의학 석사 학위논문

Establishment and Characterization of 30 Human Colorectal Cancer Cell lines: Analysis of Signal Pathway and Mutations with Drug Response

대장암 유래 세포주 30종의 수립과 특성분석 : 항암제 반응성과 함께 신호 전달 체계적 접근과 돌연변이분석

2022년 8월

서울대학교 대학원

의과학과 의과학 전공

이 승 범

Establishment and Characterization of 30 Human Colorectal Cancer Cell lines: Analysis of Signal Pathway and Mutations with Drug Response

지도 교수 구 자 록

이 논문을 의학석사 학위논문으로 제출함

2022년 8월

서울대학교 대학원

의과학과 의과학 전공

이 승 범

이승범의 의학석사 학위논문을 인준함

2022년 8월

위 원 장 _____ **(인)**

부위원장 _____ **(인)**

위 원 _____ **(인)**

Abstract

이 승 범 (Seungbeom Lee)

의과학과 (Medical Science Major)

The Graduate School

Seoul National University

Colorectal cancer (CRC) is one of the most common cancer worldwide. In Korea, the incidence of CRC has increased sharply in both men and women. There are diverse risk factors for CRC such as defects in DNA replications, genetic mutations and expressional aberrations.

Previous studies have discovered genomic alterations as well as expressional patterns of CRC by using next generation sequencing (NGS). Nonetheless, the direct integration of such molecular factors with drug responses has not been thoroughly investigated mainly due to lack of well-characterized preclinical model.

Here in this study, we established 30 CRC patient-derived cell lines (PDCs) They were then subjected to extensive genomic, transcriptomic sequencings and high-throughput screening using 21

clinically-relevant drugs. Our results demonstrated that the transcriptomic landscape of the 30 CRC cell lines can be categorized into three subgroups. AXL, AKAP12, AFAP1-AS1 largely accounted for the separation of the subgroups via PI3K/AKT pathway. Also, our study identified the association of molecular signatures with certain drug responses, and validated the use of PDCs in estimating patient-specific chemo-sensitivity.

Keywords: Colorectal Cancer, Drug Screening, Cohort Analysis

Student number: 2019-22429

Table of Contents

List of Tables

List of Figures

Introduction

Colorectal (CRC) is the third most common cancer worldwide, and second leading cause of tumor-related morality. In South Korea, CRC was the third common cancer with the estimated incidence rates of 54.4 per 100,000 persons in 2021 [1].

Traditionally, there have been 3 distinct ways to explain the occurrence of CRC. Microsatellite Instability (MSI), CpG island methylator phenotype (CIMP), and Chromosomal instability (CIN) are the most widely acceptable carcinogenic mechanisms of CRC. Microsatellites also known as Short Tandem Repeats (STRs) are composed of repeated sequences of 1-6 nucleotides [2]. Because of their structure, microsatellites have high mutation rates when replicating DNA. In normal tissue, DNA mismatch repair (MMR) system can detect DNA replication errors. However, defects in MMR genes cannot fix these replication errors so the possibility of gene mutations is increased [3].

CIMP was introduced by Toyota and colleagues [4] and is characterized by methylation of CpG island near promoter regions of some genes [5]. Methylation in CpG island near tumor suppressor

gene promoter could inhibit binding of some transcription factors [6]. By inhibiting these genes expression, CIMP is believed to contribute to carcinogenesis and progression of CRC [7, 8].

CIN is changes in chromosome number and structure [9–12]. Gain and loss of chromosome segments, chromosomal rearrangement, and loss of heterogeneity (LOH) are in this change. These changes affect the expression of tumor–associated genes, and/or genes that regulate cell proliferation or cell cycle that may activate pathways crucial for CRC initiation and progression [13, 14].

CRC development has also a stepwise manner, and each step is related with changes of molecular levels. The traditional steps of carcinogenesis of CRC are called 'Vogelgram' which has explained well about the progression of CRC in terms of an accumulation of mutations [15]. Suppression of tumor suppressor gene, Adenomatous polyposis coli (APC) is observed usually in the initial step as early adenoma. Then, activation of oncogene KRAS is related with transition from early adenoma to intermediate adenoma. Genomic level alterations such as loss of chromosome 18, Deleted in Colon Cancer (DCC) is observed in transition from intermediate

adenoma to late adenoma. Inactivation of tumor suppressor p53 (TP53) and gain of chromosome 8q are related with late adenoma–carcinoma transition. Additional genomic changes could lead to metastasis [16].

Although these studies have shed light on the pathogenesis of CRC, the prevalence of CRC is still increasing. Well–characterized CRC cell lines with clear genetic background with its ability to propagate infinitely have been used as a suitable preclinical model for biological and molecular research for novel therapeutics. In addition to this, cell lines have many advantages such as cost effectiveness, easiness to use, unlimited supplement of material and a pure population of cells.

In this study, we have established and characterized 30 human colorectal cancer cell lines to identify genomic mutations and expressional patterns that affect the drug response of CRC cell lines. For transcriptomic characterization of cell lines, we used principle component analysis (PCA) for separating clusters and then identified significant genes that accounted for the grouping of such cell lines. For further analysis, we investigated multiple fusion genes. We also looked into well–known oncogenes such as *KRAS*, *APC*, and *TP53*.

At the end of this study, we integrated these genomic mutations, altered signaling pathways, fusion genes for validating the availabilities of cell lines via analyzing the responses of 21 anticancer drugs.

Materials and Methods

2.1. Ethics statement

The research protocol was reviewed and approved by the institutional review board of the Seoul National University Hospital (IRB No. 1102-098-357). The study was performed in accordance with the Declaration of Helsinki. Written informed consents were obtained from all patients enrolled in this study.

2.2. Establishment and maintenance of CRC cell lines.

Cell lines from pathologically proven CRC were established. 30 CRC samples are obtained from Seoul National University Hospital. Most of tumor cells were initially cultured in Opti-MEM I (Thermo Fisher Scientific, MA, USA) supplemented with 5% fetal bovine serum (FBS). Established cell lines were sustained in RPMI 1640 medium with 10% FBS and 1% (v/v) penicillin and streptomycin (10,000U/ml). Cultures were maintained in humidified incubators at 37℃ in an atmosphere of 5% CO2 and 95% air. The initial passage was assigned when substantial tumor growth was detected, and successive passages were given at sub-confluence after

trypsinization. When one culture population contains both floating and adherent cells, floating cells were gathered by centrifuging the medium and dispersed by pipetting. Established cell lines were deposited to the Korean Cell Line Bank (Seoul, Korea).

2.3. DNA fingerprinting analysis using 26 Short Tandem Repeat (STR) loci and Amelogenin marker

The genomic DNA from each cell line was amplified using an AmpFISTR identifier polymerase chain reaction (PCR) amplification kit (Applied Biosystems, CA, USA). A single cycle of PCR amplified 26 short tandem repeat (STR) markers (CSF1PO, D2S1338, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D19S433, D21S11, FGA, TH01, TPOX and vWA) and an Amelogenin gender-determining marker containing highly polymorphic microsatellite markers. Amplified PCR products were analyzed by an ABI 3500XL Genetic analyzer (Applied Biosystems, CA, USA).

2.4. Drug sensitivity test and analyze

Cell suspensions were prepared in the previously described. According to various growth rates, 2-8 x 10^5 cells/mL were seeded

on 96-well tissue culture plate in 80uL of complete culture medium and incubated in humidified incubators at 37℃ for 24 hours in an atmosphere of 5% CO2 and 95% air. 21 anti-cancer drugs were serially diluted in DPBS and were then add to each well with a volume of 20uL. After 72 hours of incubation, 10ul of CellTiter-Glo (Promega, WI, USA) was added to each well. After 20 minutes of incubation a 37℃, the optical density was measured at fluorescence using Luminoskan Ascent^{TM} (Thermo Fisher Scientific, MA, USA). To compare and analyze between established cell lines, calculates area under curve (AUC) using R program version 3.63 (R Foundation for Statistical Computing, Vienna, Austria) with various packages.

2.5. RNA prep and RNA-Sequencing

Total RNA was isolated from cell lysate using TRIzol (Qiagen, Hilden, Germany) and Qiagen RNeasy kit (Qiagen, Hilden, Germany). Sequencing libraries were prepared using the Illumina TrueSeq stranded total RNA library prep kit.

Paired end sequencing reads of cDNA libraries (101bp) generated from a NovaSeq600 instrument were verified its sequence

quality with FastQC v 0.11.7. For data preprocessing, low quality bases and adapter sequences in reads were trimmed using Trimmomatic v 0.38. The trimmed reads were aligned to the human genome (USCS hg 19) using HISAT v2.1.0, a splice-aware aligner. And then, transcript assembly of known transcripts, novel transcripts, and alternative splicing transcripts was processed by StringTie v1.3.4. Based on the result of that, expressional abundance of transcript and gene were calculated as read count per sample. Consensus molecular subtype (CMS) of each sample was analyzed by R package, CMScaller .

2.6. Mutation profiling and enriched pathway analysis

The analysis was performed using R program version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria) with various pacakages including dplyr, rvest, stringr, tidyr, clusterProfiler, org.Hs.eg.db, ReactomePA. The R package 'ReactomePA' was also used to identify significant sub-networks of differentially mutated genes. To identifying significant mutations, we searched which was reported in ClinVar database

(https://www.ncbi.nlm.nih.gov/clinvar/) for a given rs number. After that, we predicted the weight of each mutation and selected only those with significant influence.

2.7. Whole Exome Sequencing.

Whole-exome capture was performed on all samples with the SureSelect Human All Exon V5 Kit (Agilent Technologies, Tokyo, Japan), using the Bravo automated liquid handler. The captured targets were subjected to sequencing using HiSeq 2500 (Illumina, San Diego, CA, USA) with the pair-end 100bp read option for cell line samples. The sequence data were processed through an in-house pipeline. Briefly, paired-end sequences are firstly mapped to the human genome, where the reference sequence is UCSC assembly hg19 (original GRCh 37 from NCBI, Feb.2009) using the mapping program BWA (version 0.7.12), and generated a mapping result file in BAM format using BWA-MEM. Then, Picard-tools (ver.1.130) were applied in order to remove PCR duplicates, The local realignment process is performed to locally realign reads with BAM files reducing those reads identically match to a position at start into

a single one. Using MarkDuplicates.jar, which requires reads to be sorted. By using Genome Analysis Toolkit, base quality score recalibration (BQSR) and local realignment around insertion, deletions (indels) were performed. Haplotype Caller of GATK (GATKv.3.4.0) was used for variant genotyping for each sample based on the BAM file previously generated (SNP and short indels candidates are detected). Those variants are annotated by SnpEff v4.1g, to vcf file format, filtering with dbSNP for the version of 142 and SNPs from the 1000 genome project. Then, SnpEff was applied to filter databases, including ESP6500, Clinvar, dbNFSP 2.9.

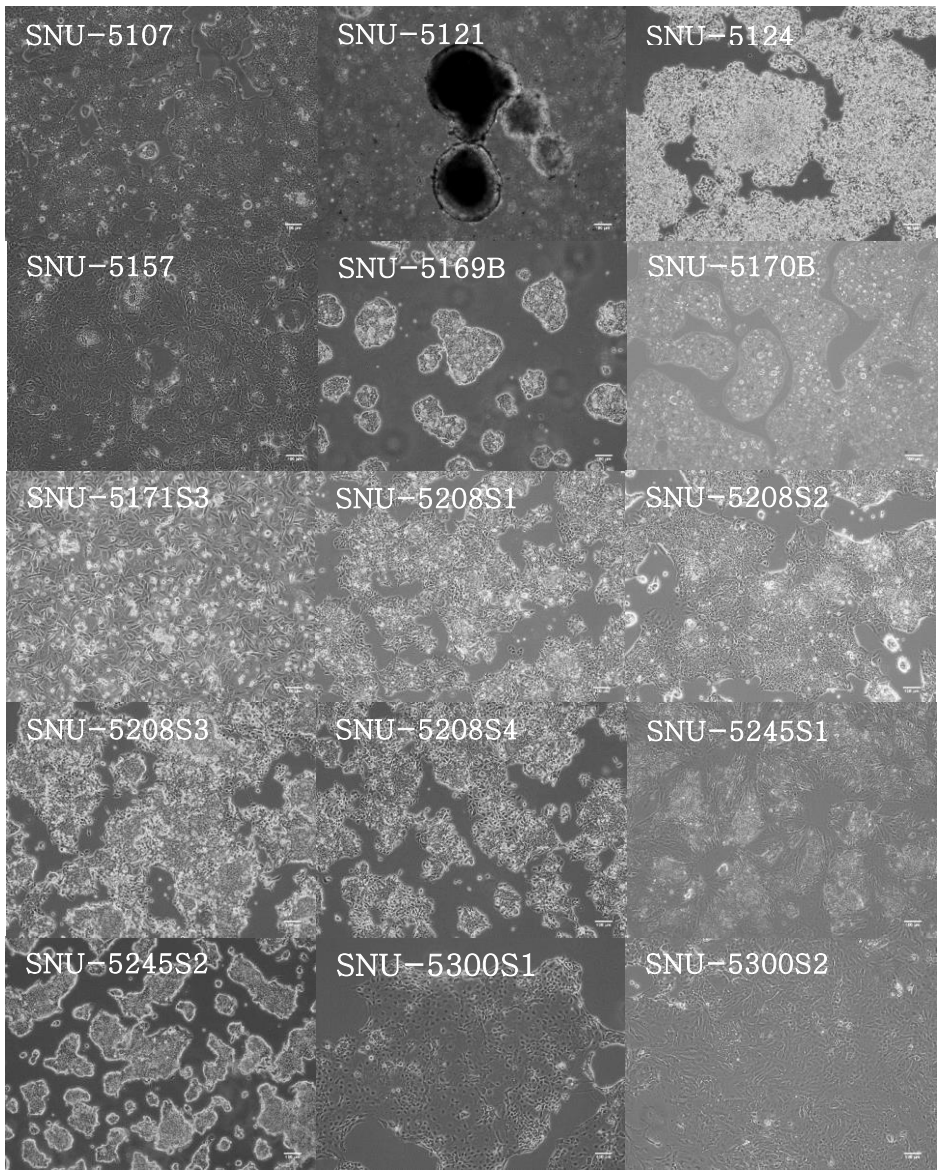
3. Results

3.1 General characteristics of the cell clines

Human specimens were obtained from CRC patients who underwent surgeries at Seoul National University (SNU) Hospital in 2018-2019. 30 CRC cell lines (SNU-5107, SNU-5121, SNU-5124, SNU-5157, SNU-5169B, SNU-5170B, SNU-5171S3, SNU-5208S1, SNU-5208S2, SNU-5208S3, SNU-5208S4, SNU-5245S1, SNU-5245S2, SNU-5300S1, SNU-5300S2, SNU-5300S3, SNU-5475, SNU-5540, SNU-5548, SNU-5584AT, SNU-5614, SNU-5621, SNU-5688, SNU-5715D, SNU-5765, SNU-5903, SNU-6023, SNU-6058S3, SNU-6184, SNU-6091) were established RPMI 1640 medium supplemented with 10% FBS. The clinical characteristic of the established cell lines is summarized in Table 1. 26 STR fingerprinting and Amelogen sex-determining markers were heterogeneously distributed and all newly established cell lines were not cross-contaminated (Table2). All cell lines were free of contamination by mycoplasma test.

Morphology and growth pattern of most cell lines showed adherent cell types but SNU-5121 and SNU-5171S3 showed floating cell

type (Figure 1). Some CRC cell lines growth colony like patten such as SNU-5169B, SNU-5245S1, SNU-5245S2, SNU-5475, SNU-5540, SNU-5715D, SNU-6058S3.



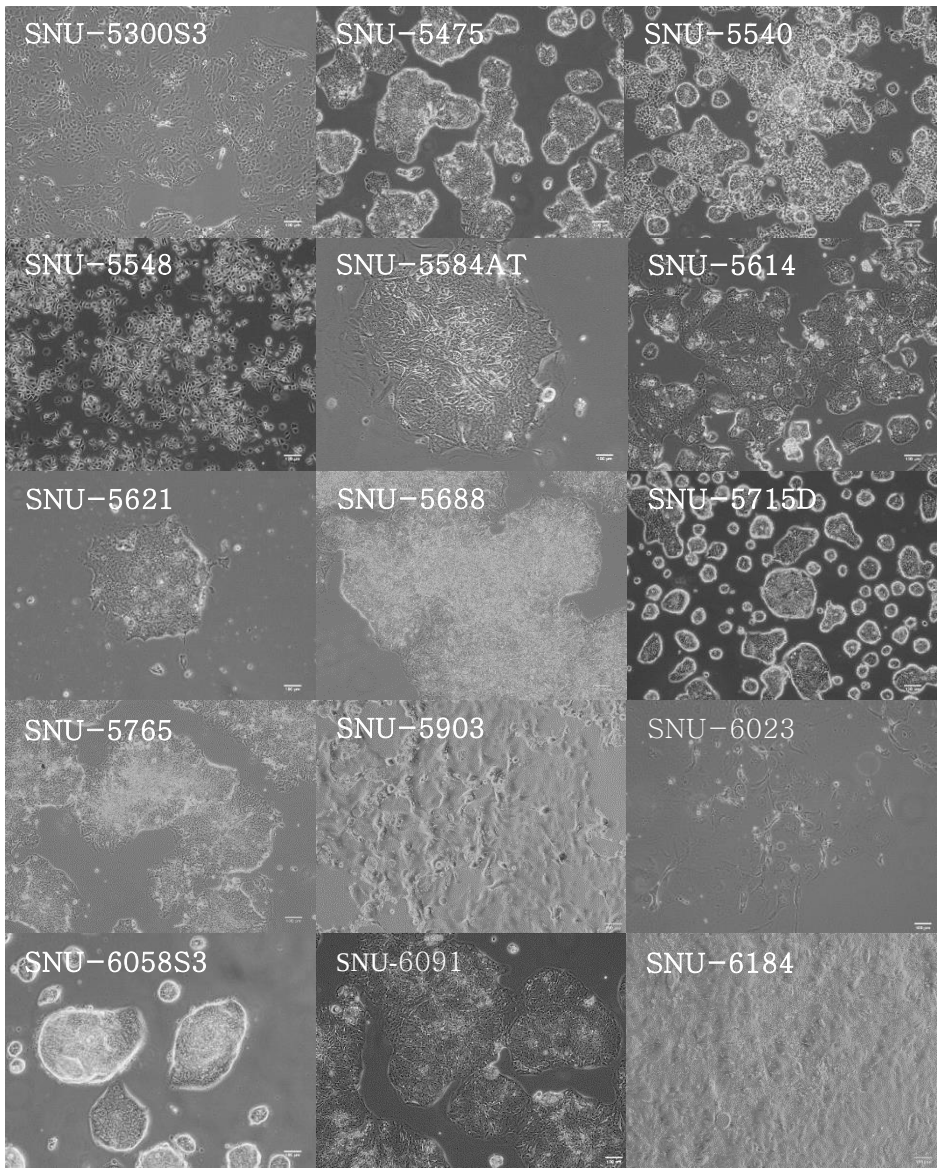


Figure 1. Phase-contrast microscopy of 30 established colorectal cancer cell lines. The white scale bar= 100㎛.

Table1. Clinical characteristics of 30 established colorectal cancer cell lines

No.	Cell line	Sex	Age
1	SNU-5107	M	78
2	SNU-5121	M	74
3	SNU-5124	F	71
4	SNU-5157	F	59
5	SNU-5169B	M	61
6	SNU-5170B	M	82
7	SNU-5171S3	F	79
8	SNU-5208S1	F	76
9	SNU-5208S2	F	76
10	SNU-5208S3	F	76
11	SNU-5208S4	F	76
12	SNU-5245S1	F	38
13	SNU-5245S2	F	38
14	SNU-5300S1	M	74
15	SNU-5300S2	M	74
16	SNU-5300S3	M	74
17	SNU-5475	M	67
18	SNU-5540	M	65
19	SNU-5548	M	60
20	SNU-5584AT	M	46
21	SNU-5614	M	62
22	SNU-5621	M	60
23	SNU-5688	M	73
24	SNU-5715D	M	40
25	SNU-5765	F	71
26	SNU-5903	M	60
27	SNU-6023	F	61
28	SNU-6058S3	M	42
29	SNU-6091	F	60
30	SNU-6184	F	69

Table2. DNA fingerprinting using 26 STR loci and Amelogenin marker

	Cell-Name	Amelogenin	D3S1358	D1S1656	D2S441	D10S1248	D13S317	Penta E	D16S539
1	SNU-5107	X	18	16,17.3	9.1,10	15	8,11	11,13	9
2	SNU-5121	X	16,18	16,17	11,14	13,16	8,9	23	11,13
3	SNU-5124	X	16	13,17	9.1,13	13,14	12	15,17	9
4	SNU-5157	X	15,16	16,17	9.1,11	14,15	11,12	11,12	9,12
5	SNU-5169B	X	15,16	14,18	11	13	8,9	11,15	9,13
6	SNU-5170B	X,Y	15,16	15,16	11,12	13,15	11,12	15	10
7	SNU-5171S3	X	15,17	12,18.3	10,14	13,15	9,10	10,22	9,11
8	SNU-5208S1	X	16	15	11,14	13	10,12	11,18	12,13
9	SNU-5208S2	X	16	15	11,14	13	10,12	18	12,13
10	SNU-5208S3	X	16	15	11,14	13	10,12	11,18	12,13
11	SNU-5208S4	X	16	15	11,14	13	10,12	11,18	12,13
12	SNU-5245S1	X	14,15	15,16	11,13	13,15	9,11	16,18	10,11
13	SNU-5245S2	X	14,15	15,16	11,13	13,15	9	17,18	10,11
14	SNU-5300S1	X,Y	15,17	16	10	13,14	10,11	15,16	10
15	SNU-5300S2	X,Y	15,17	16	10	14	10,11	15,16	10

Continued

	Cell-Name	D18S51	D2S1338	CSF1PO	Penta D	TH01	Vwa	D21S11	D7820	D5S818	TPOX
1	SNU-5107	13	25	12,13	10	6,9	17	29	7,12	12	9,11
2	SNU-5121	14	22,23	12	11	9	16,19	28,31.2	8,11	11	8,9
3	SNU-5124	15,16	19,24	12	11,13	7	18	29,30.2	11,12	13	9
4	SNU-5157	15	17,21	11	9,12	9	16,17	29,30	10,11	10	8,11
5	SNU-5169B	14,16	19,20	11	9,12	7,9	14,18	29,31.2	10	10,11	11
6	SNU-5170B	16	19,27	12	9,10	8,9	17,18	29,30	12	9	9,11
7	SNU-5171S3	13	17,26	10,12	9	9	14,18	30	8,13	10,11	8,11
8	SNU-5208S1	15	19,23	12	12,14	6,9	14	29,29.2	8,12	11	11
9	SNU-5208S2	15	19,23	12	14	6,9	14	29,29.2	8,13	11	11
10	SNU-5208S3	15	19,23	12	12,14	6,9	14	29,29.2	8,12	11	11
11	SNU-5208S4	15	19,23	12	12,14	6,9	14	29,29.2	8,12	11	11
12	SNU-5245S1	17,18	17,22	12	10,11	6,9	16,21	29,32.2	11,12	13	8
13	SNU-5245S2	17	17,22	12	10,11	6,9	16,21	29,32.2	11,12	13	8
14	SNU-5300S1	15	20,23	11,13	11,13	6,9	17	29,30	8,9	9	9,11
15	SNU-5300S2	15	20,23	11,13	11,13	6,9	14,17	29,30	8,9	9	9,11

Continued

	Cell-Name	D8S1179	D12S931	D19S433	D61043	D22S1045	DYS391	FGA	DYS576	DYS570
1	SNU-5107	12	20	14.2	18,19	11,16	-	22,23	-	-
2	SNU-5121	12,13	19,21	13,14	13,14	15	-	19,27,28	-	-
3	SNU-5124	13	18,19	13,14	12,13	17	-	21,24	-	-
4	SNU-5157	12,16	17,19	13,14.2	11,14	11	-	19,22	-	-
5	SNU-5169B	11,14	18,23	14,15.2	11,18	11,17	-	22,24	-	-
6	SNU-5170B	14,15	19	13,15.2	12,18	11,16	10	22	17	17,18
7	SNU-5171S3	12,15	18,20	13,14	17,18	16	-	20	-	-
8	SNU-5208S1	11	18	14,15.2	14,19	16	-	23	-	-
9	SNU-5208S2	11	18	14,15.2	14,19	16	-	23	-	-
10	SNU-5208S3	11	18	14,15.2	14,19	16	-	23	-	-
11	SNU-5208S4	11	18	14,15.2	14,19	16	-	23	-	-
12	SNU-5245S1	10	18,20	14,14.2	11,16	17	-	20,25	-	-
13	SNU-5245S2	10	18,20	14,14.2	11,16	17	-	20,24	-	-
14	SNU-5300S1	13,15	15	13,14	12,19	17	10	24	17	18
15	SNU-5300S2	13,15	15	13,14	12,19	16	10	24	17	18

Continued

	Cell-Name	Amelogenin	D3S1358	D1S1656	D2S441	D10S1248	D13S317	Penta E	D16S539
16	SNU-5300S3	X,Y	15,17	16	10	13,14	10,11	15,16	9,10
17	SNU-5475	X,Y	15	15,16	10,12	13	9,11	19	11
18	SNU-5540	X	15,16	12,13	11	12,16	8,12	15,17	9
19	SNU-5548	X,Y	18	15	11,11.3	17	9,11	18	10,13
20	SNU-5584AT	X	17,19	15	14	13	11	12,21	9
21	SNU-5614	X,Y	16	15	11	15	10,11	12	11,12
22	SNU-5621	X,Y	15	17	10	14	8,11	15	12,13
23	SNU-5688	X	17,19	17,18.3	11,14	13,15	8,11	19,23	9,12
24	SNU-5715D	X,Y	151,7	11,15	10	12,17	10	5	11,12
25	SNU-5765	X	15,16	13,15	11	13,16	9,12	18,20	10
26	SNU-5903	X	16,17	16,17	11	14	8,11	25	10,11
27	SNU-6023	X	15	13,16	11,14	13	8,11	19,20	10,11
28	SNU-6058S3	X	16	15	10,13	14,15	8,9	10,14	9
29	SNU-6184	X	16	12,18.3	11	12	9,12	11,12	11,12
30	SNU-6091	X	14,16	13	10,13	12,14	12	11,20	9,13

Continued

	Cell-Name	D18S51	D2S1338	CSF1PO	Penta D	TH01	Vwa	D21S11	D7820	D5S818	TPOX
16	SNU-5300S3	15	20,23	13	11,13	6,9	17	29,30	8,9	9	11
17	SNU-5475	12	20,23	10,12	9,11	9	17	31,31.2	11,12	10,11	8,11
18	SNU-5540	13	19	11	9	6,7	17	32	10	13	10
19	SNU-5548	16	24,25	11	10,11	7	18	32,32.2	8,12	10	8,11
20	SNU-5584AT	15	23,24	10,11	9	7,9	14	30	10,12	11	8,9
21	SNU-5614	14,15	25	10,11	9,13	6,9	16,18	30,31	10,11	10,11	11
22	SNU-5621	18	17,25	9,12	9,12	9	14,17	30,31	9,11	9,10	8,11
23	SNU-5688	19	18,28	10	11,13	6,8	17,18	29,30	8,11	11,12	8,9
24	SNU-5715D	17	20,23	12	13	7,9.3	14,16	32.2	11	10	8,11
25	SNU-5765	13	20,23	11,12	8,14	9	14	30,32.2	11,13	10	11
26	SNU-5903	14	18,24	11,12	6,11	6,9	15,16	29,30	10,12	12,14	11
27	SNU-6023	14	23,24	11	10,11	9	17,18	31,31.2	9	10,11	8,11
28	SNU-6058S3	15	19,23	12	11	9,9.3	16,17	30	10,12	11	8
29	SNU-6184	13	17	10	12,13	9	14	29,32.2	10,11	11	9
30	SNU-6091	12	19,20	10,11	10,11	7,9	16,19	29,31	11,13	12,13	8,11

Continued

	Cell-Name	D8S1179	D12S931	D19S433	D61043	D22S1045	DYS391	FGA	DYS576	DYS570
16	SNU-5300S3	13,15	15	13,14	12,19	16	8,10	23,24	17	18
17	SNU-5475	11,14	20,21	14.2	12,17	15	9	25,26	18	13
18	SNU-5540	10,15	18	13	18,20	15	-	23	-	-
19	SNU-5548	11	19	12	18	15	-	22,25	11	-
20	SNU-5584AT	14,16	18,21	13,15.2	14,18	16	-	22	-	-
21	SNU-5614	14,16	18,19	14	13,18	15	10	24,25	17	18
22	SNU-5621	10,11	18	13,14	18,19	15	10	23	19	18
23	SNU-5688	11,14	17,18	13,14	18,19	19	-	20	-	-
24	SNU-5715D	9,11	19	14.2,15.2	19	17	11	21,23	18	18
25	SNU-5765	10,13	19	15.2	17,18	17	-	20,24	-	-
26	SNU-5903	13,14	20	13,15	12,13	15	-	19,24	-	-
27	SNU-6023	10,15	19,20	13.2	14,15	16,18	-	23	-	-
28	SNU-6058S3	11,12	20,22	14,15.2	11	16	-	23	-	-
29	SNU-6184	12,15	20	13	18	13,15	-	23	-	-
30	SNU-6091	13,14	21,22	13,14	18	11,17	-	23	-	-

3.2 Classification of CRC cell lines

To identify the molecular characteristics of these 30 CRC cell lines, we classified them with well-known consensus molecular subtypes (CMS) classification. We used *CMScaller* R package which can make a prediction with in vivo and in vitro models [17]. Even though there are some limitations of making a prediction with cell lines, we got 6 CMS1 (20%), 9 CMS2 (30%), 3 CMS3 (10%), 9 CMS4 (30%), 3 NA (10%). 10% of CRC samples failed to make a prediction and the ratio of CMSs was not as similar as other CRC public data (Figure 2). Thus, we considered that there would be a more adequate classification way to elucidate the characteristics of 30 CRC samples.

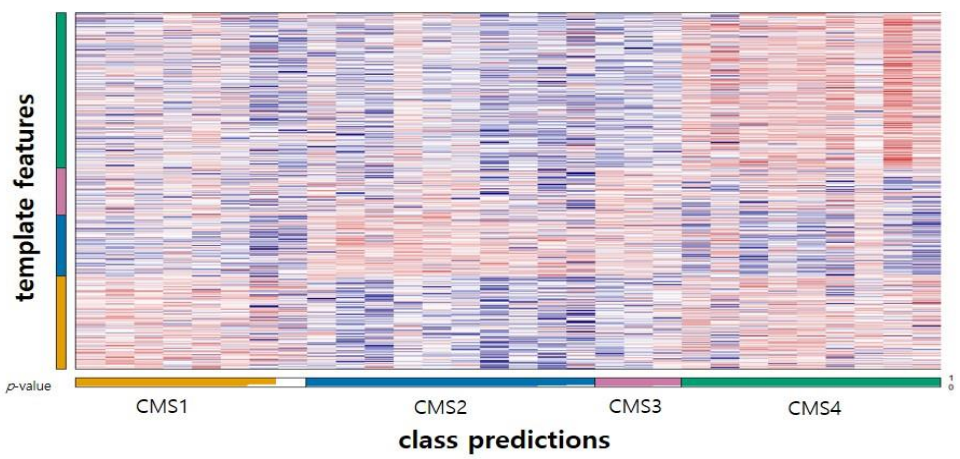


Figure 2. CMS classification of 30 CRC cell lines

Table3. CMS predictions of 30 CRC cell lines

	Cell line name	Prediction
1	SNU.5107	CMS2
2	SNU.5121	CMS2
3	SNU.5124	CMS2
4	SNU.5157	CMS4
5	SNU.5169B	NA
6	SNU.5170B	NA
7	SNU.5171S3	CMS4
8	SNU.5208S1	CMS1
9	SNU.5208S2	CMS1
10	SNU.5208S3	CMS1
11	SNU.5208S4	CMS1
12	SNU.5245S1	CMS1
13	SNU.5245S2	NA
14	SNU.5300S1	CMS4
15	SNU.5300S2	CMS4
16	SNU.5300S3	CMS4
17	SNU.5475	CMS3
18	SNU.5540	CMS2
19	SNU.5548	CMS4
20	SNU.5584AT	CMS4
21	SNU.5614	CMS2
22	SNU.5621	CMS2
23	SNU.5688	CMS2
24	SNU.5715D	CMS2
25	SNU.5765	CMS2
26	SNU.5903	CMS4
27	SNU.6023	CMS4
28	SNU.6058S3	CMS3
29	SNU.6091	CMS3
30	SNU.6184	CMS1

3.3 PCA analysis with 30 CRC cell lines

We used PCA analysis for getting adequately categorizing samples. PCA analysis provided us with significant factors which can separate clusters by reducing dimensionality. We got 3 clusters from the PCA plot as a result (Figure 3). Component scores from each cluster are listed in Table 4. We designated left, right and down clusters as Cluster A, B and C. Custer A and B were divided by PC1 component. Both Cluster A and B have diverse samples with mixed CMSs. This might indicate PCA analysis based separation is a more suitable way to contribute to generating cluster separation that can reflect the characteristic of 30 CRC samples. However, PCA plot did not give us PCA components. Thus, we investigated that which genes made this separation. We used biplot for getting the significant factors of generating clusters (Figure 4). We got AKAP12, AXL as major contributing genes for PC1 separation. But we could not get expressional aberration patterns from the PCA analysis plot. Thus, we investigated that expressional aberration patterns for 30 CRC samples.

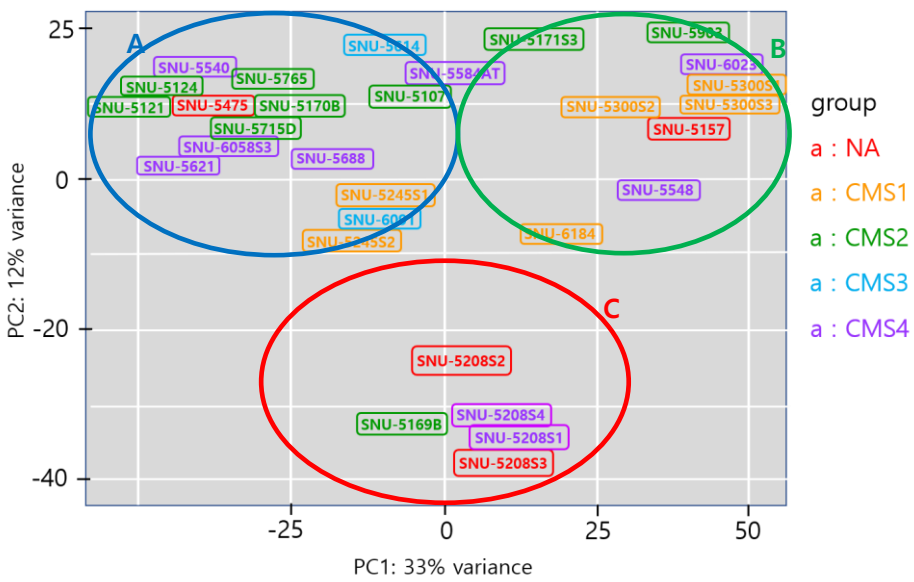


Figure 3. PCA plot of 30 CRC samples

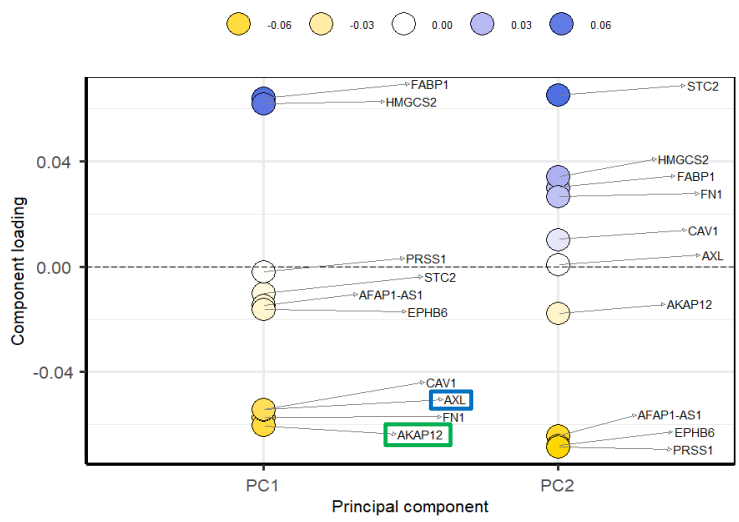


Figure 4. PCA component analysis

Table4. Result of PCA analysis and component scores

Cluster	PC1	PC2	Cell name
A	−39.9996	8.7844	SNU.5121
	−34.9573	10.1385	SNU.5124
	−34.0390	4.4780	SNU.5621
	−33.5684	12.2827	SNU.5540
	−30.7887	9.3112	SNU.5475
	−29.0092	5.0416	SNU.6058S3
	−28.6605	5.8730	SNU.5715D
	−27.2803	6.1582	SNU.5170B
	−25.6726	10.4733	SNU.5765
	−20.0259	1.4070	SNU.5688
	−15.1396	−7.4865	SNU.5245S2
	−12.2616	−5.0591	SNU.6091
	−10.9622	−1.5654	SNU.5245S1
	−6.6379	17.3880	SNU.5614
	−3.5677	6.5835	SNU.5107
B	0.9671	10.5196	SNU.5584AT
	14.2574	16.3871	SNU.5171S3
	15.6705	−5.3913	SNU.6184
	25.6800	−1.3361	SNU.5548
	38.4697	10.1263	SNU.5300S2
	40.0494	17.9850	SNU.5903
	43.8852	8.3152	SNU.5157
	47.5708	10.9417	SNU.5300S3
	48.4702	11.8149	SNU.5300S1
	48.7080	14.2156	SNU.6023
C	2.1171	−26.8370	SNU.5208S2
	7.3231	−40.1383	SNU.5208S3
	10.1564	−38.5650	SNU.5208S1
	13.1336	−37.5436	SNU.5208S4
	−3.8877	−34.3025	SNU.5169B

3.4 DEG analysis of 30 CRC

For getting expressional aberrations of 30 CRC samples, we used DESeq2 R package to normalize RNA-seq read count data. Then we sorted out 40 significant genes and represented them to heatmap (Figure 5). We found that samples in the Cluster A were inactivated in AKAP12. AKAP12 is a family of A-kinase anchor proteins (AKAP). AKAPs are a group of structurally diverse proteins, which have the common function of binding to the regulatory subunit of protein kinase A (PKA) and confining the holoenzyme to discrete locations within the cell. It associates with protein kinase A and C and phosphatase, and it serves as a scaffold protein in signal transduction. Thus, AKAP12 can work as a tumor suppressor. In previous study, re-expression of AKAP12 resulted in inhibition of cell growth and induced apoptosis [18].

On the other hand, we found AXL was overexpressed in Cluster B. AXL is kind of tyrosine protein kinase receptor. Ligand of AXL is growth-arrest-specific gene 6 (GAS6). It can stimulate cell proliferation and survival. AXL signaling is related to PI3K/AKT/mTOR, MEK/ERK, NF-kB, and JAK/STAT [19]. In

previous study, AXL overexpression was related with not only cell proliferation but also migration, cell elongation and epithelial to mesenchymal transition (EMT) [20]. Thus, AXL can work as a proto-oncogene.

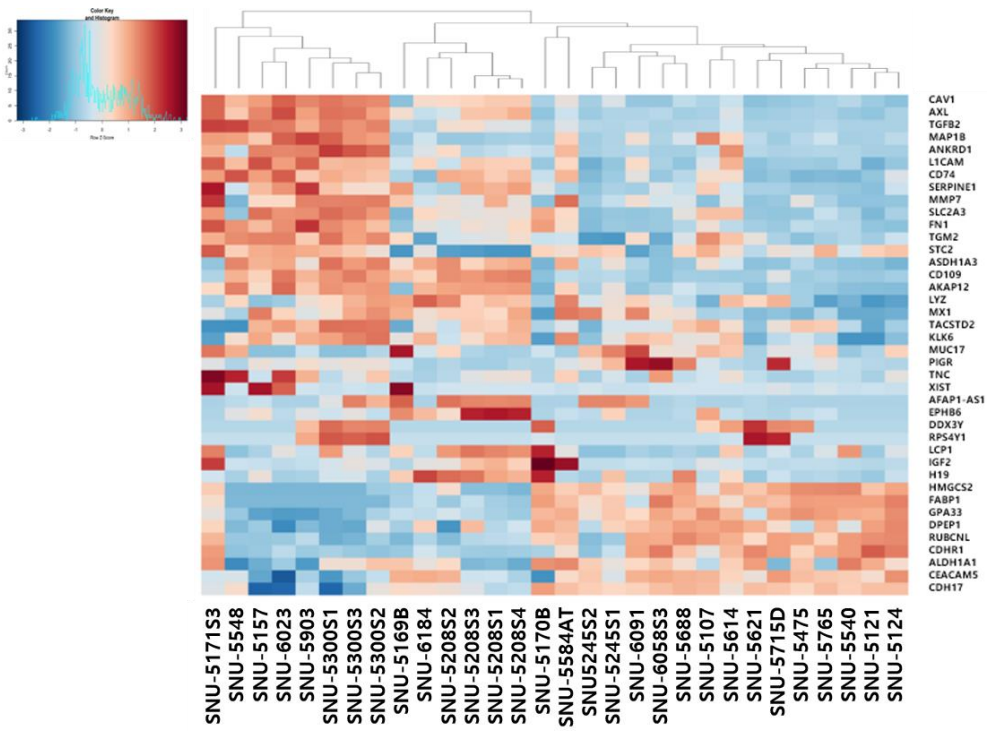


Figure 5. Heatmap of DEG 30 CRC cell lines

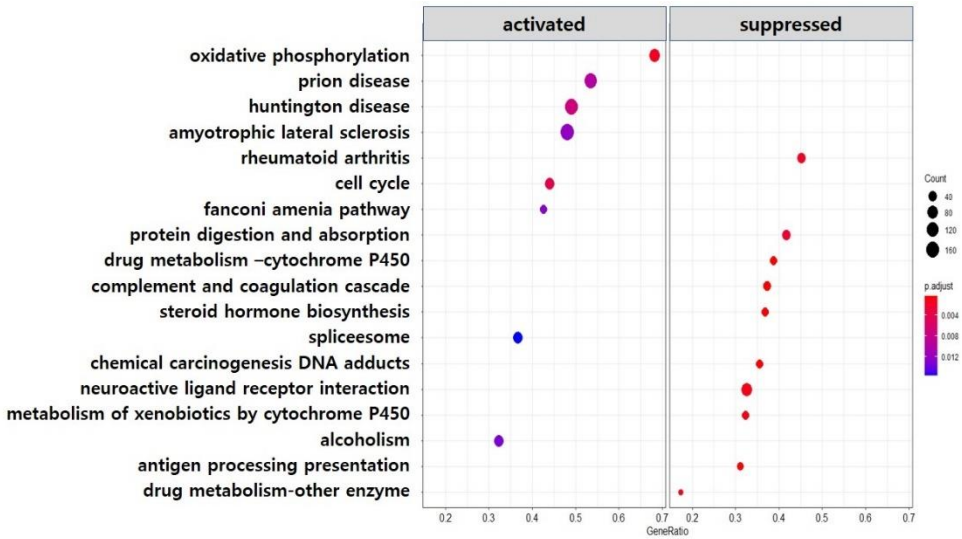
3.5 Pathway analysis of Cluster A and Cluster B

After we identified aberrant expression of AKAP12 and AXL, we investigated more features of each cluster. So we investigated more

about Cluster A for examining which pathways were enriched. We identified cell cycle was activated in Cluster A (Figure 6). We considered inactivation of AKAP12 can activate cell cycle. AKAP12 can work as a tumor suppressor by confining PKC, PKA, and Cyclins. Thus, cell cycle could be activated by suppression of AKAP12.

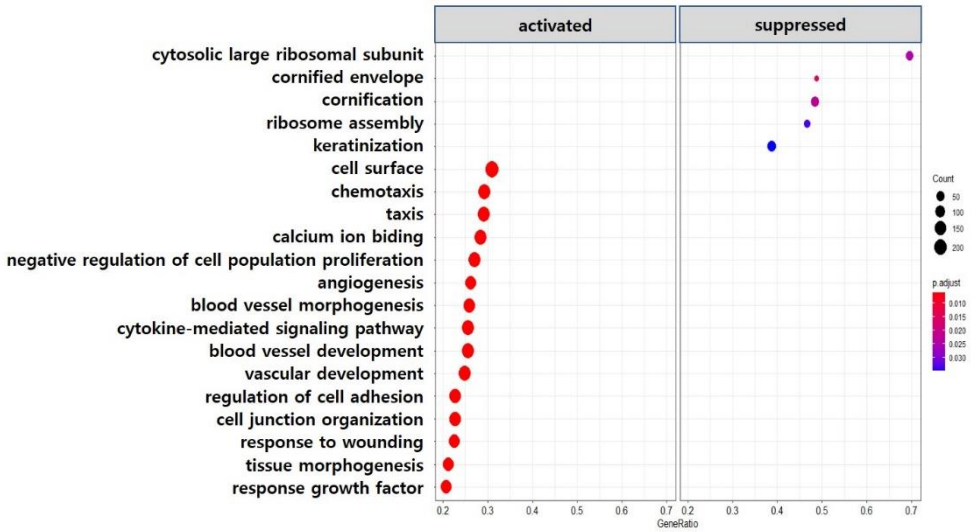
Then, we also investigated Cluster B and found out angiogenesis was activated. Activation of AXL led to stimulating not only cell survival but also cell invasion and metastasis (Figure 7). After the further analysis for Cluster A and B, we compared Cluster A and B. For comparison we used GSEA analysis (Figure 8). When we compared Cluster A with B, we found Myc signaling pathway had positive correlation with Cluster A. From this comparison, we considered suppression of AKAP12 could stimulate cell proliferation.

Then, we also compared Cluster B with A. In this comparison, we found angiogenesis, hypoxia, and EMT signaling pathway had positive correlation with Cluster B (Figure 10). So we assumed that overexpression of AXL led to invasion and metastasis pathway.



Description	setSize	NES	p.adjust	qvalues
Oxidative phosphorylation	129	1.7409	0.0014	0.0314
Cell cycle	125	1.5972	0.0039	0.0575
Fanconi anemia pathway	54	1.5066	0.0126	0.1062
Spliceosome	142	1.3881	0.0156	0.1190
Huntington disease	296	1.3272	0.0071	0.0809
Prion disease	256	1.3213	0.0097	0.0906
Alcoholism	177	1.3160	0.0134	0.1062
Amyotrophic lateral sclerosis	353	1.3107	0.0120	0.1062

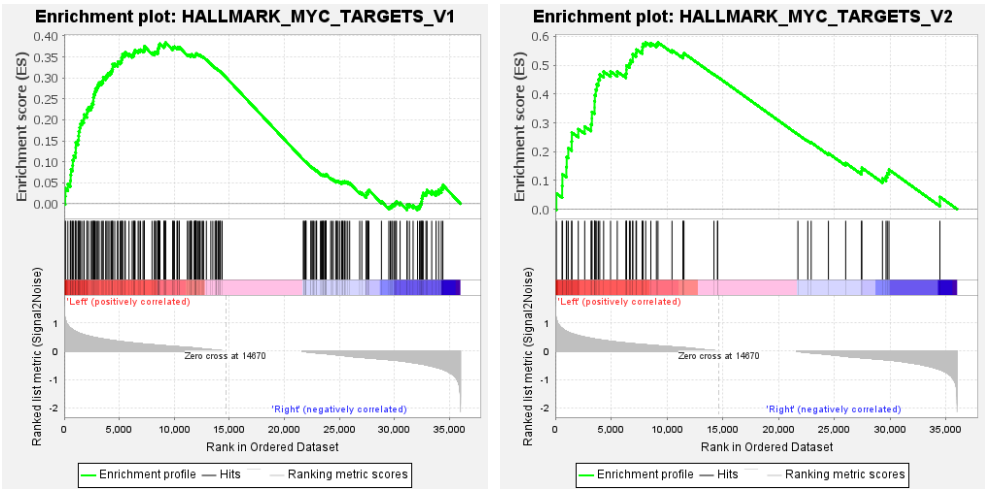
Figure 6. Enrichment plot of Cluster A. p<0.01



NAME	NES	p-val
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	2.752	0
HALLMARK_INTERFERON_GAMMA_RESPONSE	2.585	0
HALLMARK_INTERFERON_ALPHA_RESPONSE	2.354	0
HALLMARK_TNFA_SIGNALING_VIA_NFKB	2.271	0
HALLMARK_INFLAMMATORY_RESPONSE	2.220	0
HALLMARK_ALLOGRAFT_REJECTION	2.149	0
HALLMARK_UV_RESPONSE_DN	2.018	0
HALLMARK_APICAL_JUNCTION	1.979	0
HALLMARK_ANGIOGENESIS	1.979	0
HALLMARK_IL6_JAK_STAT3_SIGNALING	1.973	0
HALLMARK_TGF_BETA_SIGNALING	1.935	0
HALLMARK_MITOTIC_SPINDLE	1.915	0
HALLMARK_APOPTOSIS	1.903	0
HALLMARK_COMPLEMENT	1.860	0

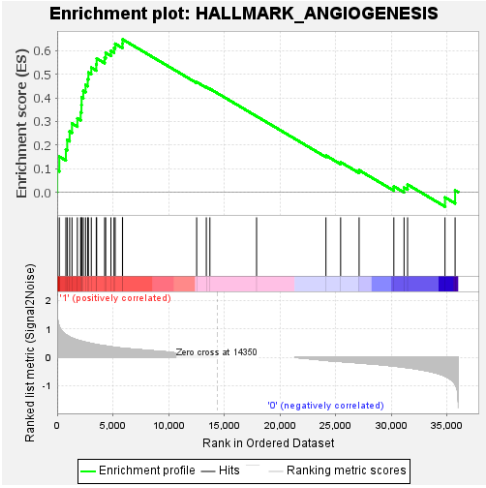
HALLMARK_IL2_STAT5_SIGNALING	1.859	0
HALLMARK_HYPOXIA	1.828	0
HALLMARK_MYOGENESIS	1.803	0
HALLMARK_KRAS_SIGNALING_UP	1.693	0
HALLMARK_COAGULATION	1.662	0

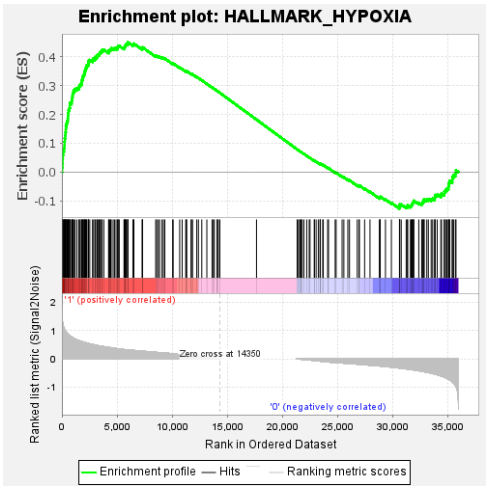
Figure 7. Enrichment plot of Cluster B. p<0.01

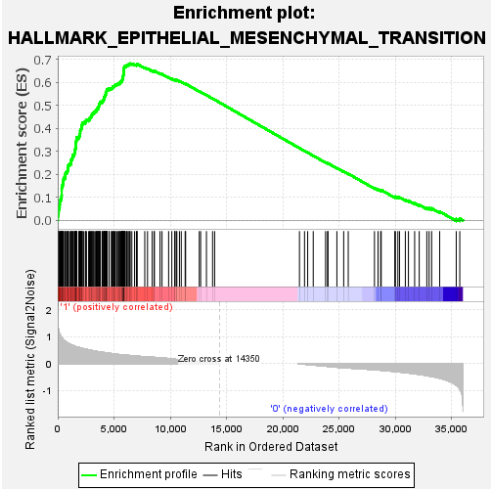


NAME	NES	p-val
HALLMARK_FATTY_ACID_METABOLISM	1.542	0.000
HALLMARK_OXIDATIVE_PHOSPHORYLATION	1.320	0.000
HALLMARK_BILE_ACID_METABOLISM	1.305	0.000
HALLMARK_XENOBIOTIC_METABOLISM	1.304	0.000
HALLMARK_MYC_TARGETS_V2	1.218	0.083
HALLMARK_PANCREAS_BETA_CELLS	1.054	0.322

Figure 8. GSEA plot comparison Cluster A with B







NAME	NES	p-val
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	2.7519	0.0000
HALLMARK_INTERFERON_GAMMA_RESPONSE	2.5855	0.0000
HALLMARK_INTERFERON_ALPHA_RESPONSE	2.3541	0.0000
HALLMARK_TNFA_SIGNALING_VIA_NFKB	2.2708	0.0000
HALLMARK_INFLAMMATORY_RESPONSE	2.2205	0.0000
HALLMARK_ALLOGRAFT_REJECTION	2.1494	0.0000
HALLMARK_UV_RESPONSE_DN	2.0183	0.0000
HALLMARK_APICAL_JUNCTION	1.9789	0.0000

HALLMARK_ANGIOGENESIS	1.9787	0.0000
HALLMARK_IL6_JAK_STAT3_SIGNALING	1.9734	0.0000
HALLMARK_TGF_BETA_SIGNALING	1.9354	0.0000
HALLMARK_MITOTIC_SPINDLE	1.9145	0.0000
HALLMARK_APOPTOSIS	1.9027	0.0000
HALLMARK_COMPLEMENT	1.8597	0.0000
HALLMARK_IL2_STAT5_SIGNALING	1.8585	0.0000
HALLMARK_HYPOXIA	1.8281	0.0000

Figure 9. GSEA plot comparison Cluster B with A

3.6 Enrichment pathway analysis of Cluster C and fusion gene

In Cluster C, the actual number of patients is just two and PCA analysis landscape revealed that comparison with other Clusters is not suitable for Cluster C. Thus, we analyzed Cluster C separately. In the previous PCA analysis, we identified Cluster C was separated by PC2. We found AFAP1-AS1 was the significant feature in Cluster C (Figure 4). AFAP1-AS1 is actin filament associated protein 1 antisense RNA1. In previous studies showed that AFAP1-AS1 was dysregulated in diverse cancers [21-23]. *Feng Wang et al.,* revealed that overexpression of AFAP1-AS1 in CRC had correlated with tumor malignant progression and poor prognosis [24]. Enrichment analysis of Cluster C revealed that PI3K/AKT and RAS pathway was

activated (Figure 10).

After enrichment pathway analysis of each Cluster, we analyzed fusion gene for further analysis for 30 CRC samples. Among 30 CRC samples, we got PTPRK-RSPO3 fusion gene in Cluster C (Table 5). In previous study verified PTPRK-RSPO3 fusion gene could be oncogenic in CRC [25]. So this fusion gene can be the specific feature of the Cluster C.

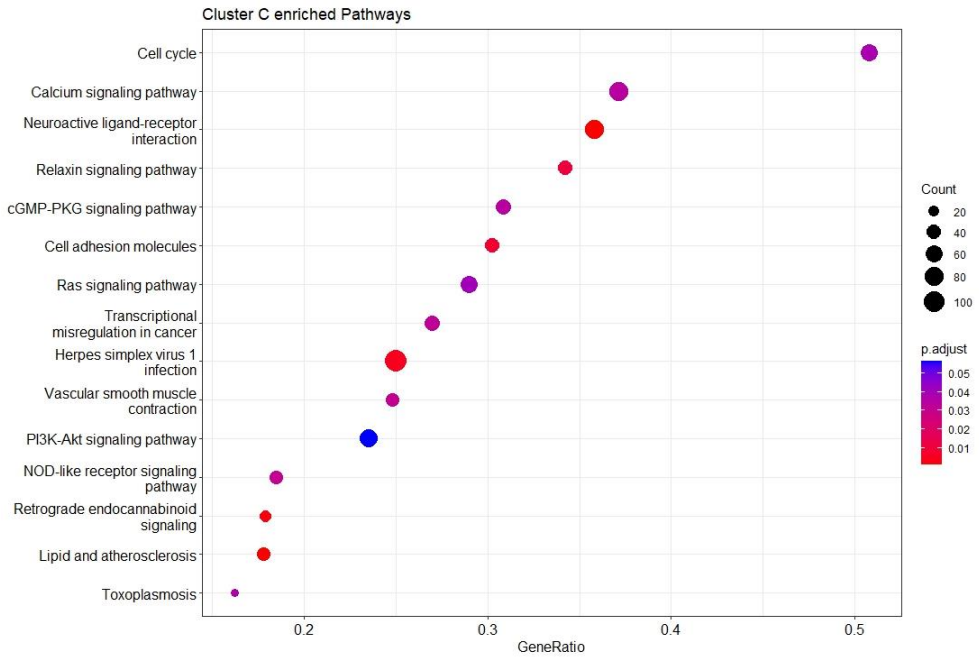


Figure 10. Pathway enrichment analysis of Cluster C. p<0.1

Table 5. Fusion genes of 30 CRC

Cell name	gene1	gene2	strand1 (gene/fusion)	strand2 (gene/fusion)	breakpoint1	breakpoint2
SNU-5107	ASL	CRCP	+/+	+/+	7:65557650	7:65592691
SNU-5107	VPS13B	POLR2K	+/+	+/+	8:100205285	8:101163575
SNU-5169B	PTPRK	RSPO3	-/-	+/+	6:128385903	6:127469793
SNU-5170B	RPL21P40(15780),GPD1L(60020)	OSBPL10	./-	-/-	3:32087983	3:31921322
SNU-5208S1	PTPRK	RSPO3	-/-	+/+	6:128841404	6:127469793
SNU-5208S2	PTPRK	RSPO3	-/-	+/+	6:128841404	6:127469793
SNU-5208S2	PTPRK	RSPO3	-/-	+/+	6:128841404	6:127471571
SNU-5208S2	PTPRK	RSPO3	-/-	+/+	6:128841404	6:127516968
SNU-5208S2	TCF7L2	VTI1A	+/+	+/+	10:114799885	10:114575049
SNU-5208S2	ZNF841	ZNF432	-/-	-/-	19:52567747	19:52550313
SNU-5208S2	ASL	CRCP	+/+	+/+	7:65557650	7:65592691
SNU-5208S3	PTPRK	RSPO3	-/-	+/+	6:128841404	6:127469793
SNU-5208S3	PTPRK	RSPO3	-/-	+/+	6:128841404	6:127471571
SNU-5208S3	PTPRK	RSPO3	-/-	+/+	6:128841404	6:127516968
SNU-5208S3	ZNF841	ZNF432	-/-	-/-	19:52567747	19:52550313
SNU-5208S4	PTPRK	RSPO3	-/-	+/+	6:128841404	6:127469793
SNU-5245S1	ASL	CRCP	+/+	+/+	7:65557650	7:65592691
SNU-5245S2	ZNF841	ZNF432	-/-	-/-	19:52567747	19:52550313

3.7 Mutational landscape of 30 colorectal cancer cell lines

After transcriptomic analysis, we investigated genomic mutations because according to Vogelgram, genomic mutation can lead to the initiation and progression of CRC. We selected 403 cancer-related genes by using PCAWG (Pancancer Analysis of whole genomes). Among those gene, we sorted cancer related pathogenic genes by using Clinvar. Then we compared 30 established cell lines with 68 CRC datasets from CCLE datasets. 30 CRC cell lines and CCLE data sets have high rates of 3 typical colorectal driver mutations such as KRAS, APC, TP53 (Figure 11), while 30 CRC cell lines also have unique mutations in NQO1. NQO1 is related to p53 stabilizing [26]. So this gene can be a novel way of p53 related target therapy. The key driver mutation genes were listed in Table 6.

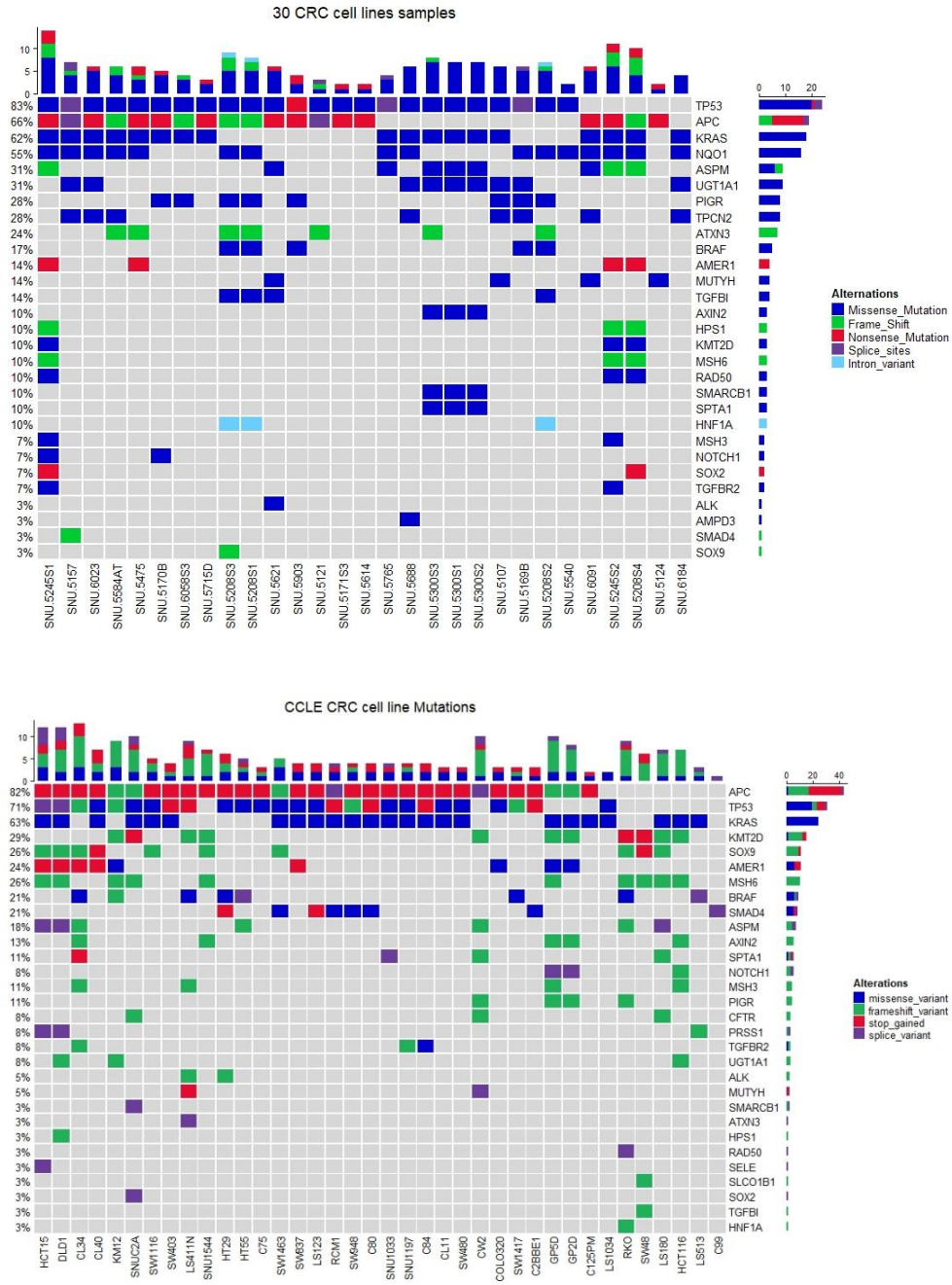


Figure 11. Mutations of 30 established colorectal cancer cell lines

Table 6. List of key mutations of 30 CRC cell lines

	SNU-5107	SNU-5121	SNU-5124	SNU-5157	SNU-5169B	SNU-5170B
KRAS	c.38G>A p.Gly13Asp Pathogenic			c.35G>T p.Gly12Val Pathogenic		c.34G>A p.Gly12Ser Pathogenic
APC		c.1312+1G>A Pathogenic	c.847C>T p.Arg283* Pathogenic	c.1312+2T>C Pathogenic		
TP53	c.857A>G p.Glu286Gly Uncertain significance	c.215C>G p.Pro72Arg drug response		c.375G>A p.Thr125Thr Pathogenic	c.673-1G>C Pathogenic	c.747G>T p.Arg249Ser Conflicting interpretations of pathogenicity
ATXN3		c.916_917insC p.Gly306fs Uncertain significance				

Continued

	SNU-5171S3	SNU-5208S1	SNU-5208S2	SNU-5208S3	SNU-5208S4	SNU-5245S1
KRAS					c.38G>A p.Gly13Asp Pathogenic	c.38G>A p.Gly13Asp Pathogenic
APC	c.3268C>T p.Gln1090* Pathogenic	c.2483delC p.Thr828fs Pathogenic	c.2483delC p.Thr828fs Pathogenic		c.4348C>T p.Arg1450* Pathogenic	c.4348C>T p.Arg1450* Pathogenic
TP53	c.844C>T p.Arg282Trp Pathogenic	c.743G>A p.Arg248Gln Pathogenic	c.743G>A p.Arg248Gln Pathogenic	c.743G>A p.Arg248Gln Pathogenic		
TGFBR2						c.1019C>T p.Thr340Met Conflicting interpretations of pathogenicity
ATXN3		c.916_917insC p.Gly306fs Uncertain significance	c.916_917insC p.Gly306fs Uncertain significance	c.916_917insC p.Gly306fs Uncertain significance		

Continued

	SNU-5245S2	SNU-5300S1	SNU-5300S2
KRAS	c.38G>A	c.436G>A	c.436G>A
	p.Gly13Asp	p.Ala146Thr	p.Ala146Thr
	Pathogenic	Conflicting interpretations of pathogenicity	Conflicting interpretations of pathogenicity
APC	c.4348C>T		
	p.Arg1450*		
	Pathogenic		
TP53		c.746G>T	c.746G>T
		p.Arg249Met	p.Arg249Met
		Uncertain significance	Uncertain significance
TGFBR2	c.1019C>T		
	p.Thr340Met		
	Conflicting interpretations of pathogenicity		

Continued

	SNU-5300S3	SNU-5475	SNU-5540	SNU-5548	SNU-5584AT	SNU-5614
KRAS	c.436G>A	c.35G>T		c.35G>A	c.35G>A	
	p.Ala146Thr	p.Gly12Val		p.Gly12Asp	p.Gly12Asp	
	Conflicting interpretations of pathogenicity	Pathogenic		Pathogenic	Pathogenic	
APC		c.2077A>T		c.4033G>T	c.4393_4394delAG	c.1660C>T
		p.Lys693*		p.Glu1345*	p.Ser1465fs	p.Arg554*
		Pathogenic		Pathogenic	Pathogenic	Pathogenic
		c.4348C>T				
		p.Arg1450*				
		Pathogenic				
TP53	c.746G>T	c.797G>A	c.818G>A	c.844C>T	c.794T>C	c.844C>T
	p.Arg249Met	p.Gly266Glu	p.Arg273His	p.Arg282Trp	p.Leu265Pro	p.Arg282Trp
	Uncertain significance	Pathogenic /Likely pathogenic	Pathogenic	Pathogenic /Likely pathogenic	Pathogenic /Likely pathogenic	Pathogenic /Likely pathogenic
ATXN3	c.916_917insC	c.916_917insC			c.916_917insC	
	p.Gly306fs	p.Gly306fs			p.Gly306fs	
	Uncertain significance	Uncertain significance			Uncertain significance	

Continued

	SNU-5621	SNU-5688	SNU-5715D	SNU-5765	SNU-5903	SNU-6023
KRAS		c.38G>A p.Gly13Asp Pathogenic	c.35G>A p.Gly12Asp Pathogenic	c.436G>A p.Ala146Thr Conflicting interpretations of pathogenicity		c.35G>A p.Gly12Asp Pathogenic
APC	c.4033G>T p.Glu1345* Pathogenic		c.1268G>A p.Trp423* Pathogenic c.4348C>T p.Arg1450* Pathogenic		c.2626C>T p.Arg876* Pathogenic	
TP53		c.475G>C p.Ala159Pro Uncertain significance	c.337T>G p.Phe113Val Uncertain significance	c.672+1G>A Pathogenic	c.637C>T p.Arg213* Pathogenic	c.743G>A p.Arg248Gln Pathogenic

Continued

	SNU-6058S3	SNU-6091	SNU-6184
KRAS	c.34G>T	c.35G>A	c.35G>T
	p.Gly12Cys	p.Gly12Asp	p.Gly12Val
	Conflicting interpretations of pathogenicity	Pathogenic	Pathogenic
APC	c.3139G>T	c.4348C>T	
	p.Glu1047*	p.Arg1450*	
	Likely pathogenic	Pathogenic	
TP53	c.844C>T		
	p.Arg282Trp		
	Pathogenic /Likely pathogenic		

3.8 Drug response of CRC cell lines

At the end of this study, we verified relation between cell lines and drug responses (Figure 12). Most of the 30 CRC cell lines were related with Akt signaling through AXL, AKAP12 and AFAP1-AS1.

In previous study, treatment of HDAC inhibitor induced AKAP12 expression [27]. For confirming that Belinostat (HDAC inhibitor) worked well in Cluster A, we selected 5 samples from Cluster A (red) and Cluster B (blue). Cell viability revealed that Belinostat worked well in Cluster A (Figure 13). And we found Cluster C had PTPRK-RSPOS fusion gene. We assumed this fusion gene activated Wnt signaling. Thus, ICG-001 worked well with Wnt overexpression cell lines.

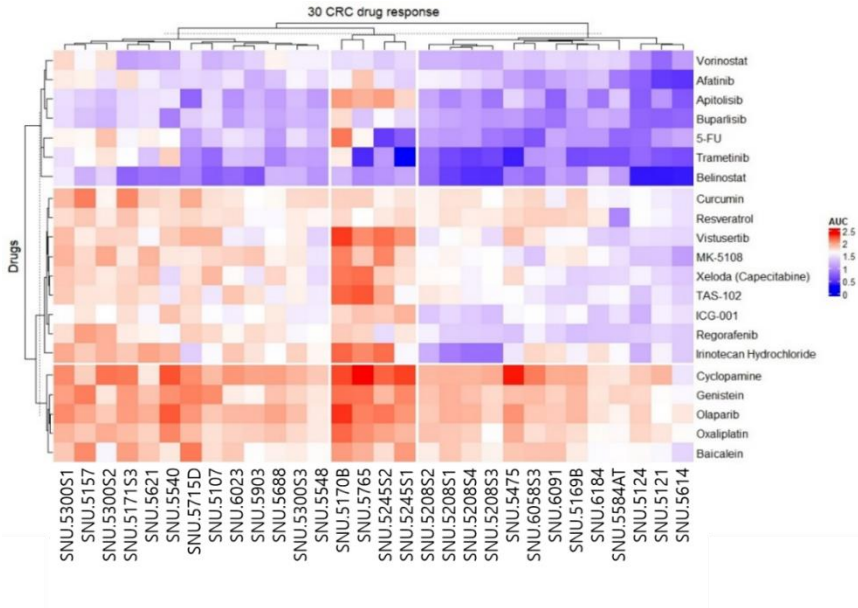


Figure 12. Heatmap of drug responses

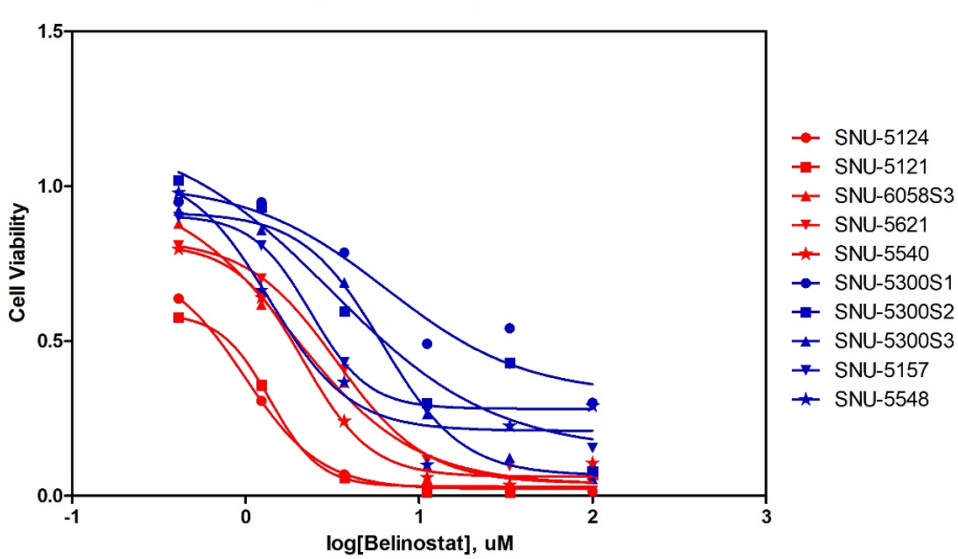


Figure 13. Belinostat response in Cluster A and B

Discussion

1.1milion patients had colorectal cancer in 2020 worldwide, and it will increase to 1.9 in 2040 [28]. Therefore, it is important to diagnose the colon cancer and treat with proper therapeutics. However, there are many factors to induce carcinogenesis. For finding carcinogenic factors, we analyzed PDCs with WES and RNA-seq.

For characterizing the 30 CRC cell lines, we should classify the samples at first. For classifying the 30 CRC cell lines, we used CMS classification type. There are 4 types of CMSs. CMS1 (microsatellite instability immune, 14%), hypermutated, microsatellite unstable and strong immune activation; CMS2 (canonical, 37%), epithelial, marked WNT and MYC signaling activation; CMS3 (metabolic, 13%), epithelial and evident metabolic dysregulation; and CMS4 (mesenchymal, 23%), prominent transforming growth factor-β activation, stromal invasion and angiogenesis [29]. While previous CMS prediction R package *CMS classifier* has limitation of CMS4 prediction with cell lines due to lack of microenvironment, the improved CMS prediction R package *CMScaller* can predict CMS more

precisely. *Anita Sveen et al.*, already suggested a preclinical cell line models by using *CMScaller* [30]. However, the accuracy of *CMScaller* also depends on sample sizes. The small number of datasets inevitably introduce prediction uncertainty. For this reason, *CMScaller* recommended using at least 40 cell lines to predict CMSs. Our study got 6 CMS1 (20%), 9 CMS2 (30%), 3 CMS3 (10%), 9 CMS 4 (30%), 3 NA (10%). This result was a little different than what we expected because of high proportion of CMS4 and 10% of NA. Thus, we used PCA analysis instead. PCA analysis is also a good way to figure out significant factors for grouping the samples by reducing the dimensionalities. As a result, we got AKAP12, AXL, AFAP1-AS1 to separate clusters. We assumed that this PCA based separation could reflect the characterization of 30 CRC samples rather than CMS at least this cohort.

Even though we could not use CMS classification due to restriction of the number of CRC cell lines, we found out dysregulated signals such as AXL, AKAP12, AFAP1-AS1 by PCA analysis and fusion gene PTPRK-RSPO3 separated 30 CRC samples. We investigated the role of 3 genes related with cancers. There has been several

studies that revealed relation between AKAP12 and CRC [31-33]. There was research that revealed reverse correlation between AKAP12 and oncomiR-183-5p [34]. miRNA-183-5p can work as oncomiRNA. It is associated with carcinogenesis, metastasis via regulation of tumor suppressor genes. The level of miRNA-183 was high in CRC [35]. It showed the inactivation of AKAP12 by overexpression of miRNA-183-5p led to increase in cell invasion. From these previous studies, we assumed that suppression of AKAP12 could lead to development of CRC and miRNA might be a way to control the carcinogenic factor of CRC. Thus, for further study, it is necessary to epigenetic analysis near AKAP12 promoter region for confirming that AKAP12 was one of the main carcinogenic factors of CRC.

AXL was a significant gene which could separate Cluster B. AXL is a kind of receptor tyrosine kinase. GAS6 is ligand of AXL so it can stimulate cell cycle, proliferation. AXL is overexpressed in various cancers related with cell proliferation, metastasis, invasion[36-38]. In previous study, AXL was also high in CRC related with AKT and ERK pathway [39]. From comparison Cluster B with A, we identified

angiogenesis signaling pathway was activated. Thus, we can confirm the angiogenesis via wound healing assay for further study.

The last feature of Cluster C was AFAP1-AS1. AFAP1-AS1 produces a long non-coding RNA that is involved in etiology of CRC. In previous study, Xu Han et al., revealed that suppression of AFAP1-AS1 inhibited the growth of CRC and reduced EMT signaling [40]. From this evidence we assumed that overexpression of AFAP1-AS1 might related with CRC carcinogenesis.

After analyzing the enrichment pathway analysis of Cluster A,B and C, we made further characterization by fusion gene detection. PTPRK-RSPO3 was presumable oncogenic fusion gene [41]. This fusion gene led to overexpression of functional R-spondin proteins.

In this study, we found out 3 significant gene expression and fusion gene via WES and RNA-seq. Even though genomic and transcriptomic analysis is a reliable way to understand the characteristic features of cancer but cancer is so complicated disease. Therefore, other aspects of analysis such as miRNA and epigenomic analysis will be needed for understanding comprehensive features of cancers. Thus, in recent cancer research, many researchers have

researched miRNA and epigenic analysis [42, 43]. For following this research tendency, we should do other analysis for further study.

Reference

1. Jung, K.W., et al., *Prediction of Cancer Incidence and Mortality in Korea, 2021.* Cancer Res Treat, 2021. **53**(2): p. 316-322.
2. Garrido-Ramos, M.A., *Satellite DNA: An Evolving Topic.* Genes (Basel), 2017. **8**(9).
3. Lower, S.S., et al., *Satellite DNA evolution: old ideas, new approaches.* Curr Opin Genet Dev, 2018. **49**: p. 70-78.
4. Toyota, M., et al., *CpG island methylator phenotype in colorectal cancer.* Proc Natl Acad Sci U S A, 1999. **96**(15): p. 8681-6.
5. Jover, R., et al., *5-Fluorouracil adjuvant chemotherapy does not increase survival in patients with CpG island methylator phenotype colorectal cancer.* Gastroenterology, 2011. **140**(4): p. 1174-81.
6. Kim, Y.S. and G. Deng, *Epigenetic changes (aberrant DNA methylation) in colorectal neoplasia.* Gut Liver, 2007. **1**(1): p. 1-11.
7. Herman, J.G. and S.B. Baylin, *Gene silencing in cancer in association with promoter hypermethylation.* N Engl J Med, 2003. **349**(21): p. 2042-54.
8. Nagasaka, T., et al., *Mutations in both KRAS and BRAF may contribute to the methylator phenotype in colon cancer.* Gastroenterology, 2008. **134**(7): p. 1950-1960.
9. Lengauer, C., K.W. Kinzler, and B. Vogelstein, *Genetic instabilities in human cancers.* Nature, 1998. **396**(6712): p. 643-9.
10. Markowitz, S.D. and M.M. Bertagnolli, *Molecular origins of cancer: Molecular basis of colorectal cancer.* N Engl J Med, 2009. **361**(25): p. 2449-60.
11. Tsang, A.H., et al., *Current and future molecular diagnostics in colorectal cancer and colorectal adenoma.* World J Gastroenterol, 2014. **20**(14): p. 3847-57.
12. Grady, W.M. and C.C. Pritchard, *Molecular alterations and biomarkers in colorectal cancer.* Toxicol Pathol, 2014. **42**(1): p. 124-39.

13. Fransen, K., et al., *Mutation analysis of the BRAF, ARAF and RAF-1 genes in human colorectal adenocarcinomas.* Carcinogenesis, 2004. **25**(4): p. 527-33.
14. Pino, M.S. and D.C. Chung, *The chromosomal instability pathway in colon cancer.* Gastroenterology, 2010. **138**(6): p. 2059-72.
15. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis.* Cell, 1990. **61**(5): p. 759-67.
16. Rao, C.V. and H.Y. Yamada, *Genomic instability and colon carcinogenesis: from the perspective of genes.* Front Oncol, 2013. **3**: p. 130.
17. Eide, P.W., et al., *CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models.* Scientific Reports, 2017. **7**.
18. Liu, W., et al., *Re-expression of AKAP12 inhibits progression and metastasis potential of colorectal carcinoma in vivo and in vitro.* PLoS One, 2011. **6**(8): p. e24015.
19. Gay, C.M., K. Balaji, and L.A. Byers, *Giving AXL the axe: targeting AXL in human malignancy.* Br J Cancer, 2017. **116**(4): p. 415-423.
20. Uribe, D.J., et al., *The receptor tyrosine kinase AXL promotes migration and invasion in colorectal cancer.* PLoS One, 2017. **12**(7): p. e0179979.
21. Shi, D.Y., et al., *LncRNA AFAP1-AS1 promotes tumorigenesis and epithelial-mesenchymal transition of osteosarcoma through RhoC/ROCK1/p38MAPK/Twist1 signaling pathway (vol 38, 375, 2019).* Journal of Experimental & Clinical Cancer Research, 2020. **39**(1).
22. Feng, Y., et al., *Increased lncRNA AFAP1-AS1 expression predicts poor prognosis and promotes malignant phenotypes in gastric cancer.* Eur Rev Med Pharmacol Sci, 2017. **21**(17): p. 3842-3849.
23. Cai, B., et al., *LncRNA AFAP1-AS1 Knockdown Represses Cell Proliferation, Migration, and Induced Apoptosis in Breast Cancer by Downregulating SEPT2 Via Sponging miR-497-5p.* Cancer Biother Radiopharm, 2020.

24. Wang, F., et al., *Overexpression of lncRNA AFAP1-AS1 correlates with poor prognosis and promotes tumorigenesis in colorectal cancer.* Biomed Pharmacother, 2016. **81**: p. 152-159.
25. Choi, Y., et al., *Integrative analysis of oncogenic fusion genes and their functional impact in colorectal cancer.* Br J Cancer, 2018. **119**(2): p. 230-240.
26. Liu, F., et al., *An NQO1-initiated and p53-independent apoptotic pathway determines the anti-tumor effect of tanshinone IIA against non-small cell lung cancer.* PLoS One, 2012. **7**(7): p. e42138.
27. He, P., et al., *Upregulation of AKAP12 with HDAC3 depletion suppresses the progression and migration of colorectal cancer.* International Journal of Oncology, 2018. **52**(4): p. 1305-1316.
28. Xi, Y. and P. Xu, *Global colorectal cancer burden in 2020 and projections to 2040.* Transl Oncol, 2021. **14**(10): p. 101174.
29. Guinney, J., et al., *The consensus molecular subtypes of colorectal cancer.* Nat Med, 2015. **21**(11): p. 1350-6.
30. Sveen, A., et al., *Colorectal Cancer Consensus Molecular Subtypes Translated to Preclinical Models Uncover Potentially Targetable Cancer Cell Dependencies.* Clin Cancer Res, 2018. **24**(4): p. 794-806.
31. Suren, D., et al., *Lack of relation of AKAP12 with p53 and Bcl-2 in colorectal carcinoma.* Asian Pac J Cancer Prev, 2014. **15**(8): p. 3415-8.
32. Xu, G.R., et al., *A 15-gene signature for prediction of colon cancer recurrence and prognosis based on SVM.* Gene, 2017. **604**: p. 33-40.
33. Yildirim, M., et al., *AKAP12/Gravin gene expression in colorectal cancer: clinical importance and review of the literature.* Journal of Buon, 2013. **18**(3): p. 635-640.
34. Hu, T., et al., *AKAP12 Endogenous Transcripts Suppress The Proliferation, Migration And Invasion Of Colorectal Cancer Cells By Directly Targeting oncomiR-183-5p.* Onco Targets Ther, 2019. **12**: p. 8301-8310.

35. Sarver, A.L., et al., *Human colon cancer profiles show differential microRNA expression depending on mismatch repair status and are characteristic of undifferentiated proliferative states.* Bmc Cancer, 2009. **9**.
36. Tai, K.Y., et al., *Axl promotes cell invasion by inducing MMP-9 activity through activation of NF-kappaB and Brg-1.* Oncogene, 2008. **27**(29): p. 4044-55.
37. May, C.D., et al., *AXL is a potential therapeutic target in dedifferentiated and pleomorphic liposarcomas.* Bmc Cancer, 2015. **15**.
38. Rankin, E.B., et al., *AXL Is an Essential Factor and Therapeutic Target for Metastatic Ovarian Cancer.* Cancer Research, 2010. **70**(19): p. 7570-7579.
39. Martinelli, E., et al., *AXL is an oncotarget in human colorectal cancer.* Oncotarget, 2015. **6**(27): p. 23281-23296.
40. Han, X., et al., *Long non-coding RNA AFAP1-AS1 facilitates tumor growth and promotes metastasis in colorectal cancer.* Biol Res, 2016. **49**(1): p. 36.
41. Hashimoto, T., et al., *Clinicopathological and molecular characteristics of RSPO fusion-positive colorectal cancer.* British Journal of Cancer, 2022.
42. Cheng, Y., et al., *Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials.* Signal Transduct Target Ther, 2019. **4**: p. 62.
43. Jung, G., et al., *Epigenetics of colorectal cancer: biomarker and therapeutic potential.* Nat Rev Gastroenterol Hepatol, 2020. **17**(2): p. 111-130.

국문 초록

대장암은 세계적으로 가장 흔한 암 중 하나이다. 한국에서 남녀 모두에게서 발병률이 가파르게 증가해왔다. 대장암이 발병되는 원인에는 유전자 복제 결함, 유전자 돌연변이, 비정상적 유전자 발현 등이 있다.

이전 연구들에서 차세대 염기서열분석(NGS)을 이용하여 비정상적 유전자 발현 뿐 아니라 유전자 변이를 발견해왔다. 그럼에도 불구하고, 잘 특징 지어진 전 임상 모델의 부족으로 인해 이런 분자 특성들을 약물 반응성과 함께 통합적으로 자세히 연구하지 못했다.

이에 이 연구에서 30개의 환자유래 대장암 세포주 (PDC)를 수립했다. 이 세포주들은 광범위한 유전자, 전사체 서열 분석과 21개의 임상 관련된 약물들을 대량으로 스크리닝 하기에 용이하다. 이 연구 결과 대장암 세포주들의 전사체의 경향 분석으로 30개의 세포주들을 3가지 하위 집단으로 분석 할 수 있다. AXL, AKAP12, AFAP1-AS1이 PI3K/AKT 경로를 통해 주로 하위 집단을 구분할 수 있다. 이 연구는 분자적 특성이 특정 약물결과와 함께 연관성을 설명하고 환자유래 세포주를 이용하는 것이 환자 특이적 항암제 민감성의 유의성을 입증할 수 있다.

주요어: 대장암, 항암제, 코호트 분석

학번: 2019-22429