNU-1460	SNU-1826	SNU-2446
KRAS				
BRAF				
TP53	P72R[c]	splice site donor[a]	P72R[c] C238Y[d]	
PIK3CA				
CTNNB1				
APC	V1822D[b]	V1822D[b]	V1822D[b]	V1822D[b]
MAPK				
TNF				
EGF	M708I[c] E920V[c]	M708I[c] E920V[c]	M708I[c] E920V[c] R431K[c] D784V[c]	M708I[c] E920V[c]
VEGF				
MLH1			exon1:c.G46C :p.V16L[d]	
MLH3	N826D[c]	N826D[c] V1050I[c]	N826D[c] P844L[c]	
MSH2				
MSH3	I79V[c] Q949R[c]		I79V[c] Q949R[c] A1045T[c]	I79V[c] Q949R[c] A1045T[c]
MSH6			G39E[b]	V1822D[b]
PMS1	D115G[d]			
PMS2	K541E[c] P470S[c]	K541E[b]	K541E[c] P470S[c] G857A[c]	
PMS6				
EXO1	H354R[c] E589K[c] E670G[c] R723C[c] P757L[c]	H354R[c] E589K[c] E670G[c] R723C[c] P757L[c]	H354R[c] E589K[c] E670G[c] R723C[c] P757L[c]	
EPCAM	M143T[b] T200M[b]	M143T[b]	M115T[b]	

[a]Pathogenic mutation, [b]benign mutation, [c]unknown mutation, [d]unreported mutation

Table 6. **Mutations shared between NCC-375 and the TCGA-GDC dataset**

GENE	CHR	POS	ID	REF	ALT	AA_CHANGE
MUC4	3	195518112	rs142781032	T	TGTCTCCTGCGTAACA	p.T113MLRRRP
FNDC1	6	159660779	rs141435210	GCCACCACCCGCCGCACGA	G	p.ATTRRTT1471A
PCLO	7	82581488	rs10630259	T	TTCA	p.E2927VK
SYNE2	14	64560092	rs2781377	G	A	p.W4001*

Abbreviations: CHR, chromosome; POS, position; REF, reference; ALT, alteration; AA_CHANGE, amino acid change

Table 7. **Mutations shared between SNU-1460 and the TCGA-GDC dataset**

GENE	CHR	POS	ID	REF	ALT	AA_CHANGE
FNDC1	6	159660779	rs141435210	GCCACCACCCGC CGCACGA	G	p.ATTRRTT1471A
THSD7A	7	11871469		A	AGCAGCG	p.35RC
PCLO	7	82581488	rs10630259	T	TTCA	p.E2927VK
NCOR2	12	124824721	rs61519723	C	CGCCGCTGCT	p.G1846GAAA
NCOR2	12	124887058	rs35831183	G	GGCT	p.511S
ZFHX3	16	72821593	rs374416547	AGCCGCCGCC	A	p.GGG3525Del
ZFHX3	16	72822563		T	TTGCTGCTGC	p.Q3204RSSK
TP53	17	7578370		C	T	n/a

Abbreviations: CHR, chromosome; POS, position; REF, reference; ALT, alteration; AA_CHANGE, amino acid change

Table 8. Mutations shared between SNU-1826 and the TCGA-GDC dataset

GENE	CHR	POS	ID	REF	ALT	AA_CHANGE
MUC4	3	195518118	rs71180965	-	TGCGTAACAG TCTCC	p.M111METVTQ
FNDC1	6	159660804	rs3842694	CCCGCCGCAC GACCACCA	-	DelCCCGCCGC A CG(ACC)2A
THSD7A	7	11871487	.	GGCAGCGGCA GC	-	p.25_29del
PCLO	7	82581493	rs10694231	-	ATC	p.D2926DD
FRMD4A	10	13698938	.	-	CGC	p.R884RG
NCOR2	12	124824739	rs61519723	-	GCCGCTGCT	p.S1834SSGS
NCOR2	12	124887095	.	-	CTG	p.Q499QQ
ZFHX3	16	72822586	.	-	CTG	p.Q2283QQ
ZFHX3	16	72831382	.	TTGTTG	-	p.818_819del
RTTN	18	67863854	rs58913700	-	TCC	p.G242GG

Abbreviations: CHR, chromosome; POS, position; REF, reference; ALT, alteration; AA_CHANGE, amino acid change

Table 9. Mutations shared between SNU-2446 and the TCGA-GDC dataset

GENE	CHR	POS	ID	REF	ALT	AA_CHANGE
TTN	2	179472617	.	C	A	p.E17633*
PCLO	7	82581488	rs10694231	T	TTCA	p.E2927VK
WDFY4	10	50165232	.	C	A	p.S2679*

Abbreviations: CHR, chromosome; POS, position; REF, reference; ALT, alteration; AA_CHANGE, amino acid change

Identification and verification of the FAM174A-WWC1 fusion gene

The fusion genes were found by three different methods and selected one of which was fused in the exon-exon boundary. Then, we screened the selected fusion genes and identified seven candidate fusion genes (Fig. 3 and Table 10).

The novel fusion gene, *FAM174A-WWC1* was detected in the NCC-375 cell line. The reads spanning the breakpoint of the fusion gene were visualized with an IGV program, which revealed a 67,892,563 bp deletion in chromosome 5 (Fig. 4a). Oligonucleotide primers were designed to encompass the breakpoint junctions of *FAM174A* and *WWC1*. Using primers complimentary to the sequence in exon 2 of *FAM174A* and exon 10 of *WWC1*, we performed reverse transcription followed by PCR (RT-PCR) of the tumor-derived cell line. Sequencing of the generated product was used to identify the in-frame junction (Fig. 4b). The same primers were used to amplify DNA from the three other cell lines, but only non-specific PCR products were generated, as confirmed by Sanger sequencing. The amplicon containing the target fusion gene was verified by direct Sanger sequencing, which showed that exon 2 of *FAM174A* was connected to exon 10 of *WWC1* (Fig. 4c).

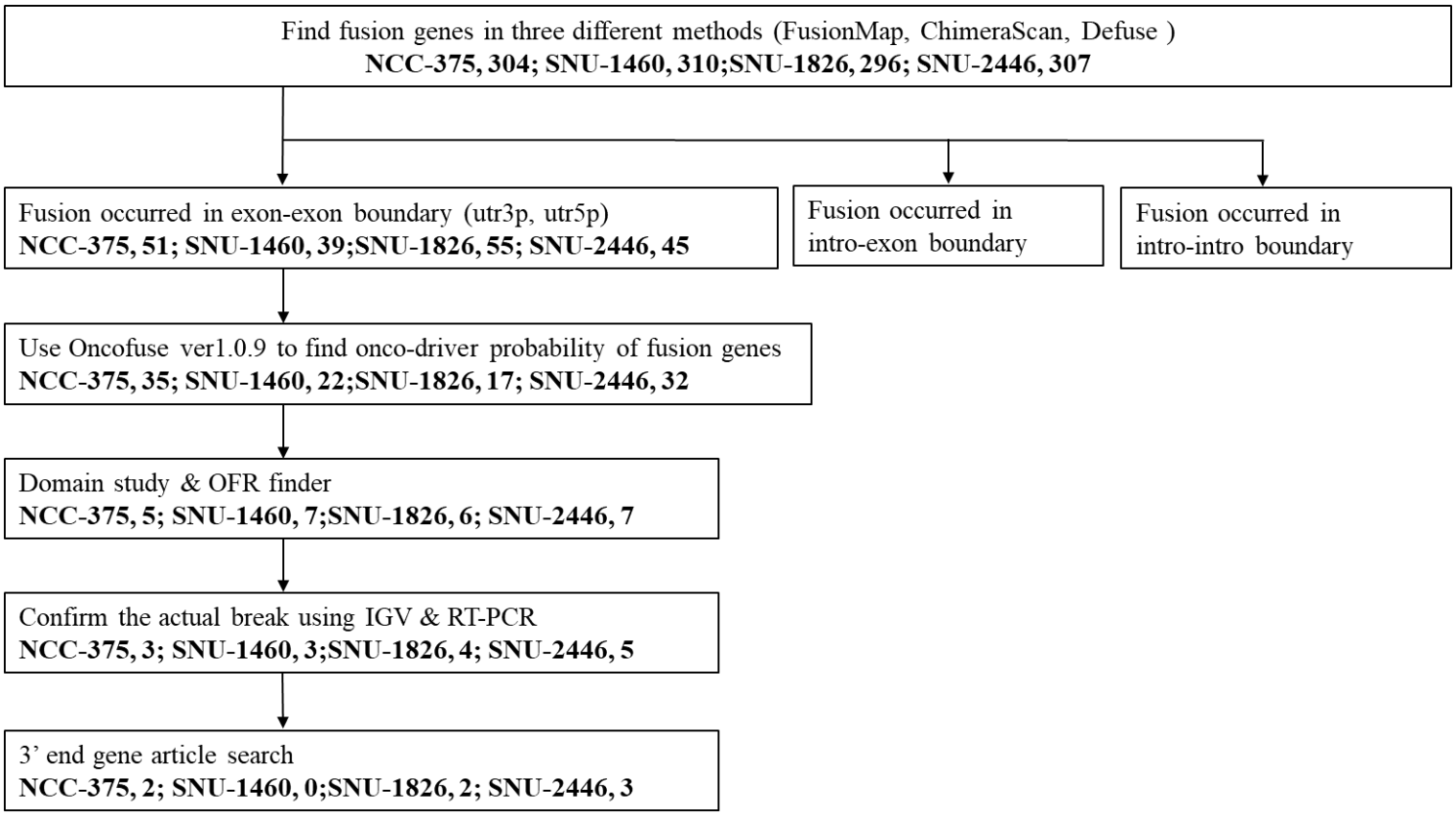
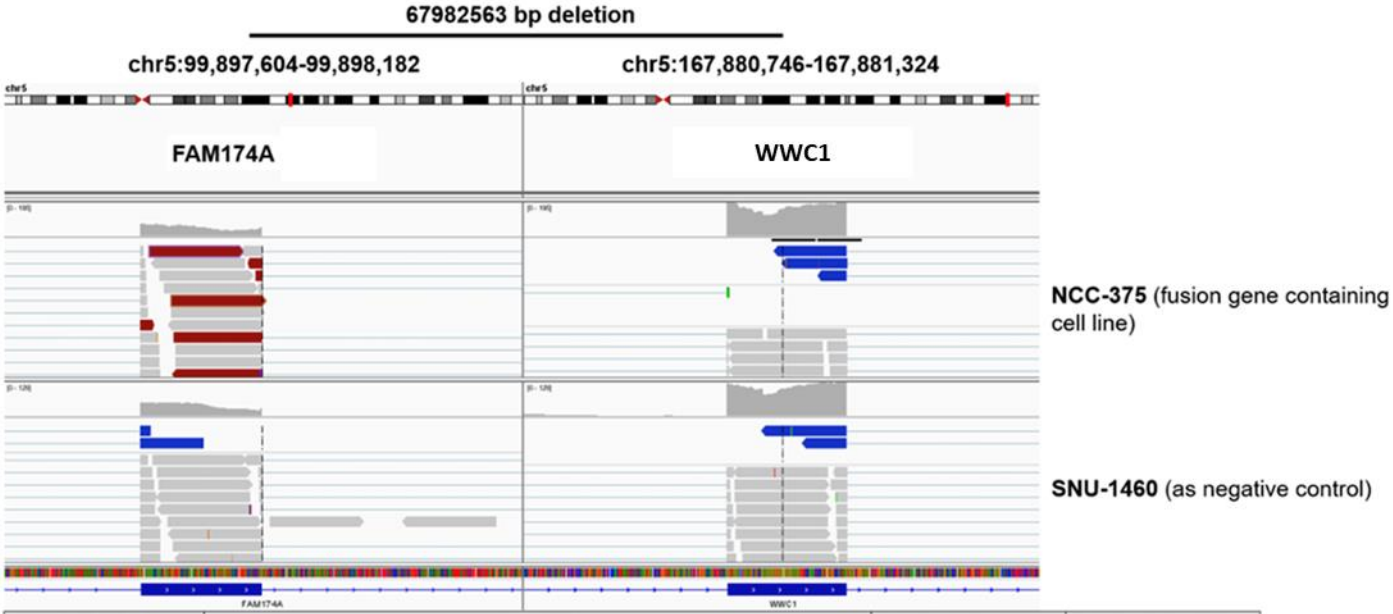


Figure 3. Fusion gene detection algorithm

Table 10. Candidate fusion genes identified by RNA sequence analysis

Cell lines	Genomic	5' end gene name	3' end gene name
NCC375	chr5:99897892>chr5:167849013	*FAM174A*	*WWC1*
	chr13:99738659>chr13:32971034	*DOCK9*	*BRCA2*
SNU1826	chr2:10269281>chr2:10281981	*RRM2*	*C2orf48*
	chr9:133499098>chr9:133541995	*FUBP3*	*PRDM12*
SNU2446	chr17:40985367>chr17:41322290	*BECN1*	*BRCA1*
	chr15:79241916>chr15:79383115	*CTSH*	*RASGRF1*
	chr1:153585621>chr1:153606873	*S100A16*	*S100A13*



SAMPLE_ID	GENOMIC	5_FPG_GENE_NAME	3_FPG_GENE_NAME
375.DEFUSE.txt	chr5:99897892>chr5:167849013	**FAM174A**	**WWC1**
375.FUSIONMAP.txt	chr5:99897892>chr5:167849013	**FAM174A**	**WWC1**
375.CHIMERASCAN.txt	chr5:99897891>chr5:167849012	**FAM174A**	**WWC1**

Figure 4. **Verification of *FAM174A-WWC1* in the NCC-375 cell line and the corresponding patient tumor sample**.

a. The reads spanning the breakpoint of the novel fusion gene, *FAM174A-WWC1*, were visualized with the IGV program.

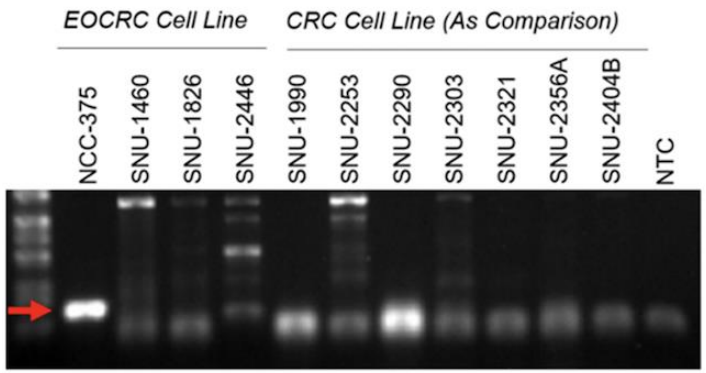


Figure 4. **Verification of *FAM174A-WWC1* in the NCC-375 cell line and the corresponding patient tumor sample.**

b. Oligonucleotide primers were designed to encompass the breakpoint junctions of *FAM174A* and *WWC1*. The amplicon with the specific oligonucleotide primers was found only in NCC-375 cell line.

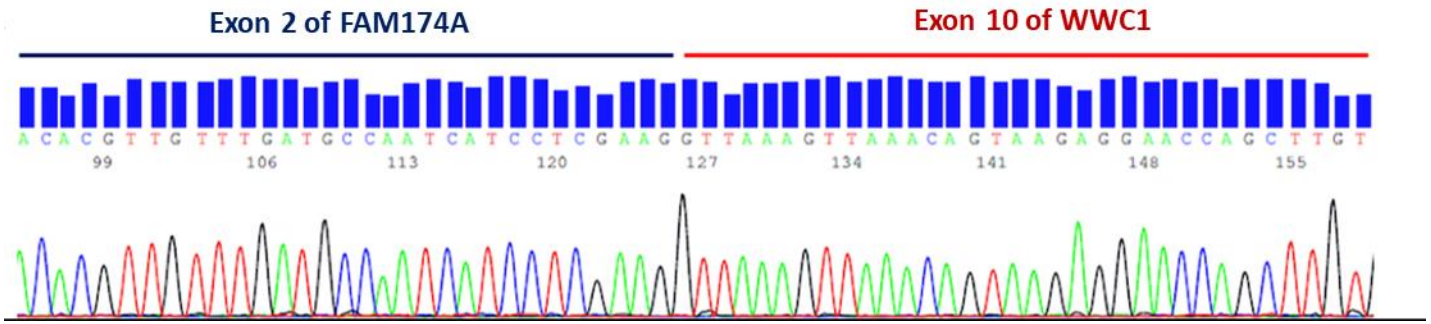


Figure 4. **Verification of *FAM174A-WWC1* in a cell line and the corresponding patient tumor sample. *c*.** Sanger sequencing of the *FAM174A-WWC1* fusion gene. The amplicon containing the target fusion gene was verified with direct Sanger sequencing, which showed that exon 2 of *FAM174A* is connected to exon 10 of *WWC1*.

Cloning and in vitro assessment of FAM174A-WWC1

We employed RT-PCR to clone the entire coding sequence of the *FAM174A-WWC1* fusion transcript from the NCC-375 cell line. In silico analysis of the *FAM174A-WWC1* coding sequence (http://web.expasy.org/compute_pi/) predicted the expression of a ～123 kDa protein. Examination of the *WWC1*-encoding sequence revealed that the truncation of *WWC1* resulted in the loss of its WW domains and preservation of its C2 and PDZ binding motifs. The truncation of the *FAM174A* portion in *FAM174A-WWC1* resulted conserved exons 1 and 2; these exons encode the DUF1180 domain, whose function is not yet known (Fig. 5a). Murine fibroblast NIH3T3 cells and HEK293 cells were transduced with packaged lentiviruses encoding *FAM174A-WWC1*. The expression of the fusion transcript was verified by RT-PCR with primers spanning the fusion breakpoint (Fig. 5b). The expression of fusion protein was assessed by Western blotting using a V5-tag antibody, which demonstrated a protein at the predicted molecular weight of 100 kDa (Fig. 5c). The expression of FAM174A-WWC1 altered the morphology of both cell lines from stellate to round and polygonal (Fig. 5d), decreased the protein expression of E-cadherin and augmented that of N-cadherin in HEK293 cells, and decreased the protein expression of E-cadherin in NIH3T3 cells (there was no change in N-cadherin, but this protein is not normally expressed in this cell line) (Fig. 5e).

Subcellular fractional protein extraction indicated that the fusion of WWC1 with the DUF1180 domain of *FAM174A* increased the nuclear levels of total and phosphorylated YAP1 (Fig. 5f). The levels of total YAP1 (Fig. 5g) and phosphorylated YAP1 (Fig. 5h) in the cytosolic and nuclear fractions were quantitatively analyzed, and the changes were found to be statistically significant.

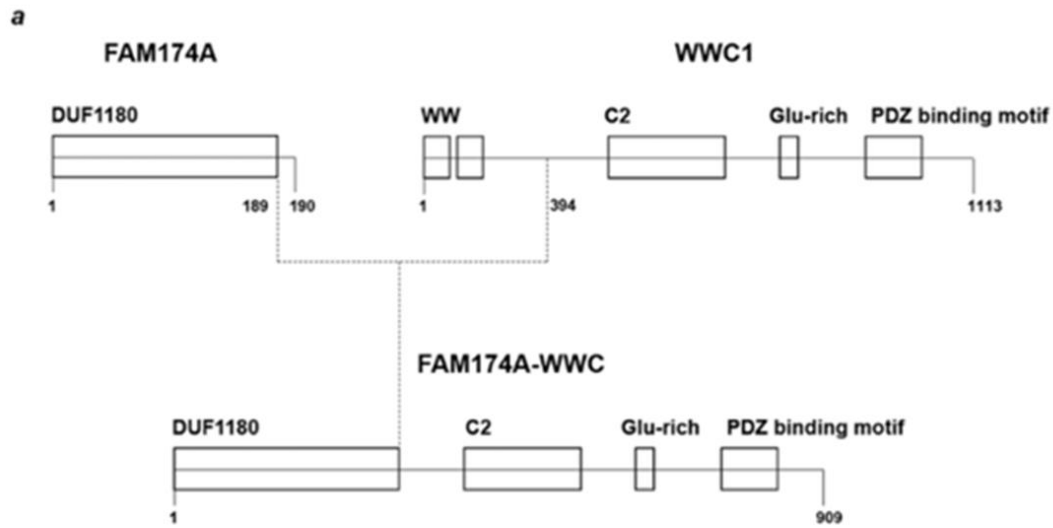


Figure 5. **Cloning and *in vitro* assessment of *FAM174A-WWC*. *a*.** Schematic diagram showing the structure of *FAM174A-WWC1*. The expression of the fusion transcript was verified by RT-PCR.

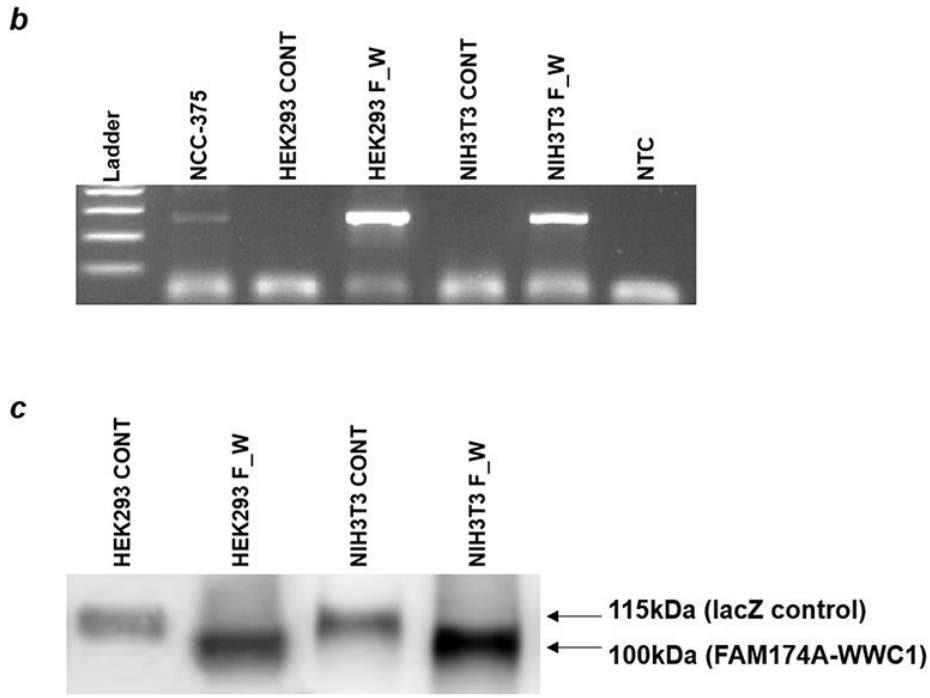


Figure 5. **Cloning and *in vitro* assessment of *FAM174A-WWC* with *b*.** primers spanning the fusion breakpoint and ***c*.** Western blotting using a V5-tag antibody, which reacted with a protein at the predicted molecular weight of 100 kDa. (continue)

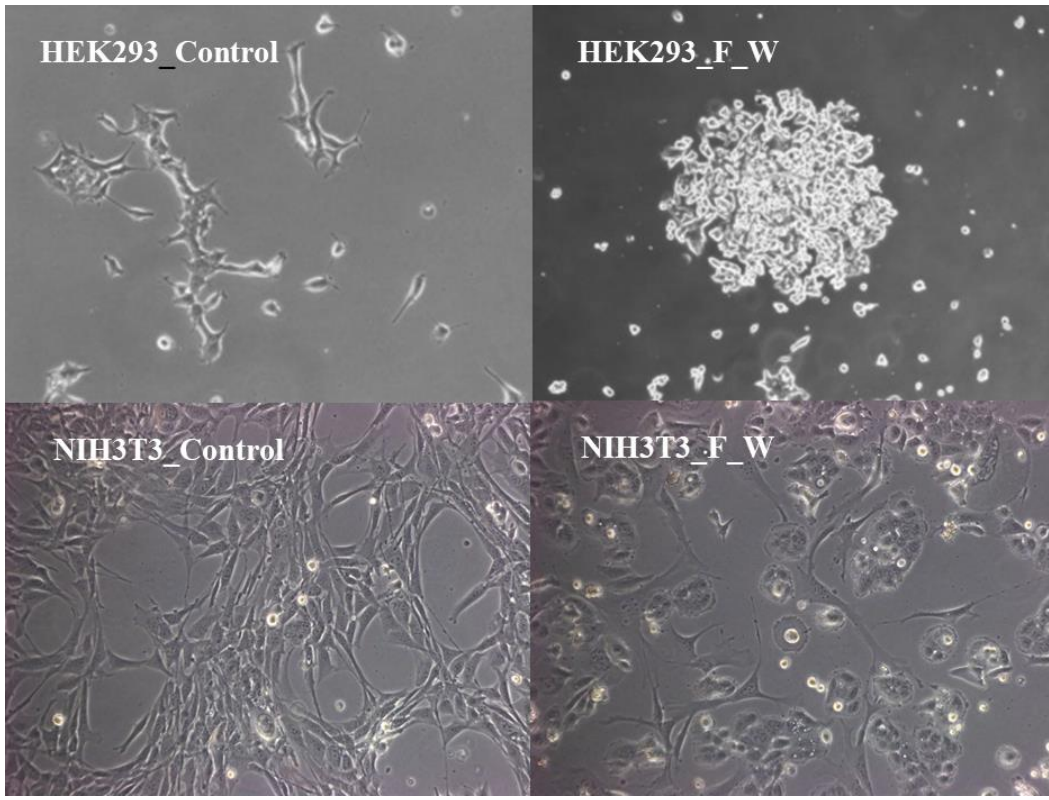


Figure 5. **Cloning and *in vitro* assessment of *FAM174A-WWC*.** ***d.*** The induction of *FAM174A-WWC1* caused morphologic alterations in both HEK293 and NIH3T3 cells with magnification of x100. (continue)

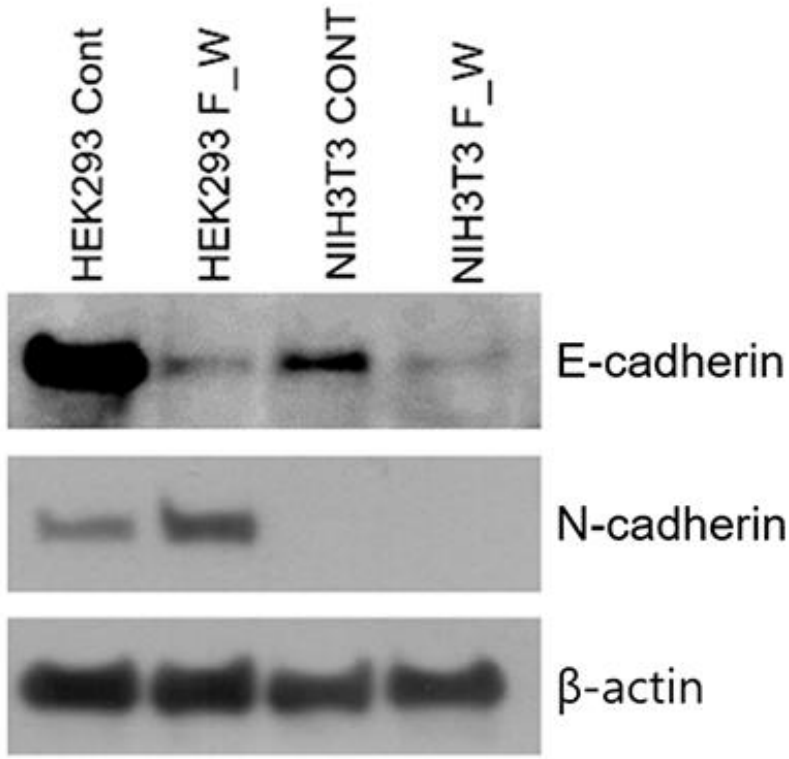


Figure 5. **Cloning and *in vitro* assessment of *FAM174A-WWC*. *e.*** Expression of *FAM174A-WWC1* decreased the protein expression of E-cadherin and augmented that of N-cadherin in HEK293 cells. (continue)

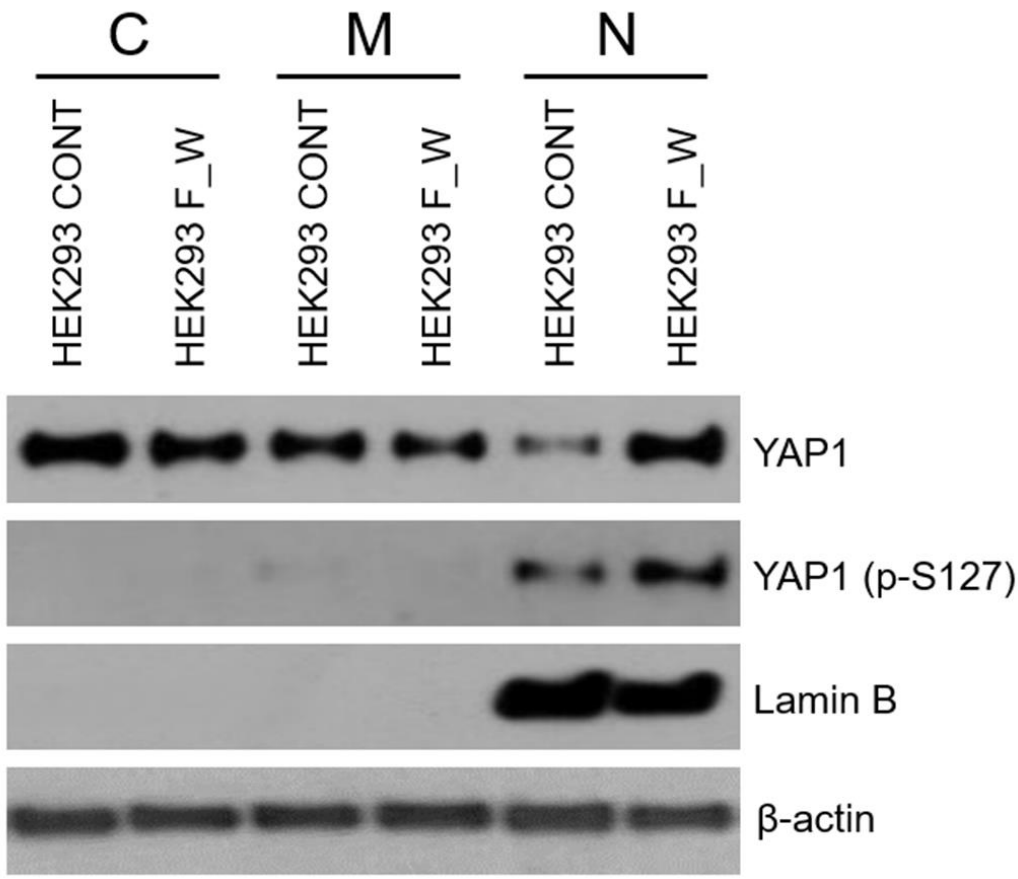


Figure 5. **Cloning and *in vitro* assessment of *FAM174A-WWC*.** ***f.*** Subcellular fractional protein extraction indicated that *WWC1* fused with the DUF1180 domain of *FAM174A* increased the levels of both total and phosphorylated YAP1 in nucleus. (continue)

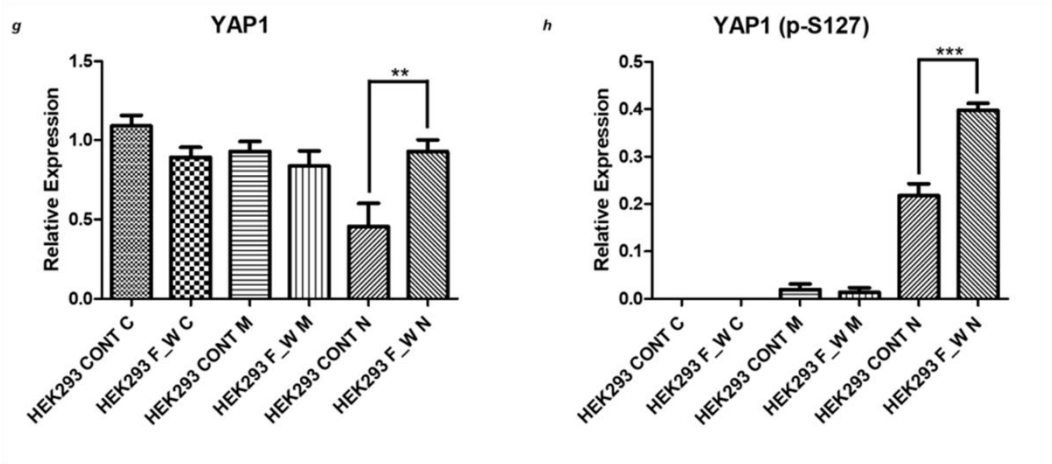


Figure 5. **Cloning and *in vitro* assessment of *FAM174A-WWC*. *g*.** Quantitative analysis of YAP1 expression in the cytoplasm, membrane, and nucleus. ***h*.** Quantitative analysis of YAP1 (p-S127) expression in the cytoplasm, membrane, and nucleus.

Oncogenic capacity of FAM174A-WWC1

The single nucleotide variation (SNV) profiling of EOCRC patients suggested that the fusion gene could have oncogenic potential. Moreover, we noted that the tumor expressing *FAM174A-WWC1* carried the fewest oncogene mutations (Fig. 6a). To further elucidate the tumorigenic capacity of *FAM174A-WWC1*, cells were grown under anoikis-promoting conditions to promote the formation of tumor-spheres, which enabled their use to examine the tumor-initiating capabilities and stem-like properties of the tumor cells. Indeed, our results indicated that *FAM174A-WWC1* expression significantly increased the tumor-sphere propagation of early-onset colorectal cancer cells (Fig. 6b).

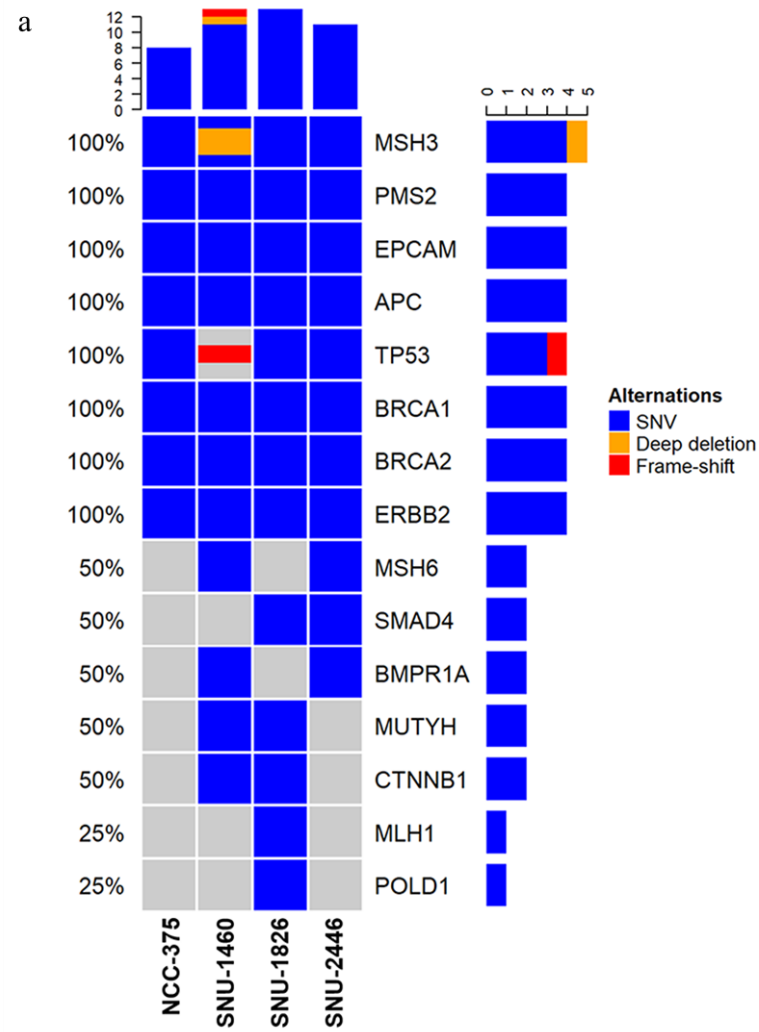


Figure 6. **Oncogenic capacity of *FAM174A-WWC1*. *a*.** SNV profiling of the enrolled EOCRC patients. The tumor expressing *FAM174A-WWC1* had the fewest oncogene mutations. (continue)

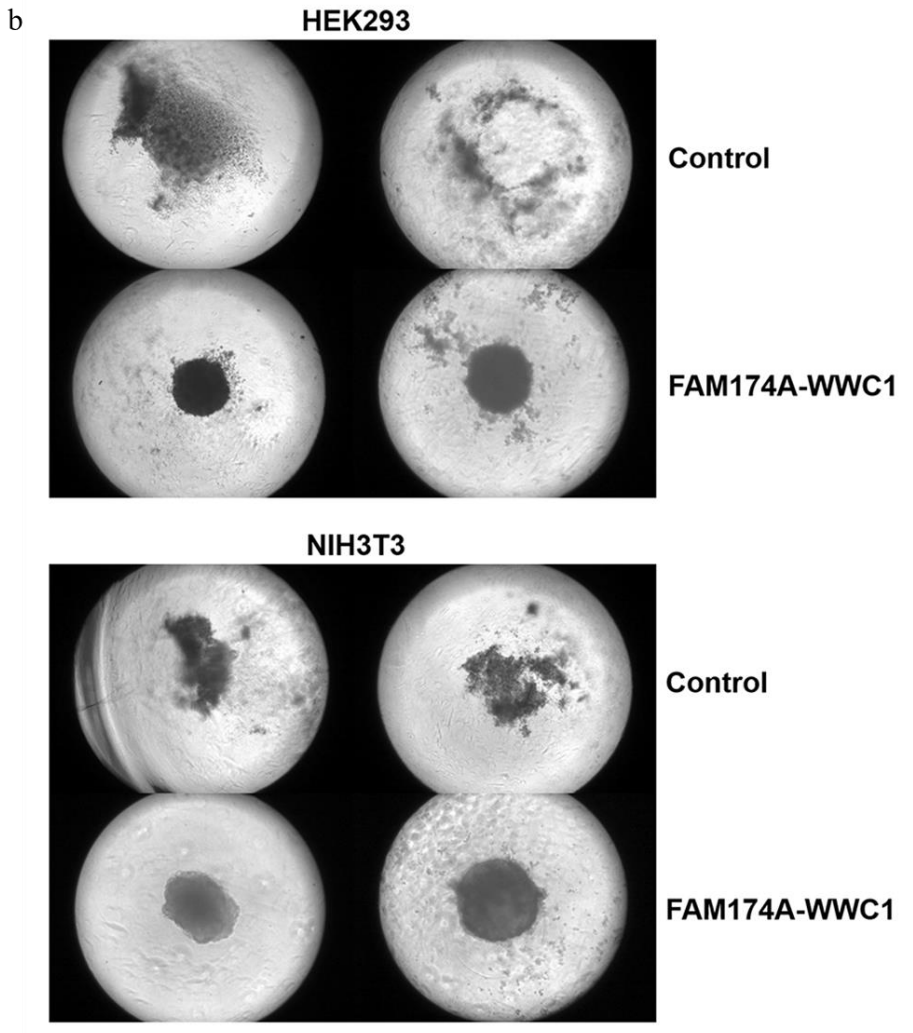


Figure 6. **Oncogenic capacity of *FAM174A-WWC1*. *b.*** 3D spheroid-formation assay showing that *FAM174A-WWC1* expression significantly increased tumor-sphere propagation.

Metastasis-promoting potential of FAM174A-WWC1

To determine the effect of *FAM174A-WWC1* on cell migration and invasion, cells were seeded into a narrow channel with a basement membrane matrix containing reduced growth factors, in a system that mimics the extravasation of tumor cells. Compared with controls, cells expressing *FAM174A-WWC1* showed significantly more invasion (Fig. 7a, 7b). To further determine if the *FAM174A-WWC1* increased all-direction invasiveness, cells were grown as 3D cyst-like forms and seeded with invasion matrix. Indeed, *FAM174A-WWC1* expression promoted the dissemination of cells into the surrounding Matrigel (Fig. 7c, 7d). Moreover, *FAM174A-WWC1* expression also enhanced 2D cell migration (Fig. 8).

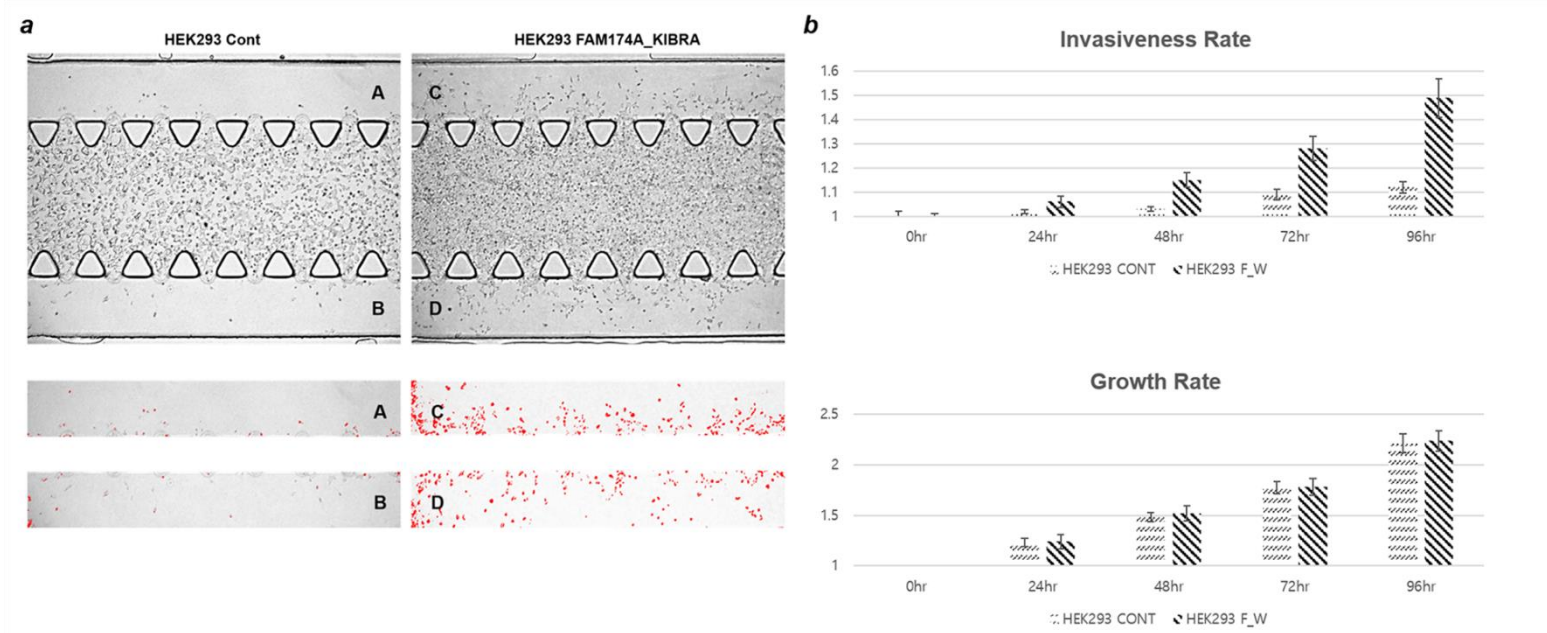


Figure 7. **Metastatic potential of *FAM174A-WWC1*. *a*, *b*.** A 3D chip invasion assay indicated that, compared with controls, cells expressing *FAM174A-WWC1* showed significantly more invaded cells, but little alteration in the cell growth rate. (continue)

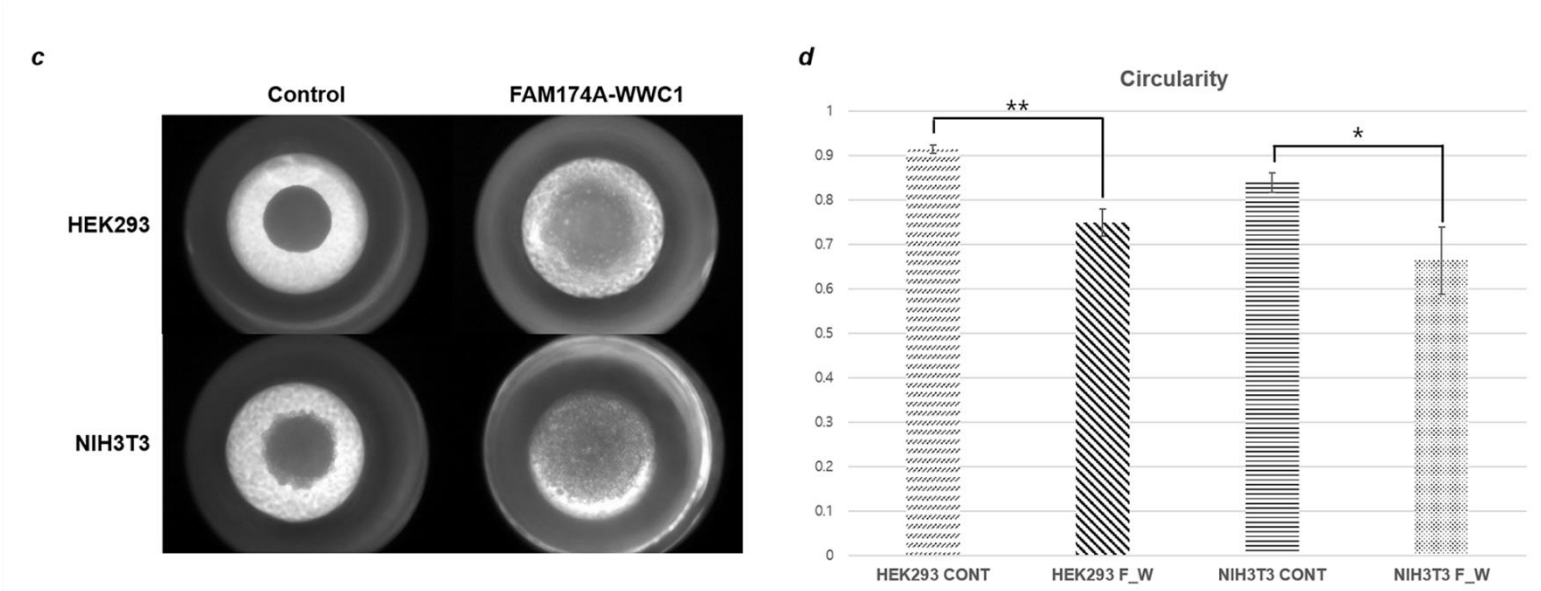


Figure 7. **Metastatic potential of *FAM174A-WWC1*. *c.*** A 3D spheroid cell invasion assay indicated that *FAM174A-WWC1* expression promoted dissemination of cells into the surrounding invasion matrix. ***d.*** The rate of intact round shape (circularity) was measured using ImageJ program. Both HEK293 and NIH3T3 cells transfected with the control vector maintained roundness of the spheroid, whereas cells expressing *FAM174A-WWC1* had disseminated into surrounding invasion matrix.

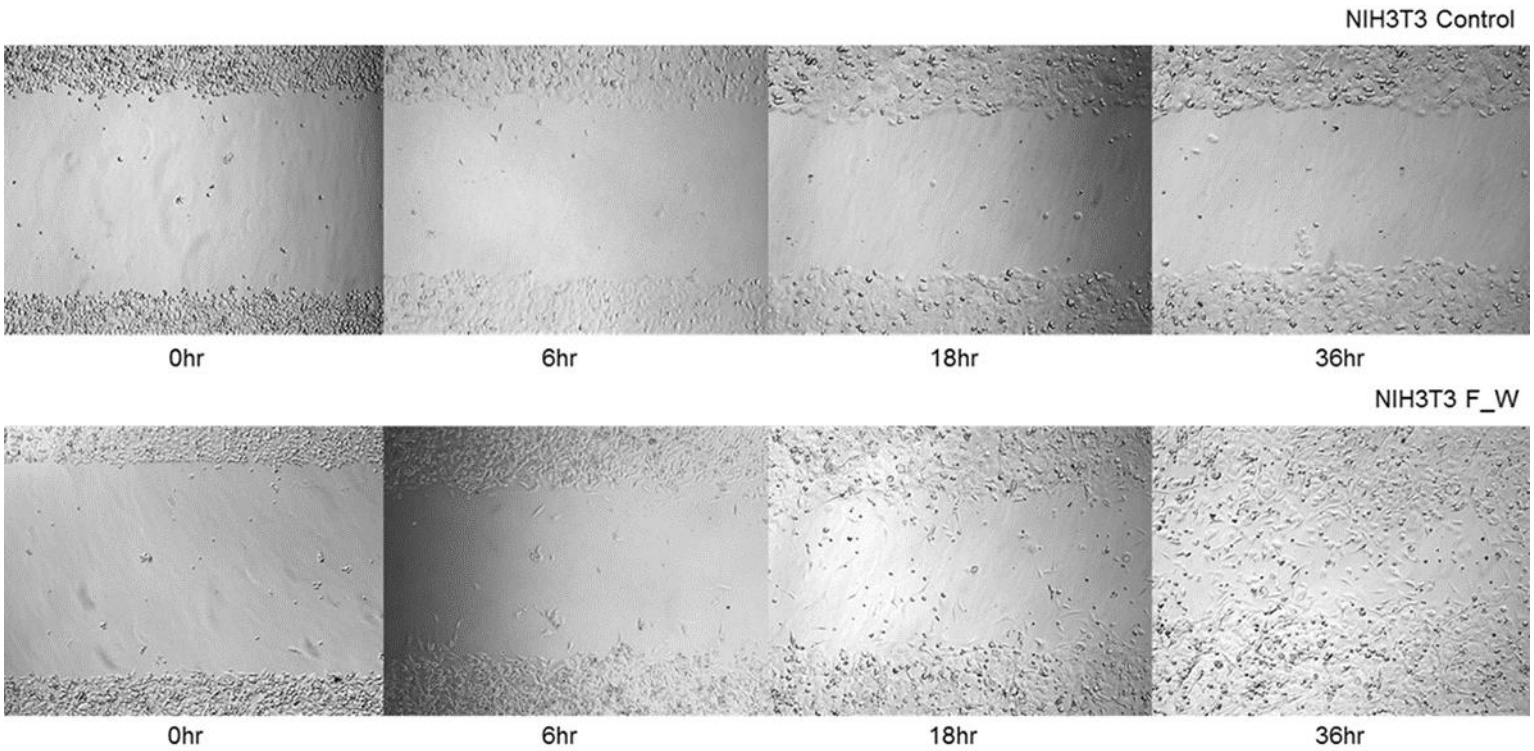


Figure 8. Effect of *FAM174A-WWC1* on wound-healing ability. A wound-healing assay demonstrated that NIH3T3 cells expressing *FAM174A-WWC1* exhibit an augmented wound-closing rate, potentially indicating that the fusion protein activates invasion and migration.

The metastatic potential of FAM174A-WWC1 may involve the focal adhesion and ECM receptor pathways

The intercellular localizations of WWC1 and FAM174A-WWC1 were analyzed with immunocytochemistry (ICC). WWC1 and FAM174A-WWC1 largely co-localized in the cytoplasmic and cytoskeletal regions. However, the fusion protein was more strongly distributed to the cytoskeletal region, which may suggest that it plays a role in focal adhesion and ECM binding (Fig. 9a). Signal intensity analysis revealed that both WWC1 and FAM174A-WWC1 were expressed in the lamellipodia, with prominent FAM174A-WWC1 expression observed at the edges (Fig. 9b, 9c). The subcellular compartmentalization of FAM174A-WWC1 was further investigated with fractional protein extraction. Consistent with the above-described results, the fusion protein was mostly detected in the cytoskeletal fraction (Fig. 9d). Finally, to identify genes and pathways that might be affected by the fusion protein, we performed RNA-sequencing. Indeed, we observed that components of the focal adhesion and ECM receptor pathways were significantly inhibited in cells expressing the FAM174A-WWC1 fusion protein. Collectively, our findings suggest that the fusion protein may contribute to activating cellular movement (Fig. 9e).

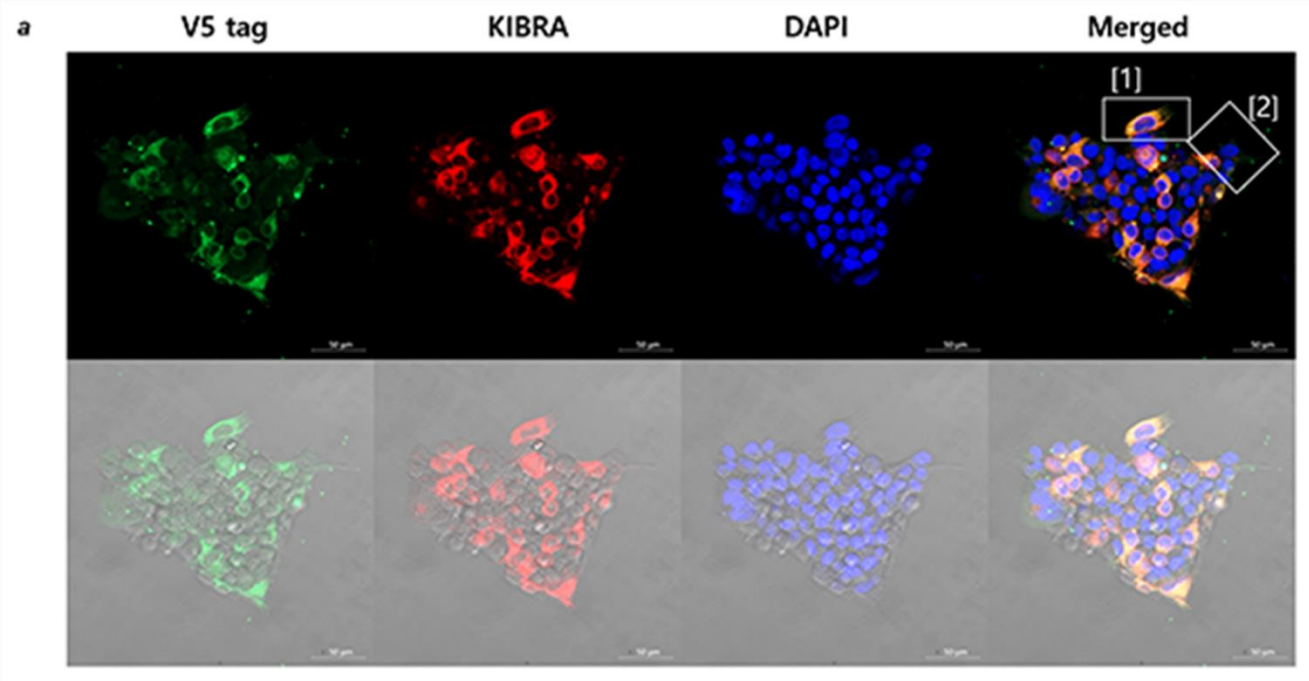


Figure 9. **Metastatic potential of *FAM174A-WWC1* appears to involve alterations in the focal adhesion and ECM receptor pathways. *a*.** WWC1 and FAM174A-WWC1 largely co-localized in the cytoplasmic and cytoskeletal region, but the latter protein was more highly distributed in the cytoskeletal region. (continue)

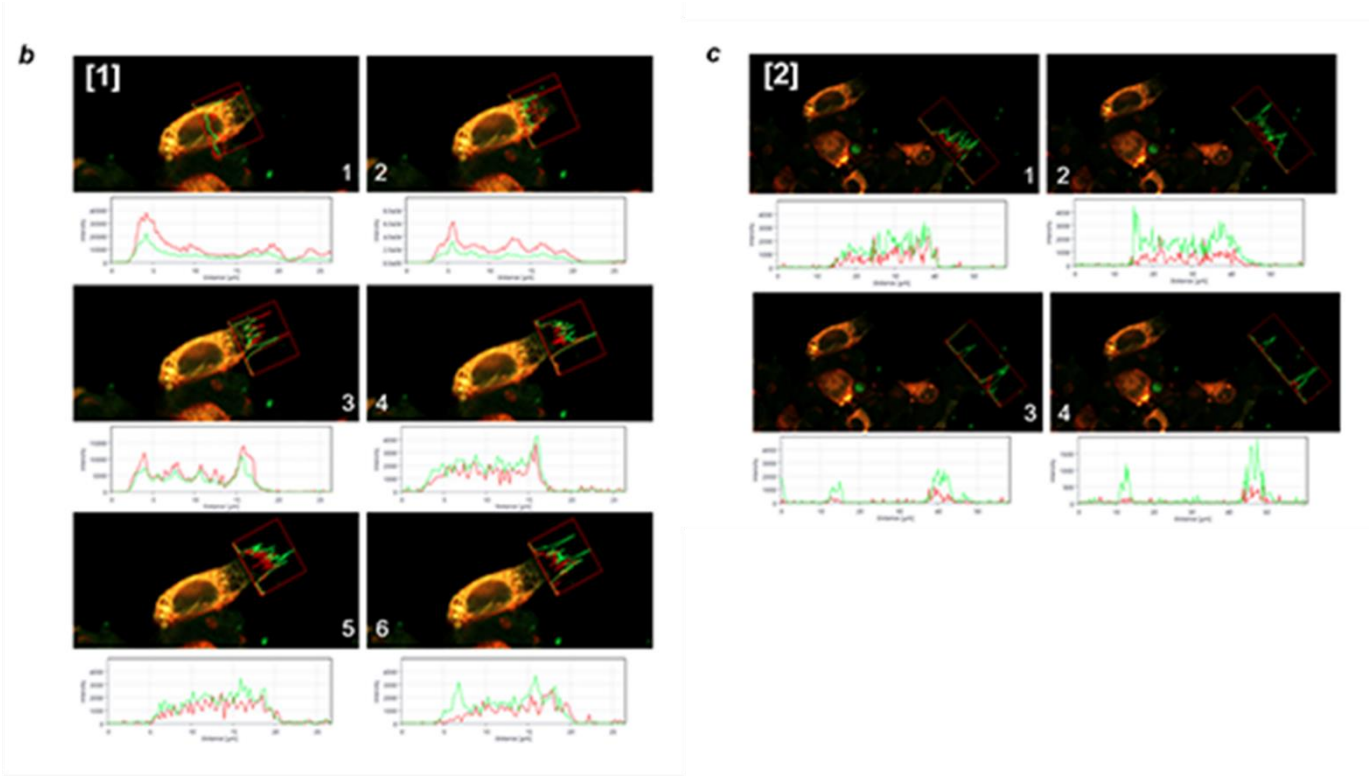


Figure 9. Metastatic potential of FAM174A-WWC1 appears to involve alterations in the focal adhesion and ECM receptor pathways. ***b, c.*** Signal intensity analysis showing that both WWC1 and FAM174A-WWC1 are expressed in lamellipodia, with prominent FAM174A-WWC1 expression seen at the edge. (continue)

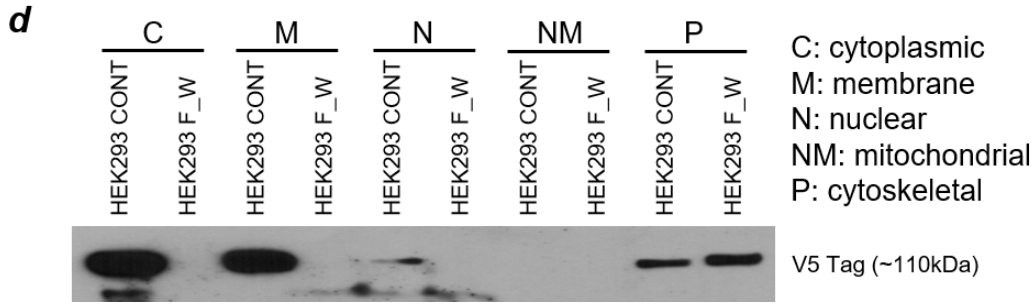
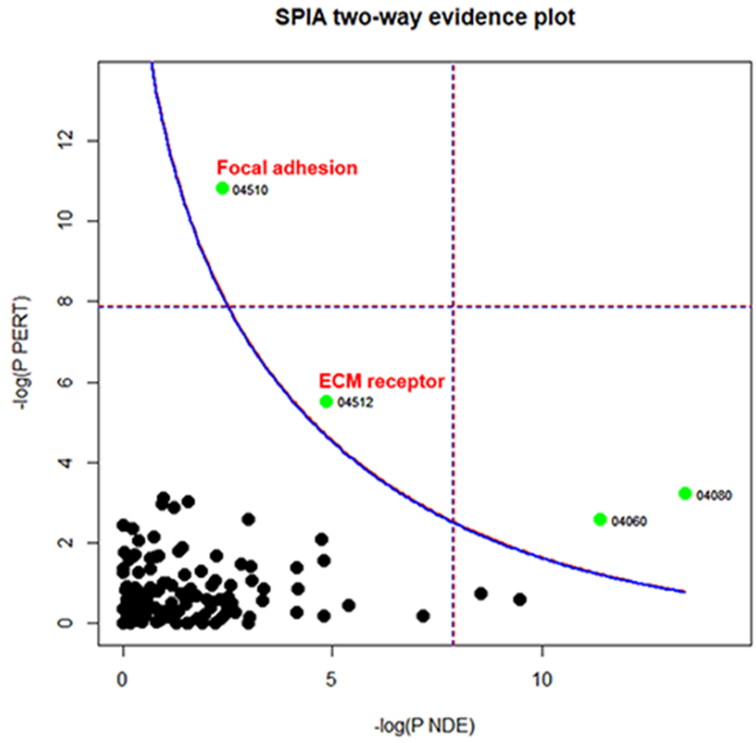


Figure 9. Metastatic potential of *FAM174A-WWC1* appears to involve alterations in the focal adhesion and ECM receptor pathways. ***d***. Fractional protein extraction showed that the fusion protein was mostly detected in the cytoskeletal fraction. (continue)

e



Name	ID	pSize	NDE	pG	pGFdr	pGFWER	Status
Neuroactiv	4080	103	69	2.80E-06	0.000383	0.000383	Inhibited
Cytokine-c	4060	107	69	3.06E-05	0.002097	0.004194	Activated
Focal adhe	4510	152	75	6.30E-05	0.002875	0.008625	Inhibited
ECM-recept	4512	56	34	0.000167	0.005727	0.022906	Inhibited

Figure 9. Metastatic potential of *FAM174A-WWC1* appears to involve alterations in the focal adhesion and ECM receptor pathways. ***e.*** RNA-sequencing results indicate that the focal adhesion and ECM receptor pathways are significantly inhibited in cells expressing FAM174A-WWC1.

Discussion

We herein established four new cell lines and performed NGS analysis on the genetic features of early-onset colorectal cancer with an unclear genetic background.

Through whole-exome sequencing, common mutation in the *PCLO* gene was observed in our cell lines and 4 EOCRC data from TCGA-GDC. The presynaptic cytomatrix protein Piccolo, encoded by *PCLO*, is frequently mutated and amplified in several cancers such as esophageal squamous cell carcinoma, glioblastoma, and chronic myeloid leukemia (44-46). However, its exact roles in tumor is unclear. The mutation is considered as a prognostic biomarker which can accurately predict sensitivity to etoposide in small cell lung cancer (47) or show poor prognosis in mesenchymal subtype of glioblastoma (48). Zhou et al. discovered that a list of dysregulated proteins involved in energy metabolism and carcinogenesis, including *PPARD, IL1-RAP, HNF, S15A2, PCLO, VA0D1, CKLF5*, were extremely upregulated in the CD26 + leukemia stem cell which persist in all patients on long-term therapy and provides a reservoir for disease progression and recurrence (46). Based on the mutational pattern of the *PCLO* gene in EOCRC, we suggest that it may be a potential driver gene for progression of EOCRC. Further studies are warranted to address the possible role of *PCLO* in EOCRC.

Although the cell lines newly generated from these cases did not harbor any of the fusion oncogenes reported previously in EOCRC, one cell line (and the

tumor from which it was derived) was found to express a novel fusion gene involving *FAM174A* and *WWC1*. We performed sequence verification and cloning of the *FAM174A-WWC1* fusion gene, followed by functional characterization of the fusion protein. We used HEK293 cells and NIH3T3 cells to examine the tumorigenic effect of *FAM174A-WWC1* expression. We found that the expression of *FAM174A-WWC1* in HEK293 and NIH3T3 cells was sufficient to trigger cellular transformation and activate cell migration. These characteristics suggest that this novel fusion gene may affect cancer cell invasion and metastasis.

Various fusion genes have been reported in CRC, including *BRAF, NTRK3, RAS,* and *RET* (34, 49); however, no previous study had specifically searched for gene fusions in EOCRC. *WWC1* is known to be involved with several fusions. For instance, the interchromosomal in-frame fusion gene, *WWC1-ADRBK2*, was found in *BRCA*-mutated breast cancer (50). Choi et al. identified an intrachromosomal in-frame fusion gene, *BOD1-WWC1*, in CRC patient tissue (33). Multiple fusions involving the *WWC1* gene may be triggered by the instability of chromosome 5q, deletions of which have been shown to be biologically and prognostically significant in myeloid malignancies (51) and triple negative breast cancers (52). The *FAM174A-WWC1* gene fusion resulted from a 67,892,563 bp deletion in the 5q arm, which may also reflect the loss of the chromosome 5q arm in this EOCRC case. Given the recurrent instability of chromosome 5q, fusion genes involving this chromosome arm may carry various breakpoints. We used the

previously identified break junctions to construct five primer sets that targeted different exons in *FAM174A* and *WWC1*, but did not find any evidence of any breakpoint other than those involving exon 2 of *FAM174A* and exon 10 of *WWC1* (data not shown).

Analysis of the *FAM174A-WWC1* sequence revealed that the *WWC1* portion of *FAM174A-WWC1* maintained the entire C2 domain and PDZ binding motif of *WWC1*, but lacks the WW domain (Fig. 5a). WWC1 has been suggested to be a key regulator of the hippo pathway, which is responsible for regulating organ size, cell contact inhibition, tissue regeneration, and tumorigenesis. WWC1 activates large tumor suppressor (LATS1/2) kinases and Yes-associated protein 1 (YAP 1) via phosphorylation on the hydrophobic motif (53, 54). Our results indicated that FAM174A-WWC1 lowered the membrane level of YAP1 while significantly increasing its level in the nucleus (Fig. 5f), suggesting that replacing the WW domain with the DUF1180 domain can facilitate the nuclear localization of YAP1. Phosphorylation at Ser127 by LATS kinases promotes the translocation of YAP1 from the nucleus to the cytoplasm, where it is sequestered through association with 14-3-3 proteins (55). The nuclear level of p-S127 YAP1 was increased in HEK293 cells expressing *FAM174A-WWC1* (Fig. 5f), potentially suggesting that the fusion protein hinders the cytoplasmic translocalization of phosphorylated YAP1. YAP1 exerts significant effects in various malignancies. It has been found to promote proliferation, invasion, and migration in colon cancer cells both *in vitro* and *in vivo*. Bioinformatics prediction, dual luciferase assays, RNA-IP,

and RNA pull-down assays demonstrated that YAP1-induced MALAT1 promotes the expression of metastasis-associated molecules, such as VEGFA, SLUG, and TWIST, by sponging miR-126-5p in CRC. These findings indicate that the YAP1–MALAT1–miR-126-5p axis could control angiogenesis and the epithelial–mesenchymal transition in CRC (56).

Our SNV profiling of EOCRC patients suggested that the fusion gene could have oncogenic potential. The tumor expressing *FAM174A-WWC1* carried the fewest oncogene mutations (Fig. 6a), which was consistent with a previous report suggesting that tumors harboring fusion genes tended to exhibit fewer oncogene mutations (34). However, it remains possible that the development of cancer in patients with *FAM174A-WWC1* may be influenced by oncogene and tumor suppressor gene mutations involving APC, in addition to the fusion gene. To further determine whether *FAM174A-WWC1* is related to the tumorigenic capacity, cells were grown as tumor-spheres; this enabled us to examine the tumor-initiating capability and stem-like properties of the various cells, since sphere-forming efficiency (57) indicates tumorigenic potential (58). Indeed, we found that *FAM174A-WWC1* expression significantly increased tumor-sphere propagation (Fig. 6b). It was previously reported that the WW1/2 domains of WWC1 mediated the ability of this protein to inhibit tumor-sphere formation (52). *FAM174A-WWC1* lacks the WW1/2 domains of *WWC1* but contains the DUF1180 domain of *FAM174A*. The *FAM174A* gene encodes a membrane protein that may contribute to cholesterol hemostasis (59, 60). Although truncation of *FAM174A* has been

observed in families with Parkinson's disease, few studies have focused on this gene (61). Thus, considering this function of *FAM174A*, our results may suggest that the gain of DUF1180 domain and loss of WW1/2 domain has the capacity to accelerate tumor progression.

The patient harboring *FAM174A-WWC1* developed multiple metastases to their lungs and liver after surgery and showed a poor prognosis. (Table 2). Consistent with this, our 2D and 3D invasion assays showed that, without exception, cells expressing *FAM174A-WWC1* exhibited increased invasiveness and dissemination compared with controls (Fig. 7 and Fig. 8). Our data suggest that *FAM174A-WWC1* can function to promote metastatic dissemination, which is contrary to the role of naïve *WWC* as a metastatic suppressor (52). Thus, the replacement of the WW domains with the DUF1180 domain may reverse the primary role of *WWC1*. Collectively, our findings indicate that *FAM174A-WWC1* expression augments the tumorigenic and metastatic capacity of cells, thereby conferring significant advantages to EOCRC.

The discovery of driver mutations among numerous passenger mutations is vital to pinpoint the cause of certain diseases. The subset of mutations that are found in only tumor from normal tissue provides the distinction of germline mutation. However, due to the unusually aggressive clinical course of EOCRC, we were unable to obtain a normal tissue sample representing patients NCC-375. To examine whether the *FAM174A-WWC1* fusion gene was a somatic mutation, we accessed this mutation in 500 CRC patient tissues

using primers that span the break junction of the fusion gene. None of the tested CRC tissues had the same fusion gene, prompting us to provisionally conclude that the fusion gene had occurred randomly.

Conclusions

Although there were fewer mutation related to carcinogenesis of CRC, *PCLO* gene mutation was observed in all EOCRC cell lines and EOCRC data from TCGA-GDC. Also, we herein used comprehensive and integrated clinicopathological analysis, and *in vitro* assessments to identify and analyze a novel fusion gene, *FAM174A-WWC1*, from a patient with EOCRC. We demonstrate that *FAM174A-WWC1* increases the oncogenic and metastatic capacity of cell *in vitro*. This fusion gene may represent a robust molecular and therapeutic target in EOCRC. Although cases of EOCRC carry an overall unfavorable prognosis, identification of young patients diagnosed with EOCRC may represent a population of high-risk cases with tumors driven by novel genes such as *FAM174A-WWC1*. It would be interesting to determine the prevalence of *FAM174A-WWC1* in archival samples of EOCRC and to correlate the *FAM174A-WWC1* status with clinical outcome.

References

1. Siegel RL, Jemal A, Ward EM. Increase in incidence of colorectal cancer among young men and women in the United States. Cancer Epidemiol Biomarkers Prev. 2009;18(6):1695-8.

2. Hubbard JM, Grothey A. Adolescent and young adult colorectal cancer. J Natl Compr Canc Netw. 2013;11(10):1219-25.

3. Stigliano V, Sanchez-Mete L, Martayan A, Anti M. Early-onset colorectal cancer: a sporadic or inherited disease? World J Gastroenterol. 2014;20(35):12420-30.

4. Ahnen DJ, Wade SW, Jones WF, Sifri R, Mendoza Silveiras J, Greenamyer J, et al. The increasing incidence of young-onset colorectal cancer: a call to action. Mayo Clin Proc. 2014;89(2):216-24.

5. Tsang WY, Ziogas A, Lin BS, Seery TE, Karnes W, Stamos MJ, et al. Role of primary tumor resection among chemotherapy-treated patients with synchronous stage IV colorectal cancer: a survival analysis. J Gastrointest Surg. 2014;18(3):592-8.

6. Vuik FE, Nieuwenburg SA, Bardou M, Lansdorp-Vogelaar I, Dinis-Ribeiro M, Bento MJ, et al. Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years. Gut. 2019;68(10):1820-6.

7. Kim SC, Shin R, Seo HY, Kim M, Park JW, Jeong SY, et al. Identification of a Novel Fusion Gene, FAM174A-WWC1, in Early-Onset Colorectal Cancer: Establishment and Characterization of Four Human Cancer Cell Lines from Early-Onset Colorectal Cancers. Translational oncology. 2019;12(9):1185-95.

8. Cavestro GM, Mannucci A, Zuppardo RA, Di Leo M, Stoffel E, Tonon G. Early onset sporadic colorectal cancer: Worrisome trends and oncogenic features. Dig Liver Dis. 2018;50(6):521-32.

9. Shida D, Ahiko Y, Tanabe T, Yoshida T, Tsukamoto S, Ochiai H, et al. Shorter survival in adolescent and young adult patients, compared to adult patients, with stage IV colorectal cancer in Japan. BMC Cancer. 2018;18(1):334.

10. Ballester V, Rashtak S, Boardman L. Clinical and molecular features of young-onset colorectal cancer. World J Gastroenterol. 2016;22(5):1736-44.

11. Dozois EJ, Boardman LA, Suwanthanma W, Limburg PJ, Cima RR, Bakken JL, et al. Young-onset colorectal cancer in patients with no known genetic predisposition: can we increase early recognition and improve outcome? Medicine (Baltimore). 2008;87(5):259-63.

12. You YN, Xing Y, Feig BW, Chang GJ, Cormier JN. Young-onset colorectal cancer: is it time to pay attention? Arch Intern Med. 2012;172(3):287-9.

13. Stoffel EM, Koeppe E, Everett J, Ulintz P, Kiel M, Osborne J, et al. Germline Genetic Features of Young Individuals With Colorectal Cancer. Gastroenterology. 2018;154(4):897-905.e1.

14. Liu B, Farrington SM, Petersen GM, Hamilton SR, Parsons R, Papadopoulos N, et al. Genetic instability occurs in the majority of young patients with colorectal cancer. Nat Med. 1995;1(4):348-52.

15. Gryfe R, Kim H, Hsieh ET, Aronson MD, Holowaty EJ, Bull SB, et al. Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. N Engl J Med. 2000;342(2):69-77.

16. Losi L, Di Gregorio C, Pedroni M, Ponti G, Roncucci L, Scarselli A, et al. Molecular genetic alterations and clinical features in early-onset colorectal

carcinomas and their role for the recognition of hereditary cancer syndromes. Am J Gastroenterol. 2005;100(10):2280-7.

17. Goel A, Nagasaka T, Spiegel J, Meyer R, Lichliter WE, Boland CR. Low frequency of Lynch syndrome among young patients with non-familial colorectal cancer. Clin Gastroenterol Hepatol. 2010;8(11):966-71.

18. Giraldez MD, Balaguer F, Bujanda L, Cuatrecasas M, Munoz J, Alonso-Espinaco V, et al. MSH6 and MUTYH deficiency is a frequent event in early-onset colorectal cancer. Clin Cancer Res. 2010;16(22):5402-13.

19. Dwyer AJ, Murphy CC, Boland CR, Garcia R, Hampel H, Limburg P, et al. A Summary of the Fight Colorectal Cancer Working Meeting: Exploring Risk Factors and Etiology of Sporadic Early-Age Onset Colorectal Cancer. Gastroenterology. 2019;157(2):280-8.

20. Bardhan K, Liu K. Epigenetics and colorectal cancer pathogenesis. Cancers (Basel). 2013;5(2):676-713.

21. Chan TL, Curtis LC, Leung SY, Farrington SM, Ho JW, Chan AS, et al. Early-onset colorectal cancer with stable microsatellite DNA and near-diploid chromosomes. Oncogene. 2001;20(35):4871-6.

22. Antelo M, Balaguer F, Shia J, Shen Y, Hur K, Moreira L, et al. A high degree of LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer. PLoS One. 2012;7(9):e45357.

23. Kirzin S, Marisa L, Guimbaud R, De Reynies A, Legrain M, Laurent-Puig P, et al. Sporadic early-onset colorectal cancer is a specific sub-type of cancer: a morphological, molecular and genetics study. PLoS One. 2014;9(8):e103159.

24. Harewood L, Fraser P. The impact of chromosomal rearrangements on regulation of gene expression. Hum Mol Genet. 2014;23(R1):R76-82.

25. Mertens F, Johansson B, Fioretos T, Mitelman F. The emerging complexity of gene fusions in cancer. Nature reviews Cancer. 2015;15(6):371-81.

26. Dai X, Theobard R, Cheng H, Xing M, Zhang J. Fusion genes: A promising tool combating against cancer. Biochimica et biophysica acta Reviews on cancer. 2018;1869(2):149-60.

27. Pietrantonio F, Di Nicolantonio F, Schrock AB, Lee J, Tejpar S, Sartore-Bianchi A, et al. ALK, ROS1, and NTRK Rearrangements in Metastatic Colorectal Cancer. J Natl Cancer Inst. 2017;109(12).

28. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood. 2016;127(20):2391-405.

29. Rejlova K, Musilova A, Kramarzova KS, Zaliova M, Fiser K, Alberich-Jorda M, et al. Low HOX gene expression in PML-RARalpha-positive leukemia results from suppressed histone demethylation. Epigenetics. 2018;13(1):73-84.

30. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448(7153):561-6.

31. Rolfo C, Raez L. New targets bring hope in squamous cell lung cancer: neurotrophic tyrosine kinase gene fusions. Laboratory investigation; a journal of technical methods and pathology. 2017;97(11):1268-70.

32. Cocco E, Scaltriti M, Drilon A. NTRK fusion-positive cancers and TRK inhibitor therapy. Nat Rev Clin Oncol. 2018;15(12):731-47.

33. Choi Y, Kwon CH, Lee SJ, Park J, Shin JY, Park DY. Integrative analysis of oncogenic fusion genes and their functional impact in colorectal cancer. Br J Cancer. 2018;119(2):230-40.

34. Kloosterman WP, Coebergh van den Braak RRJ, Pieterse M, van Roosmalen MJ, Sieuwerts AM, Stangl C, et al. A Systematic Analysis of Oncogenic Gene Fusions in Primary Colon Cancer. Cancer Res. 2017;77(14):3814-22.

35. 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(30):28929-37.

36. Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy CB, et al. Recurrent R-spondin fusions in colon cancer. Nature. 2012;488(7413):660-4.

37. Ku JL, Shin YK, Kim DW, Kim KH, Choi JS, Hong SH, et al. Establishment and characterization of 13 human colorectal carcinoma cell lines: mutations of genes and expressions of drug-sensitivity genes and cancer stem cell markers. Carcinogenesis. 2010;31(6):1003-9.

38. Kim SC, Hong CW, Jang SG, Kim YA, Yoo BC, Shin YK, et al. Establishment and Characterization of Paired Primary and Peritoneal Seeding Human Colorectal Cancer Cell Lines: Identification of Genes That Mediate Metastatic Potential. Translational oncology. 2018;11(5):1232-43.

39. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic acids research. 2003;31(13):3812-4.

40. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics. 2013;Chapter 7:Unit7.20.

41. Shamsani J, Kazakoff SH, Armean IM, McLaren W, Parsons MT, Thompson BA, et al. A plugin for the Ensembl Variant Effect Predictor that uses

MaxEntScan to predict variant spliceogenicity. Bioinformatics (Oxford, England). 2018.

42. Mbarek H, Milaneschi Y, Hottenga JJ, Ligthart L, de Geus EJC, Ehli EA, et al. Genome-Wide Significance for PCLO as a Gene for Major Depressive Disorder. Twin research and human genetics : the official journal of the International Society for Twin Studies. 2017;20(4):267-70.

43. Ahmed MY, Chioza BA, Rajab A, Schmitz-Abe K, Al-Khayat A, Al-Turki S, et al. Loss of PCLO function underlies pontocerebellar hypoplasia type III. Neurology. 2015;84(17):1745-50.

44. Pansare K, Gardi N, Kamat S, Dange P, Previn R, Gera P, et al. Establishment and genomic characterization of gingivobuccal carcinoma cell lines with smokeless tobacco associated genetic alterations and oncogenic PIK3CA mutation. Sci Rep. 2019;9(1):8272.

45. Zhang W, Hong R, Xue L, Ou Y, Liu X, Zhao Z, et al. Piccolo mediates EGFR signaling and acts as a prognostic biomarker in esophageal squamous cell carcinoma. Oncogene. 2017;36(27):3890-902.

46. Zhou S, Zhu X, Liu W, Cheng F, Zou P, You Y, et al. Comparison of chronic myeloid leukemia stem cells and hematopoietic stem cells by global proteomic analysis. Biochem Biophys Res Commun. 2019.

47. Qiu Z, Lin A, Li K, Lin W, Wang Q, Wei T, et al. A novel mutation panel for predicting etoposide resistance in small-cell lung cancer. Drug Des Devel Ther. 2019;13:2021-41.

48. Park AK, Kim P, Ballester LY, Esquenazi Y, Zhao Z. Subtype-specific signaling pathways and genomic aberrations associated with prognosis of glioblastoma. Neuro Oncol. 2019;21(1):59-70.

49. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330-7.

50. Ha KC, Lalonde E, Li L, Cavallone L, Natrajan R, Lambros MB, et al. Identification of gene fusion transcripts by transcriptome sequencing in BRCA1-mutated breast cancers and cell lines. BMC Med Genomics. 2011;4:75.

51. Giagounidis AA, Germing U, Aul C. Biological and prognostic significance of chromosome 5q deletions in myeloid malignancies. Clin Cancer Res. 2006;12(1):5-10.

52. Knight JF, Sung VYC, Kuzmin E, Couzens AL, de Verteuil DA, Ratcliffe CDH, et al. KIBRA (WWC1) Is a Metastasis Suppressor Gene Affected by Chromosome 5q Loss in Triple-Negative Breast Cancer. Cell reports. 2018;22(12):3191-205.

53. Yu FX, Zhao B, Guan KL. Hippo Pathway in Organ Size Control, Tissue Homeostasis, and Cancer. Cell. 2015;163(4):811-28.

54. Park HW, Guan KL. Regulation of the Hippo pathway and implications for anticancer drug development. Trends in pharmacological sciences. 2013;34(10):581-9.

55. Zhao B, Wei X, Li W, Udan RS, Yang Q, Kim J, et al. Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. Genes & development. 2007;21(21):2747-61.

56. Sun Z, Ou C, Liu J, Chen C, Zhou Q, Yang S, et al. YAP1-induced MALAT1 promotes epithelial-mesenchymal transition and angiogenesis by sponging miR-126-5p in colorectal cancer. Oncogene. 2019;38(14):2627-44.

57. Sanders H, Qu K, Li H, Ma L, Barlan C, Zhang X, et al. Mutation Yield of a 34-Gene Solid Tumor Panel in Community-Based Tumor Samples. Mol Diagn Ther. 2016;20(3):241-53.

58. Weiswald LB, Bellet D, Dangles-Marie V. Spherical cancer models in tumor biology. Neoplasia. 2015;17(1):1-15.

59. Blattmann P, Schuberth C, Pepperkok R, Runz H. RNAi-based functional profiling of loci from blood lipid genome-wide association studies identifies genes with cholesterol-regulatory function. PLoS Genet. 2013;9(2):e1003338.

60. Lewis SL, Holl HM, Streeter C, Posbergh C, Schanbacher BJ, Place NJ, et al. Genomewide association study reveals a risk locus for equine metabolic syndrome in the Arabian horse. Journal of animal science. 2017;95(3):1071-9.

61. Yemni EA, Monies D, Alkhairallah T, Bohlega S, Abouelhoda M, Magrashi A, et al. Integrated Analysis of Whole Exome Sequencing and Copy Number Evaluation in Parkinson's Disease. Sci Rep. 2019;9(1):3344.

요약(국문초록)

서론: 50세 이후 대장암의 발생률과 사망률이 감소하고 있음에도 불구하고, 50세 미만의 조기 발병 대장암은 증가 추세이다. 50세 이전에 발생하는 조기발병대장암은 유전성이라고 생각할 수 있지만, 조기발병대장암의 약 20%정도만 가족성 대장암 증후군(familial colorectal cancer syndrome)이고 나머지 80%정도는 원인 유전자를 모른다. 이에 저자들은 재현 가능한 생물학적인 자원을 확립하고, 조기발병대장암의 유전자 변이에 관한 유전정보 구축에 기여하기 위해 본 연구를 시행하였다.

방법: 30세 이전에 대장암을 진단받은 네 명의 조기발병대장암 환자들의 수술 후 적출된 대장암 조직으로부터 네 종류의 세포주를 수립했다. 이 세포주들에 대한 차세대염기서열 (Next-generation sequencing)을 통해, 조기발병대장암의 유전적인 특징을 분석하였다.

결과: 전체 엑솜 염기서열 분석(whole-exome sequencing)을 통해, 대장암 발생기전에 주로 작용하는 유전자의 변이를 분석해 보았을 때, SNU-1460 세포주에서 *TP53* gene의 slice 돌연변이가 발견된 것을 제외하고는 의미 있는 유전자의 변이가 발견되지 않았다. 또한, 엑솜 염기서열 분석 자료를 The Cancer Genome Atlas-Genomic Data Commons (TCGA-GDC)에서 추출한 35세 미만의 조기발병대장암 환자에서의 유전자 변이와 비교해서, *PCLO* 유전자에 공통적으로 변이가 있는 것을 발견하였다. RNA 염기서열 분석(RNA sequencing)을 통해서는

*FAM174A-WWC1*이라는 새로운 융합 유전자(fusion gene)을 발견했고, 그 기능을 알아보았다. *FAM174-WWC1*의 발현으로 인해, 정상 섬유아세포의 형태의 변형이 있었고, 이 형질 도입이 된 섬유아세포들이 종양 구체(tumor-sphere)모양으로 자라는 것을 확인하였다. 또한, 세포내 E-cadherin의 감소와 N-cadherin의 증가가 관찰되었다. 이 융합유전자는 세포막과 세포질에서의 YAP1단백질의 발현을 감소시키고, 세포핵에서의 YAP1 단백질의 발현을 증가시켰다. FAM174-WWC1 단백질은 주로 세포질에서 발견되며, 세포골격에서 정상 WWC1 단백질보다 많이 존재함을 발견했다.

결론: 이 연구를 통해, 조기발병대장암에 공통적으로 PCLO 유전자의 변이가 있으며, 새로운 융합유전자인 *FAM174A-WWC1*의 발현은 세포의 발암성(oncogenic capacity) 및 침습성(invasiveness)이 증가시키는 것을 밝혔다.

주요어: 조기발병 대장암, 차세대염기서열, 융합유전자, *WWC1*

학번: 2013-30549