

RESEARCH ARTICLE



# Genomic and transcriptomic analysis of Korean colorectal cancer patients

Sol A Jeon<sup>1,2,5</sup> · Ye Jin Ha<sup>3</sup> · Jong-Hwan Kim<sup>1,2</sup> · Jeong-Hwan Kim<sup>1</sup> · Seon-Kyu Kim<sup>1,5</sup> · Yong Sung Kim<sup>1,2,5</sup> · Seon-Young Kim<sup>1,2,5</sup> · Jin Cheon Kim<sup>3,4</sup>

Received: 4 March 2022 / Accepted: 1 June 2022 / Published online: 25 June 2022  
© The Author(s) 2022

## Abstract

**Background** Colorectal cancer (CRC) is the third most common type of diagnosed cancer in the world and has the second-highest mortality rate. Meanwhile, South Korea has the second-highest incidence rate for CRC in the world.

**Objective** To assess the possible influence of ethnicity on the molecular profile of colorectal cancer, we compared genomic and transcriptomic features of South Korean CRCs with European CRCs.

**Methods** We assembled a genomic and transcriptomic dataset of South Korean CRC patients (KOCRC; n = 126) from previous studies and European cases (EUCRC; n = 245) selected from The Cancer Genome Atlas (TCGA). Then, we compared the two datasets in terms of clinical data, driver genes, mutational signature, gene sets, consensus molecular subtype, and fusion genes.

**Results** These two cohorts showed similar profiles in driver mutations but differences in the mutation frequencies of some driver genes (including *APC*, *TP53*, *PABPC1*, *FAT4*, *MUC7*, *HSPG2*, *GNAS*, *DENND5B*, and *BRAF*). Analysis of hallmark pathways using genomic data sets revealed further differences between these populations in the WNT, TP53, and NOTCH signaling pathways. In consensus molecular subtype (CMS) analyses of the study cases, no *BRAF* mutations were found in the CMS1 subtype of KOCRC, which contrasts with previous findings. Fusion gene analysis identified oncogenic fusion of *PTPRK-RSPO3* in a subset of KOCRC patients without *APC* mutations.

**Conclusions** This study presents insights into the genomic landscape of KOCRCs and reveals some similarities and differences with EUCRCs at the molecular level.

**Keywords** Colorectal cancer · South Korea · European · Genomic landscape · Ethnicity

---

These two authors contributed equally to this work: Sol A. Jeon and Ye Jin Ha.

✉ Seon-Young Kim  
kimsy@kribb.re.kr

✉ Jin Cheon Kim  
jckim@amc.seoul.kr

<sup>1</sup> Personalized Genomic Medicine Research Center, Daejeon, South Korea

<sup>2</sup> Korea Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, South Korea

<sup>3</sup> Asan Institute for Life Sciences, Asan Medical Center, Seoul 05505, South Korea

<sup>4</sup> Department of Surgery, Asan Medical Center, University of Ulsan College of Medicine, Seoul 05505, South Korea

<sup>5</sup> Department of Functional Genomics, University of Science and Technology (UST), Daejeon, South Korea

## Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed cancer in the world and has the second-highest mortality rate, accounting for about 1 out of 10 cancer mortalities worldwide. Moreover, the global burden of CRC is expected to increase by 60% to more than 2.2 million new cases and 1.1 million deaths by 2030 (Arnold et al. 2017). Notably, in this regard, South Korea has a CRC rate of 44.5 (age-standardized rate per 100,000), which was the second-highest global rate in 2018 (Bray et al. 2018).

Over the past three decades, molecular genetic studies have provided important genomic insights into the pathogenesis of both sporadic and hereditary CRC (Fearon 2011). Alterations in oncogenes and tumor suppressor genes are closely related to CRC subsets, and a larger collection of pathway genes has also been defined for these tumors

(Fearon 2011). Various targets have been subsequently explored concerning personalized treatments, and these targeted therapies are regarded as a novel approach to improving individual survival outcomes in CRC patients (Xie et al. 2020).

According to prior large-scale genomic investigations (Cancer Genome Atlas Network 2012; Lu et al. 2019; Nagahashi et al. 2016), well-known driver gene mutations including *APC*, *TP53*, *SMAD4*, *PIK3CA*, and *KRAS*, are significantly involved in the tumorigenesis of CRC. Furthermore, the cancer genome atlas (TCGA) has revealed the role of several new driver genes and potential target pathways in these cancers (Cancer Genome Atlas Network 2012). However, the genomic knowledge of CRC has mainly been acquired from European cohorts, and little information is available on the genomic landscape in Asian CRC populations, including Korean CRC cohorts (KOCRC). Multiple genomic studies have revealed new therapeutic approaches to CRC (Ellis and Perou 2013; Horibata et al. 2020; Nagahashi et al. 2016), uncovering the specific genomic and molecular profiles of KOCRC cohorts will likely assist with the tailoring of diagnostic and therapeutic modalities for Korean cases.

The present study aimed to identify specific molecular and genetic features of KOCRCs using an integrated approach that combined clinical data comparisons with a well-defined European CRC population (EUCRC).

## Materials and methods

### Dataset establishment and public data processing

Genomic and transcriptomic data sets of KOCRC patients ( $n = 126$ ) were obtained from three previous studies (Kim et al. 2016, 2019, 2014) by the Korea Research Institute of Bioscience and Biotechnology (KRIBB, Daejeon, Republic of Korea) and Asan Medical Center (Seoul, Republic of Korea). Whole exome sequencing (WES) of normal samples was carried out using normal tissues or blood samples ( $n = 42$  and  $n = 84$ , respectively). All patients provided voluntary written formal consent to be included in the study. The study protocol strictly conformed to the Declaration of Helsinki and was approved by the Institutional Review Board of Asan Medical Center (registration numbers: 2009–0091, 2014–0150, 2018–0087). The data sets used in this study are available from GEO (GSE50760, GSE107422, GSE132024) and KoNA (PRJKA210050).

To examine possible ethnic differences in the molecular profiles of CRC between our Korean cases and a European

cohort, we downloaded a CRC dataset from The Cancer Genome Atlas (TCGA), and exclusively selected Caucasian cases for our present analyses (EUCRC;  $n = 245$ ) as the European ancestry cohort. The information for our EUCRC cases, including MAF, gene expression count, and clinical data, were acquired from the TCGA colon adenocarcinoma (TCGA-COAD) and TCGA rectum adenocarcinoma (TCGA-READ) project through the GDC Data Portal (Cancer Genome Atlas Network 2012). We used MAF files as an alternative to bam files for WES data and gene expression count files as an alternative to raw RNA sequencing (RNA-seq) fastq files. For further information about sample collection, histology method, library preparation, and bioinformatics analysis of both cohorts, please see Supplementary Table 1.

### Identification of somatic SNVs, indels, and gene fusion events

In the KOCRC cohort, exome sequencing reads were mapped to the human reference genome GRCh38 (primary assembly) using *bwa-mem* (version 0.7.17-r1188) with default parameters, followed by sorting of the bam files with *samtools* (version 1.10). As the TCGA databases had been preprocessed using *GATK* (McKenna et al. 2010), our databases were processed following *GATK* best practices (*GATK* version 4.1.4.0). PCR duplicates were removed via *Picard MarkDuplicates* (version 2.21.2), and base recalibration was conducted using *GATK BaseRecalibrator & ApplyRecalibration*. Candidate variants were called via *GATK Mutect2* and filtered using *GATK FilterMutectCalls*. *ANNOVAR* (Wang et al. 2010) was used for the annotation steps.

Fusion genes and positions were predicted using *STAR-Fusion* (version 1.9.1). We used trimmed KOCRC RNA-seq fastq files as the input. We filtered and determined fusion genes identified in 4-time repeats in a sample. Fusion genes, including non-coding RNA or immunoglobulin-related genes, were excluded from the final selection. The reported and non-reported fusion genes were distinguished using previous reports.

### Driver gene and mutational signature identification

*MutSigCV* (Lawrence et al. 2013) (version 1.3.5) software was used to detect driver genes in our CRC subjects. Briefly, the KOCRC cases were lifted from GRCh38 to GRCh37 via the *CrossMap* (version 0.3.8) for *MutSigCV* processing. The *maftools* (Mayakonda et al. 2018) R package (version 2.6.0) was consecutively used to prepare MAF files for

the MutSigCV analysis, which was finally completed on the GenePattern (Reich et al. 2006) online platform using default settings.

The nonnegative matrix factorization (NMF) R package (version 0.23.0) and maftools R package (version 2.6.0) were used to identify de novo mutation signatures. The number of signatures was estimated based on a cophenetic correlation matrix. Mutational signatures were then extracted from the trinucleotide context and decomposed into the designated number of signatures.

### Gene set enrichment analysis (GSEA) and consensus molecular subtyping

Transcriptomic data from the KOCRC and EUCRC cases were used to conduct GSEA. Trimmed RNA-seq fastq files were mapped to GRCh38 (primary assembly) on STAR (Dobin et al. 2013) (version 2.7.3a), concurrently estimating the expression counts. The edgeR (Robinson et al. 2010) R package (version 3.32.0) was used to obtain log<sub>2</sub> fold-changes in gene expression between normal and tumor tissues. The fgsea (Korotkevich et al. 2021) R package was used to perform GSEA with the 50 hallmark gene set (v7.2) from MSigDB (Liberzon et al. 2015). Significantly enriched gene sets were filtered and acquired based on a cutoff level at  $q < 0.01$ . Enriched known oncogenic pathways were examined on a maftools R package. Oncogenic signaling pathways were derived from TCGA cohorts. The values of “fraction mutated samples” were used to compare the influence in oncogenic pathways between the KOCRC and EUCRC cohorts.

To identify consensus molecular subtypes (CMS) of CRC samples, we used the CMSclassifier R package (Guinney et al. 2015). Transcriptomic data was initially normalized to counts per million bases (CPM). Log transformations were subsequently conducted by adding one pseudo-count transformed into a log<sub>2</sub> scale. A random forest classifier method was used to arrange the KOCRC and TCGA samples into four CMS classes. The ambiguous subtypes were designated as ‘unspecified’.

### Statistics

A Wilcoxon signed-rank test was used to determine differences between two dependent samples with unknown distribution, while continuous variables were compared using paired Student's *t*-tests. The chi-square test was used to compare clinical datasets on oncogenic pathways, whereas mutational frequencies between KOCRC and EUCRC gene sets were compared with a Fisher's exact test. All statistical analyses were performed using the limma (Ritchie et al. 2015) R package (ver.

3.48.0), with a two-sided  $p < 0.05$  defined as statistically significant.

## Results

### General clinical features of the KOCRC and EUCRC cohorts

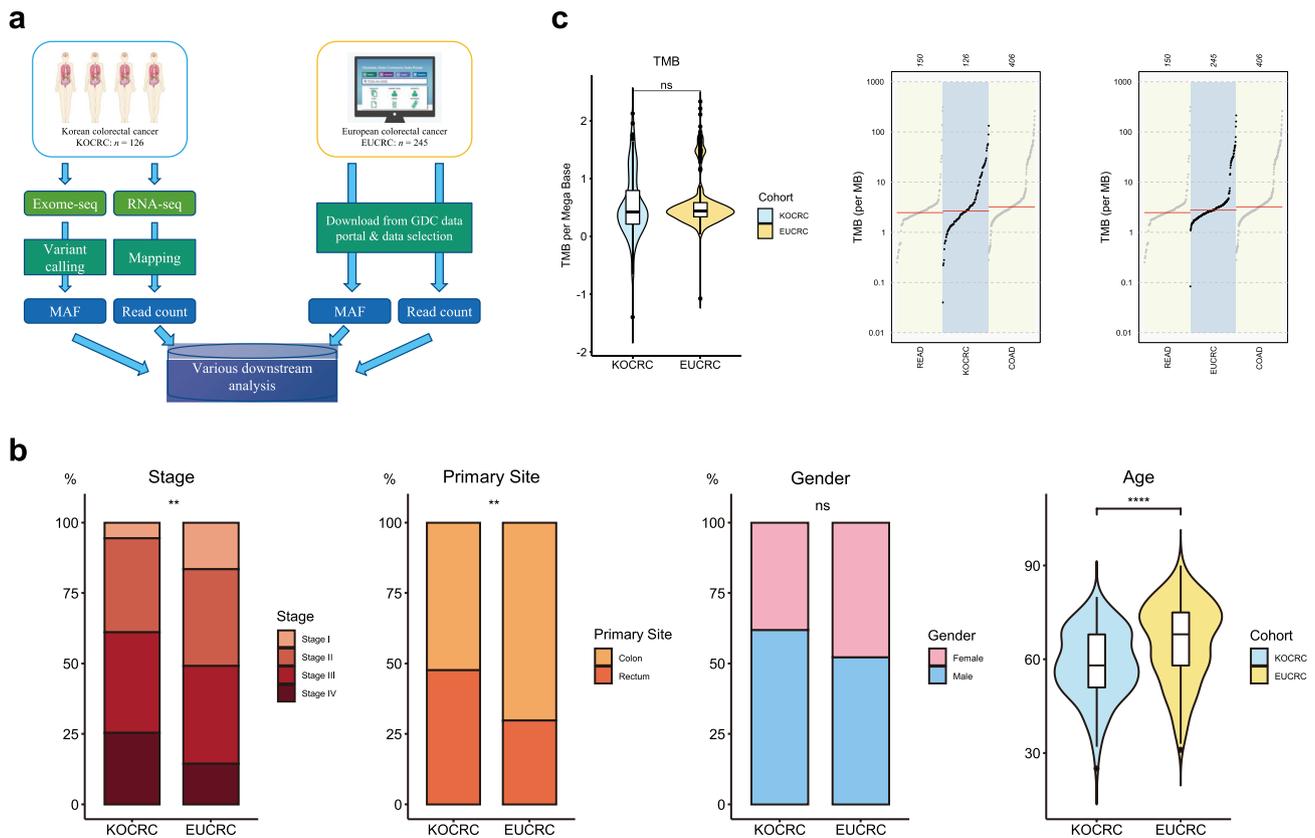
This study was designed to enable genomic comparisons of CRC patients of Korean and European descent, i.e., KOCRC and EUCRC cohorts (Fig. 1a). The clinical features of these cases were also compared, including cancer stage, primary tumor site, and patient demographics (Fig. 1b). The gender ratios were similar between the cohorts ( $p = 0.1$ ), but differences were evident in the cancer stage, primary site, and age ( $p = 0.004$ ,  $0.001$ , and  $3.86 \times 10^8$ , respectively). Age differences were particularly noticeable, with the KOCRC cohort having a median age of about 58, which was ten years younger than of the EUCRC patients.

We estimated the tumor mutation burden (TMB) of the two cohorts (Fig. 1c) and found a median TMB per megabase (TMB/MB) of 2.65 and 2.76, respectively, for the KOCRC and EUCRC populations. It appeared from our analyses that the higher proportion of rectum adenocarcinoma (READ) in the KOCRC cohort may have affected the median TMB/MB (the READ proportions for the KOCRC and EUCRC groups were about 47.6% and 29.8%, respectively) but this was not statistically significant ( $p = 0.13$ ).

### Mutation analysis centered on driver genes

Using the driver detecting software, MutSigCV, we found six previously well-known CRC driver genes (*APC*, *TP53*, *KRAS*, *FBXW7*, *SMAD4*, and *AMER1*) common between the two cohorts. In contrast, three putative novel CRC driver genes (*MUC7*, *PABPC1*, and *B2M*) were identified in the KOCRC cohort at a false discovery rate (FDR) of 0.05. Additionally, we adopted well-known CRC driver genes from Integrative Onco Genomics (Martinez-Jimenez et al. 2020) (intOgen) and other previous studies for these comparative analyses (Hanna et al. 2013; Lu et al. 2019). A gene set of 25 driver genes was used in further analyses (Fig. 2a).

The most frequently mutated driver genes in the KOCRC cohort were *APC* (53%), *TP53* (42%), *KRAS*



**Fig. 1** Workflow and clinical data comparisons. **(a)** Workflow of this study. **(b)** Clinical data comparison between the KO CRC and EUCRC cohorts. Asterisks are labeled according to the  $p$ -values calculated. The  $p$ -values for stage, primary site, gender, and age were 0.004032, 0.001053, 0.09634, and 3.86e-08, respectively (KO CRC:

n=126, EUCRC: n=245). **(c)** TMB comparisons. The first plot shows a direct comparison between the KO CRC and EUCRC populations, and the next two plots compare each cohort with TCGA-COAD and TCGA-READ

(37%), *PABPC1* (21%), and *PIK3CA* (21%) (Fig. 2a). In terms of mutation frequency, most of the driver genes showed similar tendencies between the two cohorts, except for *APC*, *TP53*, *PABPC1*, *FAT4*, *MUC7*, *HSPG2*, *GNAS*, *DENND5B*, and *BRAF* (Fig. 2b, c and Supplementary Fig. 1). Mutations in the *APC*, *TP53*, *FAT4*, and *BRAF* genes were more frequent in the EUCRC cases, whereas those of *PABPC1*, *MUC7*, *HSPG2*, *GNAS*, and *DENND5B* were more frequent in the KO CRC series (Fig. 2b, c). Regarding the three putative novel drivers identified in the KO CRC cohort, *MUC7*, *PABPC1*, and *B2M* were mutated in 19, 27, and 7 samples, respectively, out of the 126 total KO CRC samples.

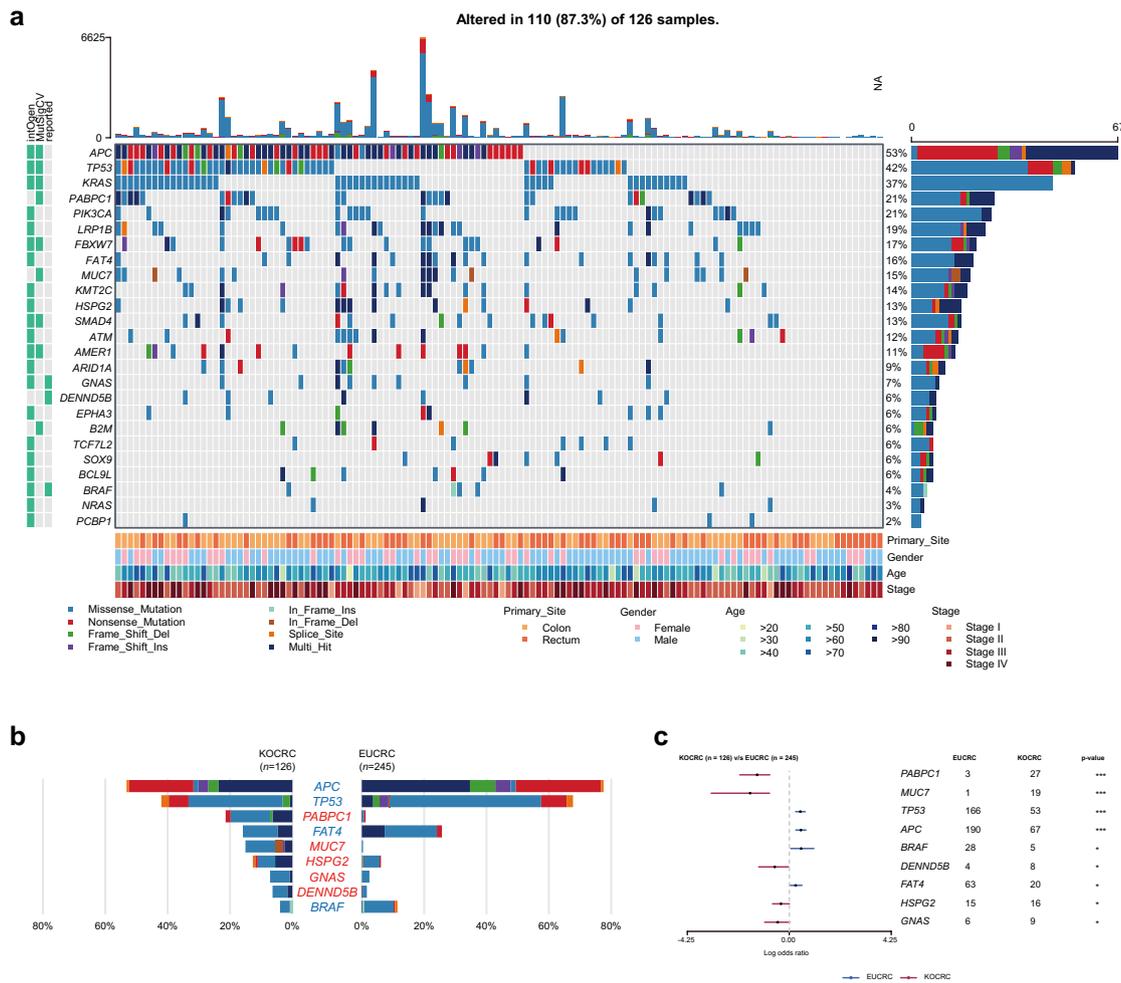
### Mutational signature analysis

We used the NMF algorithm to identify mutational signatures in the KO CRC and EUCRC patients and calculated

cosine similarities against single base substitution (SBS) COSMIC (Tate et al. 2019) signatures to identify the best matches (Fig. 3a, b). We thereby identified ‘defective DNA mismatch repair (dMMR)’ (COSMIC Signature 6), ‘POLE’ (COSMIC Signature 10), ‘unknown’ (COSMIC Signature 5), and ‘sequencing artifact’ (COSMIC Signature 45) in the KO CRC cohort, and ‘aging’ (COSMIC Signature 1), ‘dMMR’, and ‘POLE’ signatures in the EUCRC populations. Both cohorts have ‘dMMR’ and ‘POLE’ signatures, which have also been verified in many other cancer types. The ‘unknown’ signature, COSMIC Signature 5, also arises in all cancer types but remains to be verified.

### GSEA and oncogenic pathways

Using transcriptomic data, we conducted GSEA using 50 hallmark gene sets from MSigDB. To identify significantly



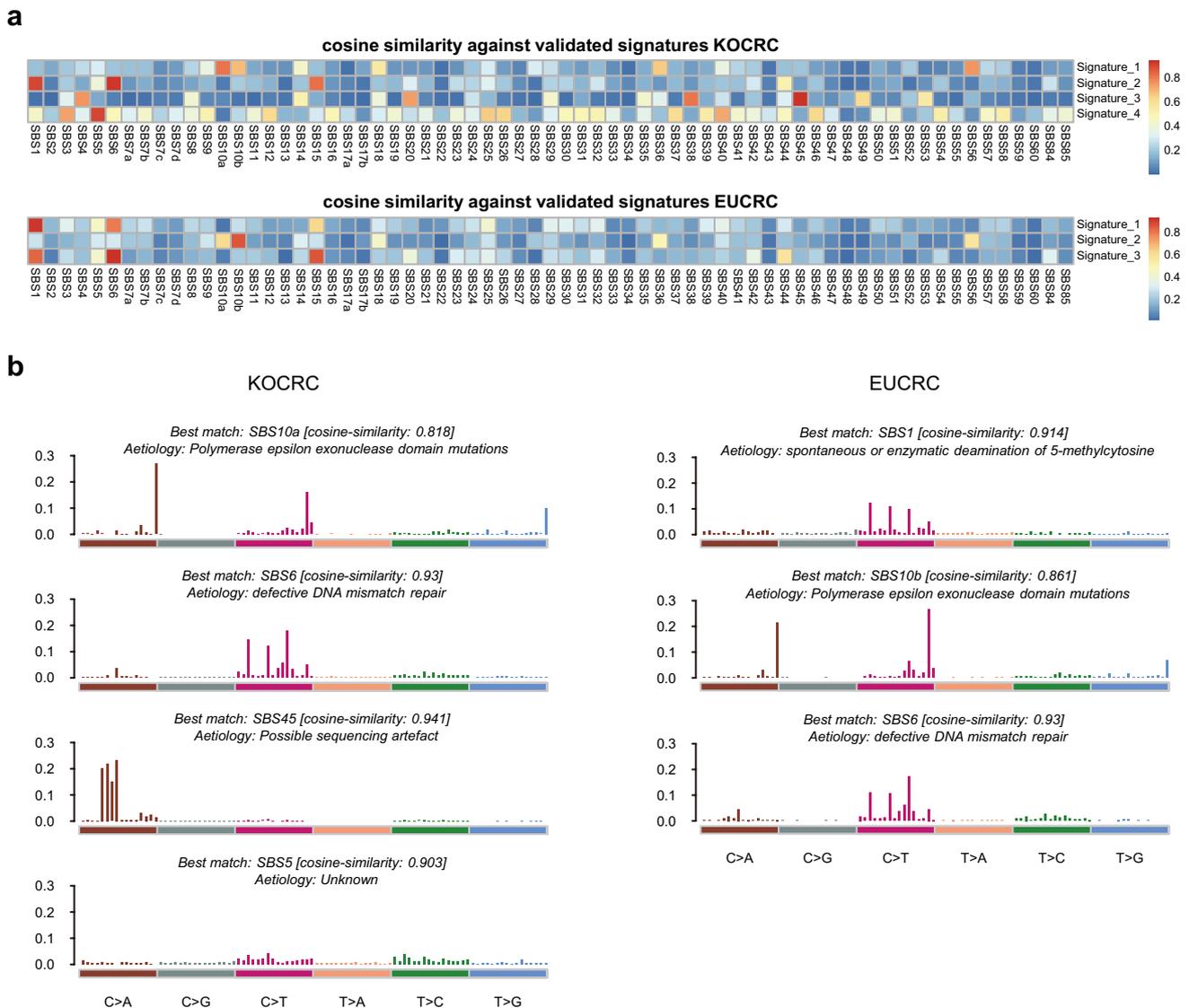
**Fig. 2** Mutation analysis of driver genes. **(a)** Mutational profiles of the KOCRC cohort are shown with clinical data. The annotations for driver genes (intOgen, MutSigCV, and reported) are indicated on the left side. **(b)** Comparison of the mutation frequencies of driver genes between the KOCRC and EUCRC cohorts. Only genes with sig-

nificant differences in frequency are shown ( $p$ -value < 0.05). A  $2 \times 2$  Fisher’s exact test was performed for each gene. **(c)** Forest plot of differently mutated genes for  $p$ -values < 0.05 between the KOCRC and EUCRC groups

enriched gene sets, we applied an FDR cutoff of 0.01 (Fig. 4a). The results indicated that seven hallmark gene sets were significantly enriched in both cohorts, whereas another 12 and 5 were exclusively enriched only in the KOCRC and EUCRC groups, respectively (Fig. 4b). The 12 gene sets enriched and up-regulated only in the KOCRC cohort were ‘mitotic spindle’, ‘G2M checkpoint’, ‘adipogenesis’, ‘myogenesis’, ‘interferon gamma response’, ‘unfolded protein response’, ‘PI3K/AKT/mTOR signaling’, ‘MYC targets v2’, ‘epithelial mesenchymal transition’, ‘inflammatory response’, ‘IL2 STAT5 signaling’, and ‘peroxisome’.

The five gene sets enriched and down-regulated only in the EUCRC cohort were ‘TNF $\alpha$  signaling via NF $\kappa$ B’, ‘protein secretion’, ‘apical surface’, ‘oxidative phosphorylation’, and ‘reactive oxygen species pathway’.

We used ten canonical oncogenic signaling pathways derived from TCGA cohorts (Sanchez-Vega et al. 2018) (Fig. 4c) to perform pathway analysis. Pathway analyses were performed using genomic data. In most pathways, the frequencies of affected samples were similar in both cohorts. However, in the  $\beta$ -catenin/WNT and p53 signaling pathways, significantly more fractions of samples were affected in the



**Fig. 3** Mutational signatures among the KOCRC and EUCRC patients. **(a)** Heat maps of cosine similarities between a group of SBS COSMIC signatures (v3) and the mutational signatures of each cohort. The mutational signatures for the KOCRC and EUCRC popu-

lations were divided into four groups using the NMF algorithm. Each mutational signature found by this algorithm was compared to the SBS COSMIC signature (v3). **(b)** Plots of decomposed mutational signatures for the KOCRC and EUCRC cohorts

EUCRC cohort, whereas the Notch signaling pathway had a higher fraction of affected samples in the KOCRC cohort.

### CMS classification

A prior study established four CMSs for CRC and developed a tool named ‘CMSclassifier’ (Guinney et al. 2015). To investigate how well our data fitted with existing findings, we utilized ‘CMSclassifier’ to analyze our transcriptomic data from both the KOCRC and EUCRC cohorts.

We first compared the proportions of each CMS in the two cohorts after deducting the ‘unspecified’ subtype (Fig. 5a, b, Supplementary Fig. 2). The prevalence of CMS1-4 for the KOCRC cohort were 11.3%, 30.0%, 13.8%, and 45.0%, respectively. For the EUCRC cohort, these percentages were 13.1%, 30.6%, 19.7%, and 36.6%, respectively ( $p=0.5215$ ).

We next compared the reported features of each CRC CMS with our current data (Fig. 5a, Supplementary Fig. 2). CMS1 is known as an MSI high and *BRAF* mutation enriched subtype (Guinney et al. 2015). In the EUCRC

cohort, CMS1 samples ( $n=24$ ) showed this expected high MSI and *BRAF* mutation rate (62%), whereas the KOCRC CMS1 samples ( $n=9$ ) showed MSI high features but no *BRAF* mutations. CMS3 is known to have a high frequency of *KRAS* mutation (Guinney et al. 2015). This fact was also found in our cohort, in which CMS3 samples showed the highest *KRAS* mutation frequencies of the four subtypes. The *KRAS* mutation frequency was 55% for KOCRC CMS3 ( $n=11$ ) and 61% for EUCRC CMS3 ( $n=36$ ) (Fig. 5c). Additionally, even though some variation may be anticipated because of the limited number of samples, CMS4 samples for both of our present cohorts tended to have higher proportions of cancer stage IV cases than other subtypes (Fig. 5d).

### Fusion genes in the KOCRCs patients

We used STAR-Fusion software to identify fusion genes in the KOCRC cohort present in at least four patients. Four intrachromosomal fusion genes (*SEPTIN7P2-PSPH*, *OR51S1-TP53I11*, *PTPRK-RSPO3*, and *PMS2P6-CCDC146* in 47, 20, 7, and 7 cases, respectively) and two interchromosomal fusion genes (*YAF2-RYBP* and *FBXO25-SEPTIN14* found in 13 and 6 patients, respectively) were thereby identified.

We then examined whether these six fusion genes had any effects on gene expression. The samples harboring a *PTPRK-RSPO3* fusion showed a dramatic increase in *RSPO3* expression (Wilcoxon test,  $p=2.0357 \times 10^{-5}$ ; Supplementary Fig. 3, Fig. 6a). We observed two different kinds of *PTPRK-RSPO3* fusions that contained either exon1 or exon7 of *PTPRK* and exon2 of *RSPO3* (Fig. 6b). Furthermore, the *PTPRK-RSPO3* fusion showed a mutually exclusive pattern with the *APC* mutation (Fig. 6c).

### Discussion

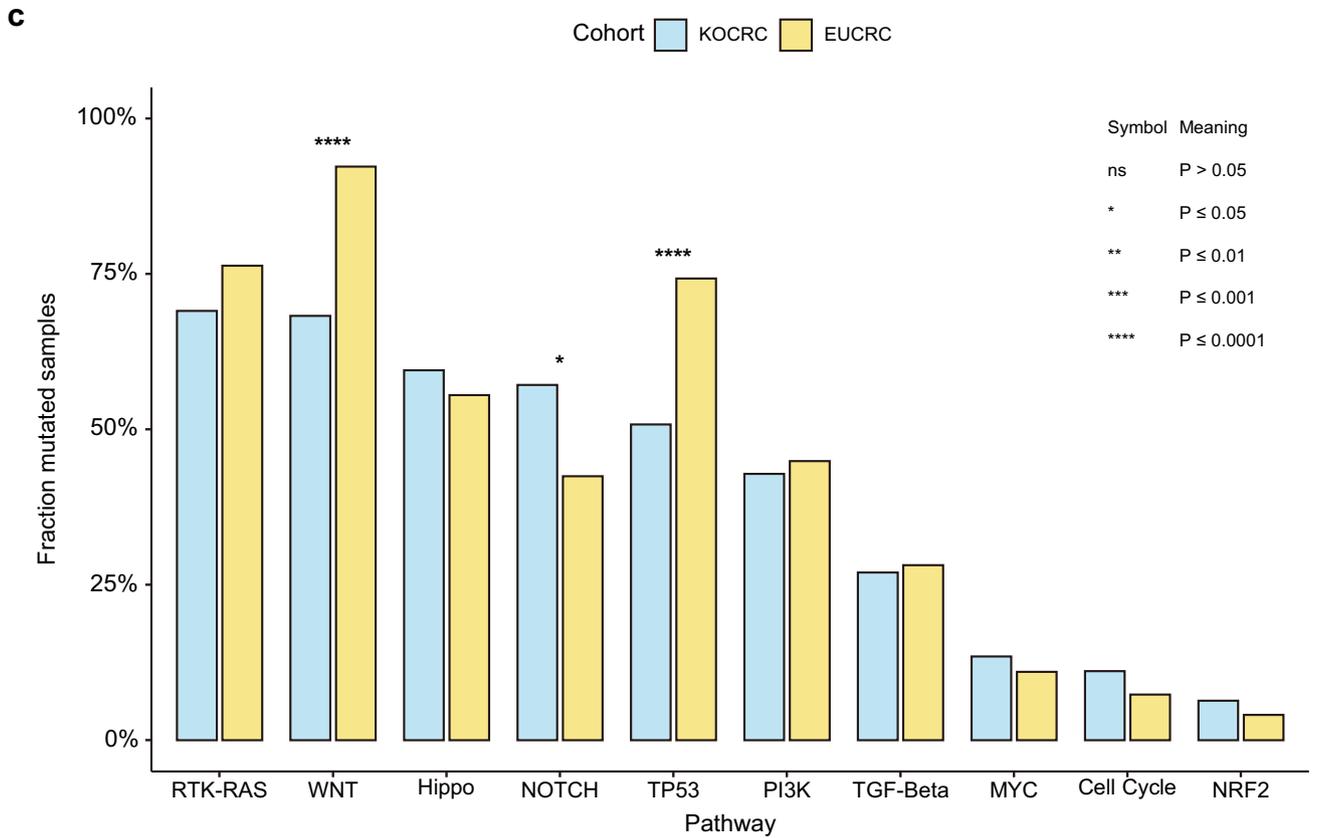
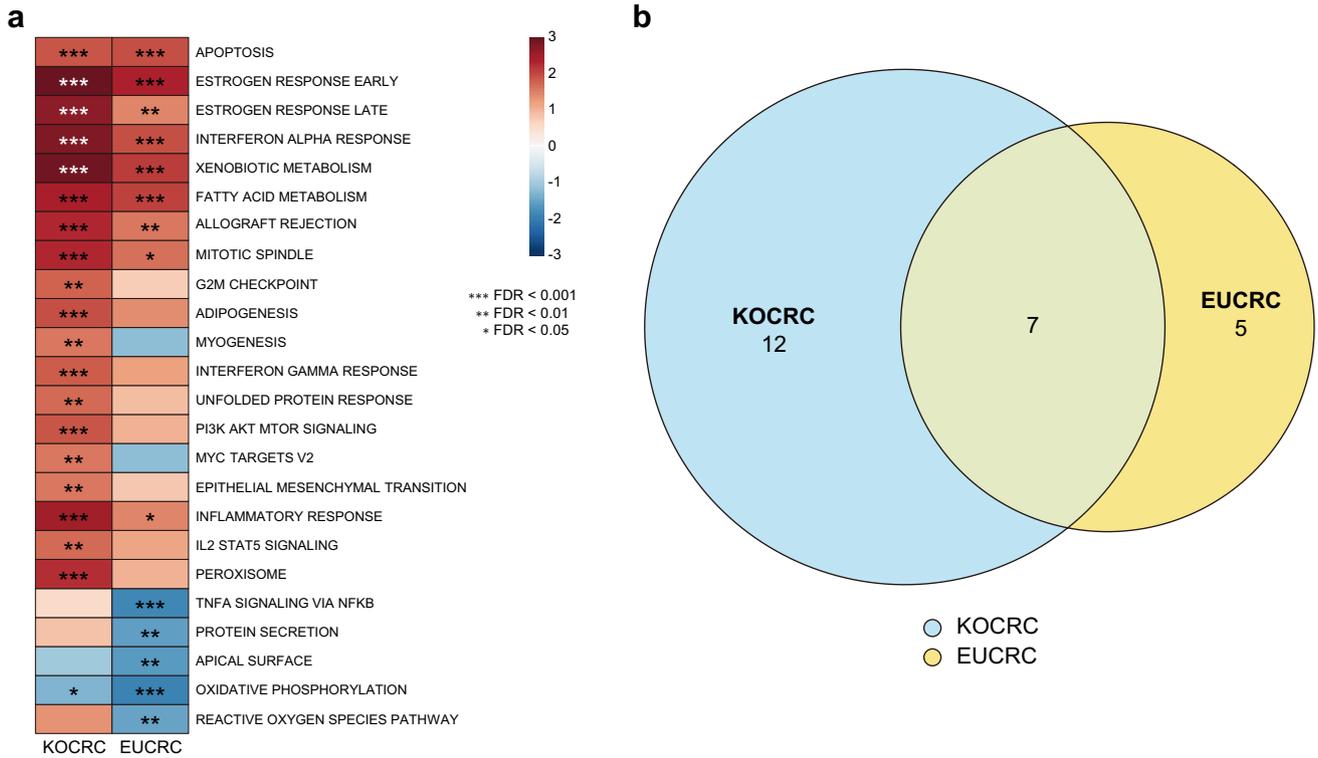
By comparing large cohorts and establishing the genomic landscape of KOCRCs, the commonalities and differences between CRC patients of Korean and European ancestry could be identified and discussed. In the comparative analyses of the clinical data for these populations, it was notable that the KOCRC and EUCRC cohorts showed significant age differences, with a median age of about 58 and 68, respectively. The lower median age of the KOCRC patients is likely to be related to the higher prevalence of this cancer in Korea and the national health checkups for all Korean citizens over the age of 50. These checkups include a CRC

screen using a stool occult blood test and a colonoscopy, which can improve the early diagnosis of CRC.

The KOCRC and EUCRC cohorts in our present study showed differences in the mutation frequencies in several driver genes. Of note, the lower mutation frequency of the *BRAF* gene in our Korean subjects is consistent with another study of CRCs from distinct ethnic groups that also found variations in the *BRAF* mutation frequency (Hanna et al. 2013). In addition, the higher mutation frequencies observed in the *GNAS* and *DENND5B* genes in our KOCRC cases is supported by another study that identified 13 loci that were significantly associated with the risk for CRC in Asians. Two of these 13 loci were located inside or near the protein-coding regions of *GNAS* and *DENND5B* (Lu et al. 2019).

We additionally identified three new putative driver genes (*MUC7*, *PABPC1*, *B2M*) in our KOCRC population. *MUC7* has often been associated with other cancer types, particularly bladder cancer, and its expression levels have been assayed in many tumor types (Retz et al. 1998). However, the significance of *MUC7* mutations in CRC remains uncertain. *PABPC1* (poly A binding protein cytoplasmic1) is known to play a role in the post-transcriptional control of mRNA and may be involved in tumorigenesis (Takashima et al. 2006). In addition, several studies have revealed that this gene has important roles in tumor progression and carcinogenesis in both esophageal and gastric cancer (Takashima et al. 2006; Zhu et al. 2015). *B2M* mutations are often reported in high-level microsatellite instability (MSI-H) CRCs (Tikidzhieva et al. 2012). Robust evidence is available that correlates *B2M* variations and immune escape in CRC (Grasso et al. 2018; Ozcan et al. 2018), and this gene also acts as a driver in diffuse large B cell lymphoma (DLBC) (Fan et al. 2020).

The most frequently mutated genes in our EUCRC cohort were *APC*, *TP53*, *FAT4*, and *BRAF*. These four genes are involved in major carcinogenesis pathways, including the Wnt, Hippo, and MAPK signaling pathways. Of the genes most frequently mutated in the KOCRC cohort, the activating mutation in *GNAS* has been reported previously in *APC* deficient mice to promote intestinal tumorigenesis by activating the Wnt and ERK1/2 MAPK pathways (Wilson et al. 2010). In another prior study, the *GNAS* mutation functioned as an alternative activator of the Wnt/beta-catenin signaling pathway in gastric adenocarcinoma (Nomura et al. 2014). These results suggest that the Wnt/beta-catenin pathway is activated in Korean CRC patients by a *GNAS*-mediated alternative pathway and a canonical *APC* pathway. We speculate that this alternative mechanism of Wnt pathway activation by *GNAS* may partially explain the lower mutational frequency of the *APC* gene in the KOCRC compared to the



**Fig. 4** Analyses of gene sets and pathways among the different CRC cohorts. **(a)** Heat map of the GSEA results for hallmark gene sets. The heatmap was drawn according to normalized enrichment scores (NES). Asterisk labeling is based on FDR values. **(b)** Venn diagram of enriched hallmark gene sets in the KOCRC and EUCRC cohorts. **(c)** Comparison of the mutation frequency of genes in 10 hallmark pathways across the KOCRC and EUCRC patient subjects. Asterisks indicate significant differences based on a chi-square test. The *p*-values for the WNT, NOTCH, and TP53 pathways were 1.64e-09, 4.88e-06, and 0.011, respectively

EUCRC cohort in our current study. However, we predict that the *PTPRK-RSPO3* fusion gene likely plays a role in an alternative mechanism of Wnt pathway activation. The Wnt-dependent endogenous *Rspo2* and *Rspo3* chromosomal rearrangements can initiate and maintain colorectal carcinogenesis (Han et al. 2017). Another previous study has suggested a role for the *PTPRK-RSPO3* fusion gene in activating Wnt/beta-catenin signaling because it showed a mutually exclusive pattern with *APC* or beta-catenin mutations (Hao et al. 2016), which is in line with our present data indicating its mutual exclusiveness with *APC* mutations. Taken together, the cumulative evidence now suggests that two alternative pathways, including *GNAS*-mediated and *PTPRK-RSPO3* fusion-mediated mechanisms, may play an important role in the activation of Wnt/beta-catenin signaling in place of *APC* mutations in Korean CRC lesions. Additionally, *DENND5B*, a guanine nucleotide exchange factor that activates *RAB39A* and *RAB39B*, was previously identified as one of 13 loci significantly associated with risk for CRC in Asians (Lu et al. 2019). Further studies are needed to determine the roles of *DENND5B* in colorectal carcinogenesis.

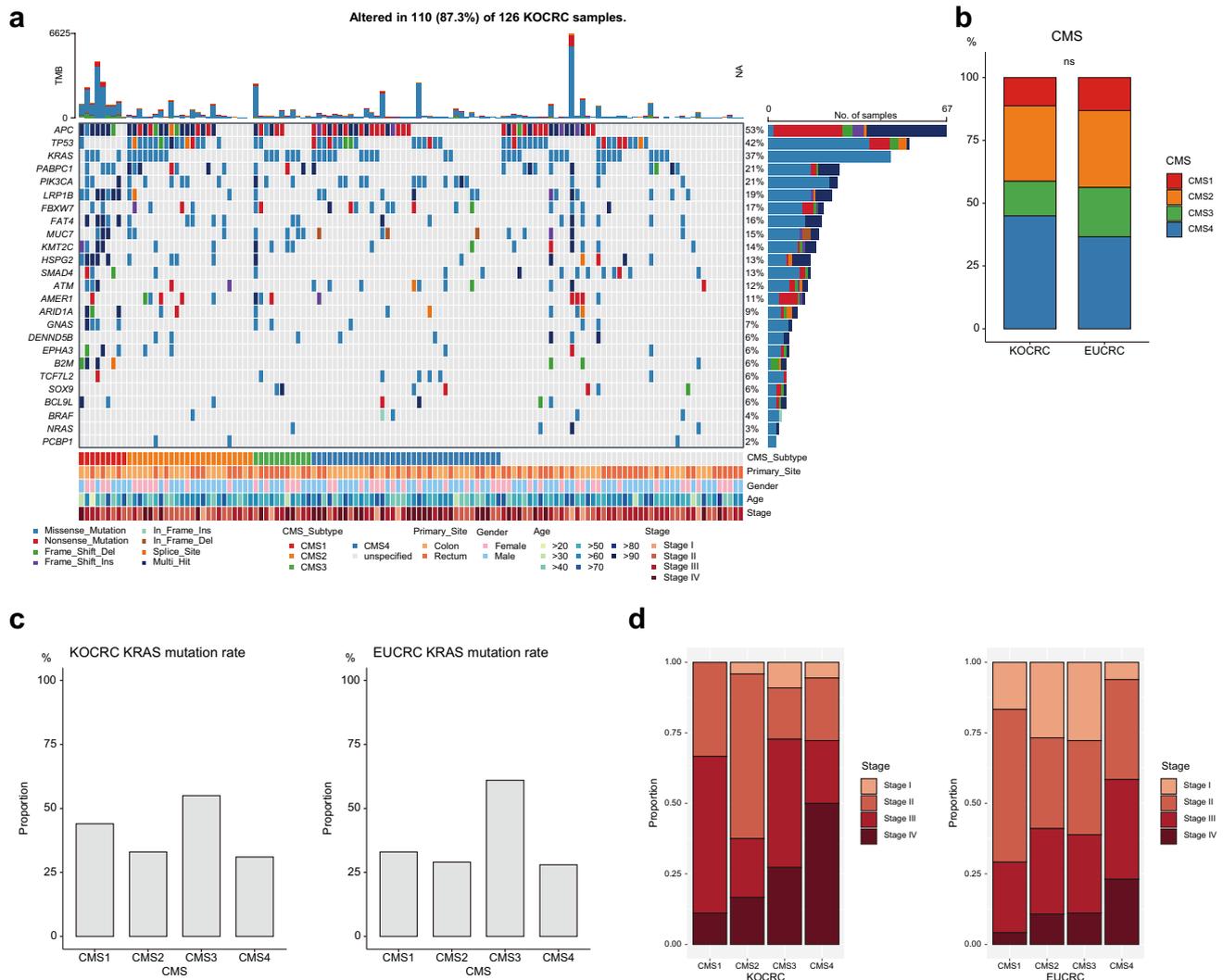
Our current mutational signature analysis results suggested that KOCRCs and EUCRCs are very similar except for the unknown signature (COSMIC Signature 5), indicating that the major mutational signatures are conserved among these two cohorts. The aging signature (COSMIC Signature 1) was evident in EUCRC cases which were not surprising since the median age of the EUCRC cohort was older than that of the KOCRC cohort. *POLE* has a crucial role in chromosomal DNA replication due to its proofreading capacity and is known to be mutually exclusive with dMMR. Somatic mutations in the proofreading domains of *POLE* have been identified in relation to microsatellite instability (MSI), which has been found to occur in CRC due to a dMMR system with key MMR genes inactivated by various mechanisms (Domingo et al. 2016; Kim et al. 2013). Moreover, mutations in polymerase proofreading-associated

syndrome involving *POLE* and *POLD1* constitute 0.3–0.7% of familial cancer cases when only CRC and polyposis are considered (Mur et al. 2020).

In our GSEA and pathway analysis for mutated genes, we identified significant differences in some hallmark gene sets and pathways between the KOCRC and EUCRC patients. These results indicate that Korean CRC cases may require different therapeutic approaches than the current conventional methods. Among the gene sets enriched in KOCRC were upregulated immune-related gene sets such as ‘interferon gamma response’, ‘inflammatory response’, and ‘IL2 STAT5 signaling’, indicating the possibility that immunotherapy-based approaches could be effective in these cases.

In the CMS analysis we conducted in our present series, we assessed the previously established four CRC subtypes (CMS1–4) (Guinney et al. 2015). CMS1 is the MSI immunogenic type, CMS2 is the canonical type, CMS3 is a metabolic type and CMS4 is a mesenchymal type. CMS1 was enriched for MSI tumors and *BRAF*-mutations. CMS2 tumors had epithelial characteristics with marked WNT and *MYC* signaling augmentation and a high CIN. CMS3 cancers also had epithelial features but a lower CIN, were enriched for *KRAS* mutations and presented with evident metabolic dysregulation. The CMS4 grouping was the mesenchymal subtype with prominent TGF- $\beta$  activation, stromal invasion, angiogenesis, and an inflammatory, immunosuppressive phenotype. CMS analyses of CRCs is a new modality that includes knowledge of molecular factors, tumor stroma, and signaling pathways to facilitate personalized, patient-orientated systemic treatments, i.e., precision medicine (Ten Hoorn et al. 2021). In our present study, the proportions of each subtype in the two cohorts did not show differences, implying that they are conserved among different ethnic groups. Additionally, the conserved proportions of each subtype indicated no fundamental differences in the molecular carcinogenesis processes between the two cohorts.

Since gene fusions are closely associated with specific tumor phenotypes, they represent ideal targets for anticancer treatments and risk stratification. A previous study reported that the fusion of *NAV2* and *TCF7L1* is a new marker for aggressive CRC and has an important role in *MYC*-directed transcriptional activation and repression (Cancer Genome Atlas Network 2012). We identified several new fusion genes that may become oncogenic candidates for CRC in this present study. A previous report identified a *PTPRK-RSPO3* fusion gene in CRC and demonstrated that targeting *RSPO3*

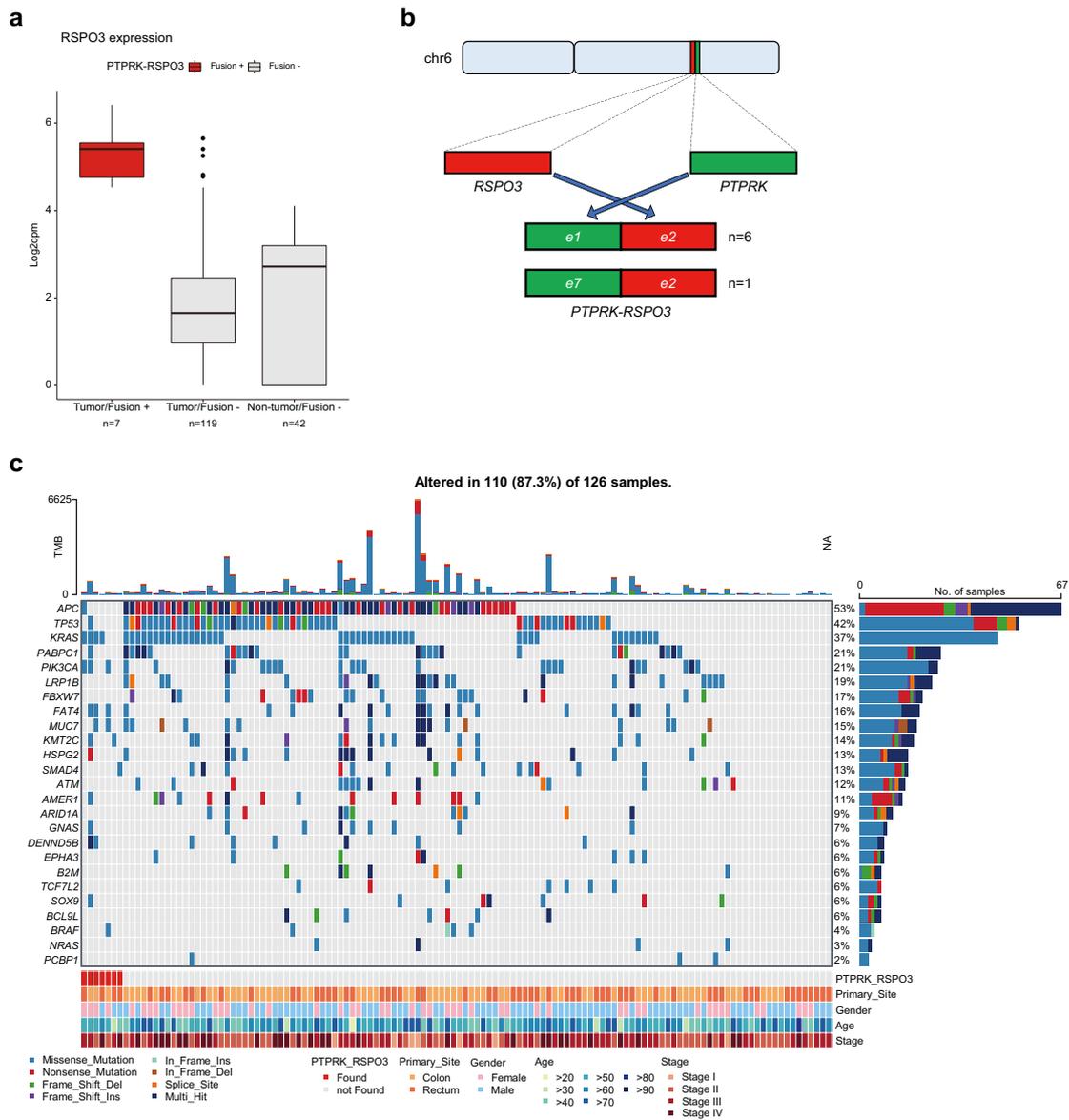


**Fig. 5** Consensus molecular subtype (CMS) analysis. **(a)** Mutational profiles of the KOCRC cases stratified by CMS. **(b)** Stacked bar plot showing CMS distribution in the KOCRC and EUCRC cohorts. **(c)** Bar plots showing the *KRAS* mutation rate in each cohort according to the CMS. **(d)** Stacked bar plots showing the cancer stage distributions for each cohort according to the CMS

in *PTPRK-RSPO3* fusion-positive human tumor xenografts inhibited tumor growth and promoted differentiation (Storm et al. 2016). Although the precise functions of the fusion genes found in CRC remain to be defined, our current data in two ethnically different cohorts suggest that gene fusion events may contribute to tumorigenesis in this cancer type.

Overall, we suggest from our present analyses that further studies involving larger populations of Korean CRC cases are needed to validate our current data. In addition, as the data from the KOCRC and EUCRC cohorts in our series were

processed using partially different computational procedures, caution should be exercised in interpreting our results which may have been affected by this. However, the effect would be trivial, as we followed most of the computational procedures as GDC Data Portal stated (Supplementary Table 1). Notwithstanding these limitations, our present study suggests that distinct molecular and genomic differences exist between Korean and European CRCs, and our analyses provide an important



**Fig. 6** Fusion gene analysis. (a) Boxplot showing *RSPO3* expression on a  $\text{Log}_2$  scale in cpm, according to the presence of the *PTPRK-RSPO3* fusion gene. (b) Schematic diagram of the *PTPRK-RSPO3*

fusion gene. (c) Mutational profiles of the KOCRC patients with additional information on the presence of the *PTPRK-RSPO3* fusion gene

reference point for the future genetic testing of cancer risk and potential targeted treatments in Korean CRC patients.

writing: JSA, HYJ, KSY, KJC. All authors have read and approved the manuscript.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s13258-022-01275-4>.

**Data availability** The data sets used in this study are available from GEO (GSE50760, GSE107422, GSE132024) and KoNA (PRJKA210050).

**Acknowledgements** This work was supported by grants from the National Research Foundation of Korea (NRF-2020M3E5D7085172 and NRF-2021M3H9A1030267 to KSY).

**Declarations**

**Author contributions** Study design: KSY and KJC. Data collection and analysis: JSA, HYJ, KJH, KJH, KSK, and KSY. Manuscript

**Conflict of interest** SAJ, YJH, JHK, JHK, SKK, YSK, SYK, and JCK declare that they have no conflict of interest.

**Ethical approval** This study had been approved by the Institutional Review Board of Asan Medical Center (registration numbers: 2009–0091, 2014–0150, 2018–0087). Informed consent was obtained from all individual participants included in the study.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F (2017) Global patterns and trends in colorectal cancer incidence and mortality. *Gut* 66:683–691
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68:394–424
- Cancer Genome Atlas Network (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487:330–337
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21
- Domingo E, Freeman-Mills L, Rayner E, Glaire M, Briggs S, Vermeulen L, Fessler E, Medema JP, Boot A, Morreau H et al (2016) Somatic POLE proofreading domain mutation, immune response, and prognosis in colorectal cancer: a retrospective, pooled biomarker study. *Lancet Gastroenterol Hepatol* 1:207–216
- Ellis MJ, Perou CM (2013) The genomic landscape of breast cancer as a therapeutic roadmap. *Cancer Discov* 3:27–34
- Fan Z, Pei R, Sha K, Chen L, Wang T, Lu Y (2020) Comprehensive characterization of driver genes in diffuse large B cell lymphoma. *Oncol Lett* 20:382–390
- Fearon ER (2011) Molecular genetics of colorectal cancer. *Annu Rev Pathol* 6:479–507
- Grasso CS, Giannakis M, Wells DK, Hamada T, Mu XJ, Quist M, Nowak JA, Nishihara R, Qian ZR, Inamura K et al (2018) Genetic Mechanisms of Immune Evasion in Colorectal Cancer. *Cancer Discov* 8:730–749
- Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Sonesson C, Marisa L, Roepman P, Nyamundanda G, Angelino P et al (2015) The consensus molecular subtypes of colorectal cancer. *Nat Med* 21:1350–1356
- Han T, Schatoff EM, Murphy C, Zafra MP, Wilkinson JE, Elemento O, Dow LE (2017) R-spondin chromosome rearrangements drive Wnt-dependent tumour initiation and maintenance in the intestine. *Nat Commun* 8:15945
- Hanna MC, Go C, Roden C, Jones RT, Pochanard P, Javed AY, Javed A, Mondal C, Palessandolo E, Van Hummelen P et al (2013) Colorectal cancers from distinct ancestral populations show variations in BRAF mutation frequency. *PLoS ONE* 8:e74950
- Hao HX, Jiang X, Cong F (2016) Control of Wnt receptor turnover by R-spondin-ZNRF3/RNF43 signaling module and its dysregulation in cancer. *Cancers (basel)* 8:54
- Ten Hoorn S, de Back TR, Sommeijer DW, Vermeulen L (2021) Clinical value of consensus molecular subtypes in colorectal cancer: a systematic review and meta-analysis. *J Natl Cancer Inst*
- Horibata S, Alyateem G, DeStefano CB, Gottesman MM (2020) The evolving AML genomic landscape: therapeutic implications. *Curr Cancer Drug Targets* 20:532–544
- Kim TM, Laird PW, Park PJ (2013) The landscape of microsatellite instability in colorectal and endometrial cancer genomes. *Cell* 155:858–868
- Kim SK, Kim SY, Kim JH, Roh SA, Cho DH, Kim YS, Kim JC (2014) A nineteen gene-based risk score classifier predicts prognosis of colorectal cancer patients. *Mol Oncol* 8:1653–1666
- Kim JC, Ha YJ, Tak KH, Roh SA, Kim CW, Kim TW, Kim SK, Kim SY, Cho DH, Kim YS (2016) Complex behavior of ALDH1A1 and IGFBP1 in liver metastasis from a colorectal cancer. *PLoS ONE* 11:e0155160
- Kim SK, Kim SY, Kim CW, Roh SA, Ha YJ, Lee JL, Heo H, Cho DH, Lee JS, Kim YS et al (2019) A prognostic index based on an eleven gene signature to predict systemic recurrences in colorectal cancer. *Exp Mol Med* 51:1–12
- Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A (2021) Fast gene set enrichment analysis. *bioRxiv*
- Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA et al (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499:214–218
- Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P (2015) The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst* 1:417–425
- Lu Y, Kweon SS, Tanikawa C, Jia WH, Xiang YB, Cai Q, Zeng C, Schmit SL, Shin A, Matsuo K et al (2019) Large-scale genome-wide association study of east asians identifies loci associated with risk for colorectal cancer. *Gastroenterology* 156:1455–1466
- Martinez-Jimenez F, Muinos F, Sentis I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, Mularoni L, Pich O, Bonet J, Kranas H et al (2020) A compendium of mutational cancer driver genes. *Nat Rev Cancer* 20:555–572
- Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP (2018) Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* 28:1747–1756
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M et al (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
- Mur P, Garcia-Mulero S, Del Valle J, Magraner-Pardo L, Vidal A, Pineda M, Cinnirella G, Martin-Ramos E, Pons T, Lopez-Doriga A et al (2020) Role of POLE and POLD1 in familial cancer. *Genet Med* 22:2089–2100
- Nagahashi M, Wakai T, Shimada Y, Ichikawa H, Kameyama H, Kobayashi T, Sakata J, Yagi R, Sato N, Kitagawa Y et al (2016) Genomic landscape of colorectal cancer in Japan: clinical implications of comprehensive genomic sequencing for precision medicine. *Genome Med* 8:136
- Nomura R, Saito T, Mitomi H, Hidaka Y, Lee SY, Watanabe S, Yao T (2014) GNAS mutation as an alternative mechanism of activation of the Wnt/ $\beta$ -catenin signaling pathway in gastric adenocarcinoma of the fundic gland type. *Hum Pathol* 45:2488–2496
- Ozcan M, Janikovits J, von Knebel DM, Kloor M (2018) Complex pattern of immune evasion in MSI colorectal cancer. *Oncoimmunology* 7:e1445453

- Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP (2006) GenePattern 2.0. *Nat Genet* 38:500–501
- Retz M, Lehmann J, Roder C, Plotz B, Harder J, Eggers J, Pauluschke J, Kalthoff H, Stockle M (1998) Differential mucin MUC7 gene expression in invasive bladder carcinoma in contrast to uniform MUC1 and MUC2 gene expression in both normal urothelium and bladder carcinoma. *Cancer Res* 58:5662–5666
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47
- Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140
- Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, Dimitriadoy S, Liu DL, Kantheti HS, Saghaforia S et al (2018) Oncogenic signaling pathways in the cancer genome atlas. *Cell* 173(321–337):e310
- Storm EE, Durinck S, de Sousa MF, Tremayne J, Kljavin N, Tan C, Ye X, Chiu C, Pham T, Hongo JA et al (2016) Targeting PTPRK-RSPO3 colon tumours promotes differentiation and loss of stem-cell function. *Nature* 529:97–100
- Takashima N, Ishiguro H, Kuwabara Y, Kimura M, Haruki N, Ando T, Kurehara H, Sugito N, Mori R, Fujii Y (2006) Expression and prognostic roles of PABPC1 in esophageal cancer: correlation with tumor progression and postoperative survival. *Oncol Rep* 15:667–671
- Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E et al (2019) COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res* 47:D941–D947
- Tikidzhieva A, Benner A, Michel S, Formentini A, Link KH, Dippold W, von Knebel DM, Kornmann M, Kloor M (2012) Microsatellite instability and Beta2-Microglobulin mutations as prognostic markers in colon cancer: results of the FOGT-4 trial. *Br J Cancer* 106:1239–1245
- Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38:e164
- Wilson CH, McIntyre RE, Arends MJ, Adams DJ (2010) The activating mutation R201C in GNAS promotes intestinal tumorigenesis in *Apc<sup>Min/+</sup>* mice through activation of Wnt and ERK1/2 MAPK pathways. *Oncogene* 29:4567–4575
- Xie YH, Chen YX, Fang JY (2020) Comprehensive review of targeted therapy for colorectal cancer. *Signal Transduct Target Ther* 5:22
- Zhu J, Ding H, Wang X, Lu Q (2015) PABPC1 exerts carcinogenesis in gastric carcinoma by targeting miR-34c. *Int J Clin Exp Pathol* 8:3794–3802

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.