의학석사 학위논문

Integrated Multi-Omics Analysis of Colorectal Liver Metastasis Using Colorectal Organoids Identifies Putative Molecular Targets

대장암 오가노이드의 다중 오믹스 분석을 통한 대장암 간 전이의 잠재적 표적 발굴

2023년 8월

서울대학교 대학원

의과학과 의과학전공

김 채 영

Integrated Multi-Omics Analysis of Colorectal Liver Metastasis Using Colorectal Organoids Identifies Putative Molecular Targets

지도 교수 구 자 록

이 논문을 의학석사 학위논문으로 제출함
2023년 4월

서울대학교 대학원
의과학과 의과학전공
김 채 영

김채영의 의학석사 학위논문을 인준함
2023년 7월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

Abstract

김 채 영 (ChaeYoung Kim)
의과학과 의과학전공 (Biomedical Sciences Major)
The Graduate School
Seoul National University

Metastasis remains an obstacle to the treatment of colorectal cancer (CRC). Liver is one of the most common sites of colorectal cancer metastases (CLM). However, the number of experimental models that retain the biological and molecular characteristics of CLM is insufficient to pinpoint therapeutic targets. Exosomal miRNA, one of the many causes of liver metastasis (LM), promotes distant metastasis by altering biological processes such as epithelial–mesenchymal transition (EMT), immunosuppression, and tumor microenvironment re–composition.

In this study, five pairs of organoids established from CLM cell lines and tissues were subjected to genetic, translational and pharmacological analysis. Then, multi–omics layers were integrated with specific molecular markers centered on the transcriptomic and miRNA profiles. Decision tree and LARS model indicated that regulation of KRAS signaling was highly associated with the

responses of anti-tumor drugs.

My approach not only identified molecular biomarkers that were specifically regulated in LM organoids but also connected those markers to a certain therapeutic regimen.

Keywords: Colorectal liver metastasis, Organoid, Exosomal miRNA, Multi-omics, Drug screening, Prediction model

Student Number: 2021-28472

Table of Contents

List of Tables

List of Figures

Figure 20. Molecular targets of Regorafenib and Apitolisib.

Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed malignancy worldwide, and the third leading cause of cancer death in Korea [1−3]. More than half of CRC patients develop colorectal liver metastasis (CLM) largely due to the adjacent anatomical relation [4, 5]. The prognosis of metastatic CRC patients remains dismal, as only 11.7% survive within the 5−year follow up period. The benefits of hepatectomy would be marginal because more than 90% of CLM patients are diagnosed as unresectable due to tumor location, size, or inadequate residual liver [6−8]. Accordingly, identifying suitable chemotherapeutic targets would improve the prognosis of CLM patients. Nevertheless, current research in CLM has been focused on only the genetic and transcriptomic layers, limiting understanding of tumor dissemination in association with pharmacological comprehension [9−11].

Regardless of resectability, treatment of CLM mainly depends on adjuvant chemotherapy. Combination therapies based on cytotoxic agents such as CAPOX, FOLFIRI, and FOLFOX with certain targeted drugs, including Bevacizumab and Cetuximab, are considered first−line treatments [12]. Nevertheless, only a small portion of CLM

patients benefit from first-line chemotherapies, highlighting the importance of identifying CLM-specific drugs in accordance with multi-omics analysis. In this view point, establishing *in-vitro* CLM models contributes to comprehending the sequential cascades of tumor dissemination to the liver as well as the overall drug response to certain drugs. Especially, the unique ability of patient-derived organoids (PDOs) to capture both clinicopathological and molecular features of the original tumor is highly valuable [13].

Recent studies report that tumor-derived exosomal microRNAs (miRNAs) induce metastases by regulating multiple oncogenic pathways such as epithelial-mesenchymal transition (EMT), angiogenesis, and immunosuppression in CRC [14]. Exosomes are small (30-90 nm) cell-derived vesicles that promote intercellular communication by delivering bioactive components such as DNA, mRNA, protein, and various kinds of non-coding RNA between cells [15]. Exosomes can be detected in most body fluids, therefore, they are considered a valuable non-invasive liquid biomarker [16, 17]. miRNAs are short (18-25 nt) non-coding RNAs involved in the post-transcriptional regulation of gene expression [18]. They bind to the 3'-untranslated region (3' UTR) of target messenger RNAs (mRNAs) and induce translational repression or target degradation

[19].

In this study, I established eleven PDOs and developed a platform for extracting miRNA from the three-dimensional (3D) model. The CLM organoids and miRNAs were subjected to next-generation sequencing (NGS) and high-throughput drug screening (HTS). Subsequently, multi-omics analyses were conducted using supervised learning algorithms, including DIABLO, LARS, and Random Forest, to identify specific features in liver metastasis (LM). The findings of my study suggest a significant correlation between miR-3613-5p, *eEF2K*, and the anti-cancer drug TAS-102. Furthermore, it was discovered that there is a relationship between the down-regulated gene set regulated by KRAS signaling and the response to anti-cancer drugs in liver metastasis (LM).

Material and Methods

Ethics statement

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

Establishment of tumor derived organoids

The dissociated tumor cells were taken up in Basement Membrane Extract (BME) (Cultrex (R)PC BME RGF type 2, Aimbio) and plated the mixture into a flat-bottomed plate. After the BME gel is polymerized, Human Intestinal Stem Cell medium (HISC) was added for organoid culture. HISC minus Wnt is composed of basal culture medium with 20% R-Spondin conditioned medium, 10% Noggin conditioned medium, 1 x B27, 1.25 mM n-Acetyl Cysteine, 10 mM Nicotinamide, 50 ng/ml human EGF, 10 nM Gastrin, 500 nM A83-01, 3 μM SB202190, 10 nM Prostaglandine E2 and 100 mg/ml Primocin (Vivogen). Culture medium was exchanged every 2 or 3 days. For

passaging cultured organoids, BME was dissolved by applying a physical force by pipetting. Dissolved cells were collected in a tube and were centrifuged at 1,000 rpm for 3 min. 5 ml Triple Express (Invitrogen) was added and the organoids were incubated at 37° C for approximately 5 min. FCS and medium were added and cells were spun down at 1,500 rpm for 3 min. The pellet was resuspended in BME and cells were plated in droplets of 50–100 μL each. After the BME gel is polymerized, Human Intestinal Stem Cell medium (HISC) was added for organoid culture. All organoids were cultured in a humidified incubator at 37 °C containing 5% CO2 and 95% air. Organoids and cell lines established in this study will be governed and distributed worldwide by the leading contact, Dr. Ku Ja-Lok and Korean Cell Line Bank (KCLB).

Establishment of cell line derived organoids

Initial passage (passage 0 ~ 2) cell lines were selected for producing CLM organoid (Table 1). Cell lines were trypsinized by Trypsin. FCS and medium were added, and cells were spun down at 1,500 rpm for 3 min. The dissociated cell pellet was resuspended in BME and plated the mixture into a flat-bottomed plate at different densities.

Subsequent culture methods were identical to those previously mentioned in the *section "Establishment of tumor derived organoids"*.

Table 1. The passage number of cell lines for establishing PDCO

Case Number	Sample Name	Cell line passage
Case 1	SNU-2337A-CO	p.2
Case 1	SNU-2337B-CO	p.2
Case 2	SNU-2536B-CO	p.1
Case 2	SNU-2536C-CO	p.1
Case 3	SNU-2600A-CO	p.2
Case 3	SNU-2600B-CO	p.2
Case 3	SNU-2600C-CO	p.0
Case 4	SNU-3546A-CO	p.1
Case 4	SNU-3546B-CO	p.1

Whole-exome sequencing

Whole-exome capture was performed on all PDOs and PDCOs with the SureSelect Human All Exon V5 Kit (Agilent Technologies, Tokyo, Japan). The captured targets were subjected to sequencing using HiSeq 2500 (Illumina, San Diego, CA, USA) with the pair-end 100 bp read option for organoid samples. The sequence data were processed through an in-house pipeline. In brief, paired-end sequences were first mapped to the human genome, where the reference sequence was UCSC assembly hg19 (original GRCh37

from NCBI, Feb. 2009), using the mapping program BWA (version 0.7.12), and a mapping result file in BAM format was generated using BWA-MEM. Then, Picard-tools (ver. 1.130) were applied in order to remove PCR duplicates. The local realignment process was performed to locally realign reads with BAM files, reducing those reads that identically match a position at start into a single one, using MarkDuplicates.jar, which requires reads to be sorted. By using the Genome Analysis Toolkit, base quality score recalibration (BQSR) and local realignment around indels were performed. The Haplotype Caller of GATK (GATK v3.4.0) was used for variant genotyping for each sample based on the BAM file previously generated (SNP and short indels candidates are detected). Somatic mutations were identified by providing the reference and sequence alignment data of organoids to MuTect2 (involved in GATK v3.8.0) with default parameters using tumor-normal mode. Those variants are annotated by SnpEff v4.1g in vcf file format, filtering with dbSNP for version 142 and SNPs from the 1000 genome project. SnpEff was applied to filter additional databases, including ESP6500, ClinVar, and dbNSFP 2.9.

RNA-sequencing

Paired-end sequencing reads of cDNA libraries (101 bp) generated from a NovaSeq6000 instrument were verified for 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 (UCSC hg19) 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 v1.3.4d. Based on the result of that, the expressional abundance of transcript and gene was calculated as a read count or TPM value (transcript per million mapped reads) per sample.

Gene expression omnibus database analysis

Gene expression omnibus (GEO) database analysis. The GEO database (https://www.ncbi.nlm.nih.gov/geo/) is a web based database providing raw gene expression data. This study used a GEO expression data (GEO accession: GDS4393) investigating responders to FOLFOX therapy of fifty-four unresectable CRC patients. The paired t-test was applied by dividing cluster into

thirty-three pCRCs and twenty-one LM

Pathway analysis

50 canonical pathway gene sets from HALLMARK described in MsigDB (v7.5.1) were used for pathway analysis. To assign pathway enrichment estimates to each organoid, GSVA in the GSVA package (version 1.46.0) was applied with standard settings. Differentially activated pathway analysis was performed using DEseq2 (version 1.38.3).

Exosome isolation

For exosome isolation, culture medium was replaced with serum-free HISC after passaging. Serum-free HISC is composed of serum minus basal culture medium with 20% R-Spondin conditioned medium, 10% Noggin conditioned medium, 1 x B27, 1.25 mM n-Acetyl Cysteine, 10 mM Nicotinamide, 50 ng/ml human EGF, 10 nM Gastrin, 500 nM A83-01, 3 μM SB202190, 10 nM Prostaglandine E2 and 33 mg/ml Primocin (Vivogen). After 10 – 14 days, culture medium was harvested and centrifuged at 2000 x g for 30 min. Supernatant was filtered with 100 um syringe filter and transferred

to new tube. 1/2 volume of exosome isolation (from cell culture media) reagent (cat no. 4478359, Thermo Fisher Scientific, Inc.) was added to filtered medium. Mixed medium-reagent mixture was incubated in 4 °C overnight. After incubation, mixture was centrifuged at 10000 x g 1h 4 °C. Exosome pellet was resuspended by PBS and stored in the deep freezer at -70 ° C.

Western blot for exosome detection

Organoid harvesting solution was added, and the organoids were incubated at 4 °C for approximately 3 h while shaking. Cells were spud down at 1500 rpm in 3 min. Cells were lysed with RIPA lysis buffer (ATTO corporation) adding proteinase inhibitors and phosphatase inhibitors. Protein concentrations were determined using the SMARTTM Micro BCA Protein Assay Kit (Intron Biotechnology, Inc.). Proteins (8 μg) were loaded on Mini-PROTEAN ® TGX Precast Gels (Bio-Rad Laboratories, Inc.) with 4× SDS buffer and transferred to PVDF membranes using the Trans-Blot Turbo™ Transfer Pack (Bio-Rad Laboratories, Inc.). The membranes were blocked at room temperature for 1 h with 2% skim milk in 0.05% TBS-Tween (BD Biosciences). The membranes

were then exposed to primary antibodies for 1–2 h at room temperature against CD63 (RRID:AB_394205, cat. no. 551458; 1:2000; Thermo Fisher Scientific, Inc.) and Calnexin (RRID:AB_2259493, cat no. 2692-1; 1:500; Abcam, UK). Subsequently, membranes were incubated with anti-mouse IgG (H+L) secondary antibody, HRP (RRID:AB_2536527, cat. no. G-21040; 1:5,000; Thermo Fisher Scientific, Inc.) and anti-rabbit IgG (H+L) secondary antibody, HRP (RRID:AB_1500696, cat. no. G-21234; 1:5,000; Thermo Fisher Scientific, Inc.). For visualized the target protein, ECL reagent (Pierce™ ECL Western Blotting Substrate; cat no. 32106; Thermo Fisher Scientific, Inc.) was used.

Exosomal miRNA extraction

Exosomal miRNA was extracted with the Total Exosome RNA & Protein Isolation Kit (cat no. 4478545, Thermo Fisher Scientific, Inc.) Exosome was mixed with 200μl of 2X denaturing solution (pre-warmed at 37 ° C) and incubated on ice for 5 min. 400μl of Acid-Phenol:Chloroform was added to the exosome mixture and centrifuged for 5 min at 10,000 x g at room temperature. The aqueous phase was collected in a new tube without lower phase

contamination. 1.25 volumes of 100% ethanol were added to the aqueous phase. The mixture was placed on the filter cartridge. The filter cartridge with the mixture was centrifuged at 10,000 X g for ~15 s and washed with wash solution. The filter cartridge was transferred to a collection tube. miRNA was eluted with (pre-heated at 95 ° C) nuclease free water.

miRNA microarray

The Affymetrix Genechip miRNA 4.0 array was used for miRNA detection. 1000 ng RNA samples were labeled with the FlashTag™ Biotin RNA Labeling Kit (Genisphere, Hatfield, PA, USA). The labeled RNA was quantified, fractionated, and hybridized to the miRNA microarray according to the standard procedures provided by the manufacturer. The labeled RNA was heated to 99 °C for 5 minutes and then to 45 °C for 5 min. RNA-array hybridization was performed with agitation at 60 rotations per minute for 16 h at 48 ° C on an Affymetrix GeneChip Hybridization over 645. The chips were washed and stained using a Genechip Fluidics Station 450 (Affymetrix, CA, United States). The chips were then scanned with an Affymetrix GCS 3000 canner (Affymetrix). Raw data (CEL file)

were generated by Affymetrix Genechip Command Console Software (AGCC). CEL files were normalized, checked for probe quality and filtered by the manufacturer's procedure. Only quality checked probes (330) were used for miRNA analysis.

miRNA target prediction

Top 20 miRNAs were selected through comparative analysis between pCRC and LM and target genes were predicted through miRWalk 3.0. Then, the target genes' expression amount was evaluated on the mRNA-seq. In detail, the target genes of the up-regulated miRNAs were sorted by only the common gene, whose $\log_2 FC \leq 1$ on the RNA-seq data. Conversely, the target genes of the down-regulated miRNAs were sorted by the common gene, which has $\log_2 FC \geq 1$. The schematic diagram of miRNA target prediction is in Figure 1.

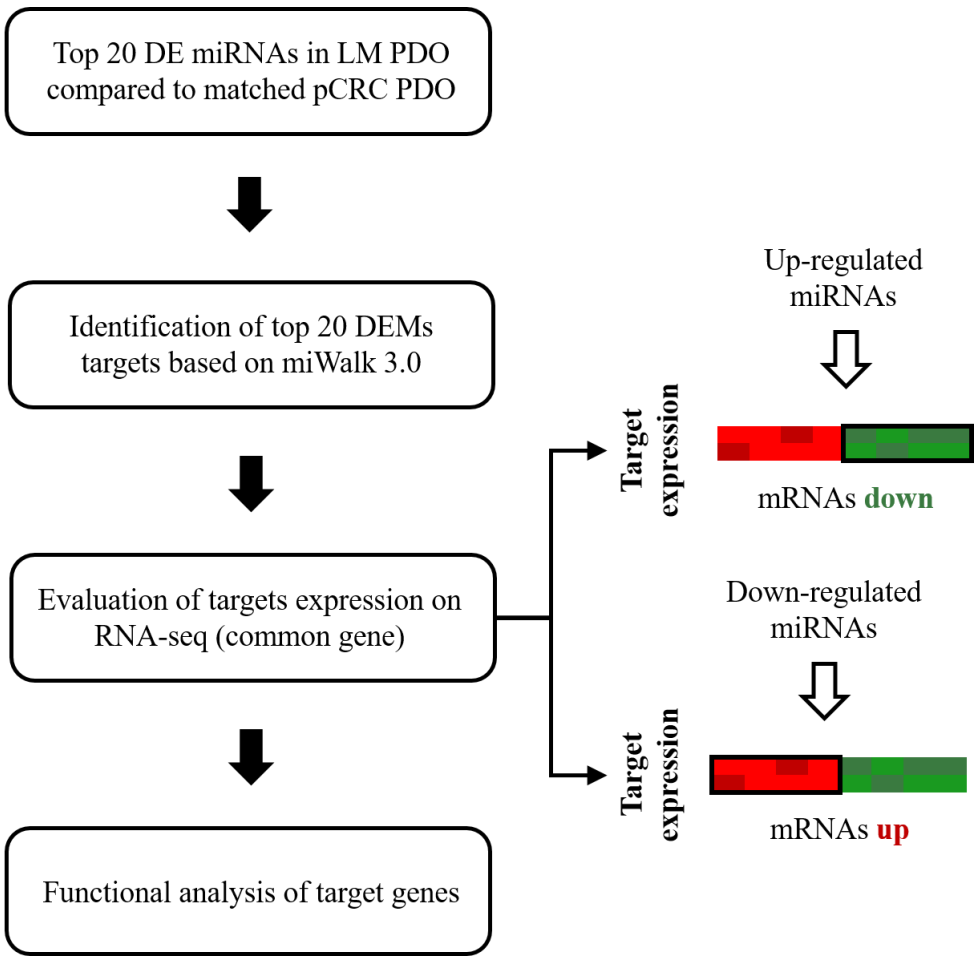

Figure 1. Scheme of miRNA target prediction.

Organoids drug sensitivity test

All drug screens were repeated more than three times. Organoids were enzymatically and mechanically dissociated into single cells by incubating and pipetting in TrypLE Express solution for 5 to 10 min. The mixture was spun down at 1500 rpm for 3 min. After dissolving

BME, the cell pellet was resuspended with 7:3 mixtures of HISC medium and BME. Cell–BME–HISC mixture (60 μL, 20 000 organoids mL−1) was plated homogeneously in clear–bottomed, white–walled 96–well plates (Corning, Cat# 3903) using F1–ClipTip Multichannel Pipettes (Thermo Fisher Scientific, MA, USA). After allowing the BME to be polymerized, pre–warmed HISC medium (20 μL) was added to each well using an F1– ClipTip Multichannel Pipettes. 72 h after seeding, drug solution (20 μL) was added to each well using F1– ClipTip Multichannel Pipettes. In the HISC medium, the drugs were serially diluted at a ratio of 1:3 from the maximum dose to produce six dose points. The mixture of HISC medium and drug–solvent solution was added to all plates for the control group. The drug list used in this study is documented in Table 2.

Table 2. Compounds list of organoid drug screening.

Drug	Commercial name	Max conc. (μM)	Target
TAS-102	LONSURF	100	nucleoside metabolism
Regorafenib	Stivarga	100	VEGFR1/2/3, PDGFR-β, FGFR, c-KIT, c-RET, RAF
Capecitabine	Xeloda	1000	nucleoside metabolism
Apitolisib (GDC-0980)		50	mTOR
Belinostat (PXD101)	Beleodaq	100	histone deacetylase(HDAC)
Trametinib (GSK1120212)	Mekinist	50	MEK1/2
cyclopamine		50	Smoothened(Smo)
ICG-001	Foscenvivint	100	β-catenin
Buparlisib	Norvatis	100	PI3K
Vorinostat	SAHA	50	histone deacetylase(HDAC)
Afatinib	Gilotrif	50	ErbB family of tyrosine kinases
Vistusertib (AZD2014)		5	mTORC 1/2
MK-5108		100	Aurora A kinase
Olaparib	LYNPARZA	50	poly ADP-ribse polymerase(PARP) enzyme
Irinotecan Hydrochloride	Camptosar	100	topoisomerase I
5-FU (Fluorouracil injection)	Efudex	20000	deoxyribonucleic acid(DNA) synthesis
Oxaliplatin	Eloxatin	100	DNA synthesis
Dabrafenib	Tafinlar	10	BRAF
MK-2206		4	AKT 1/2/3
Cabozantinib	Cometriq	10	tyrosine kinase receptor(VEGFR2, c-MET, RET)
Everolimus (RAD001)	Afinitor	10	mTOR

Multi-omics integration

DIABLO method of the mixOmics package (version 6.22.0) was used to integrate RNA-seq, miRNA expression, and drug response data. Each piece of data was normalized and filtered by its own method to be used as an input for data integration. The matrix used for data integration consisted of three blocks: RNA-seq (319 features) x miRNA microarray (296 features) x drug responses (21 features) estimating more than two thousand multi-omics combinations. While running the tuning function tune.black.splsda (), parameters called Overall.BER and centroids.dist were used for best accuracy [20].

Construction of prediction model

R packages were used for all estimates and visualization [21]. To identify the highest performance, six different machine learning models were used: Decision tree[22, 23], k-Nearest-Neighbor [24], Logistic regression [25], Naïve bayes [26], Neural network [27], and Random forest [28]. Comparison and visualization was done by modelr package [29]. The estimation of accuracy and confusion matrix is done by yardstick within tidymodels package [25] to quantify how well model fits to a data set. I checked and found that

the overall prediction results did not depend very much on the machine-learning method and parameter tuning (which requires a huge computational resource), but more on the set of meaningful biological features (data not shown). Pathway enrichment score (NES) and miRNA expression are used to select genomic features for training machine learning models. In addition, I incorporated the response of 21 drugs in the unit of AUC.

I have implanted the automatic machine learning package Lares [9] with the following settings: split = 0.7 (train set), target = "metastasis", no_outliers = TRUE, unique_train = TRUE, center = FALSE, scale = FALSE, nfolds = 5, max_models = 500, thresh = 10. All appropriate algorithms, including Distributed Random Forest (DRF), Extremely Randomized Trees (XRT), and Generalized Linear Model with regularization are used if the search stopping criteria allow.

Statistical analysis

All statistical analysis was conducted using the R program (version 4.2.2). A p-value of 0.05 was considered statistically significant. To determine the optimal number of clusters, use the elbow and

silhouette methods of the factoextra (version 1.0.7). Multidimensional scaling was conducted using limma (version 3.54.1).

Results

11 CRC and paired LM organoids were successfully established

Primary CRC (pCRC) and paired LM lesions were collected from 5 patients, resulting in five matched CLM PDOs. Four cases (SNU-2337, SNU-2536, SNU-2600, and SNU-3546 case) are derived from initial passages of previously established cell lines (-CO). The case of SNU-5455 was directly originated from the patient tissues (-TO). The name given to an established organoid indicates the institute, sample number, sampling location, and model type in order (Figure 2). STR profiling confirmed that there was no cross-contamination among samples (Table 2).

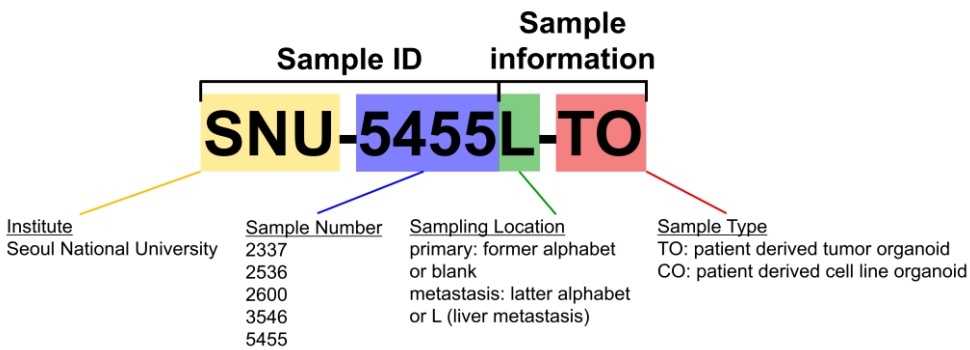

Figure 2. Nomenclature of established organoids.

PDOs had various morphologies, even from the same patient (Figure 3). One or more lumens at the center were typical morphologic features of CLM organoids. Both cystic and solid forms were also observed in all case of organoids. No standing morphological difference was observed between pCRC and LM organoids.

Eleven CLM organoids were subjected to whole-exome sequencing (WES), RNA sequencing (RNA-seq), and drug screening. The medium used during organoid culture was harvested to perform miRNA microarray. Various omics data were integrated and used to identify the characteristics between pCRC and LM that offer a clinically effective therapeutic target. The scheme of the experiment is given in Figure 4.

Table 3. STR profiles of 11 CLM organoids with matched normal, tissues and cell lines.

Sample Name	Sample Type	Amelogenin	D3S1358	D1S1656	D2S441	D10S1248	D13S317	Penta E	D16S539	D18S51	D2S1338	CSF1PO	Penta D
SNU-2337A-CO	Cell Line Derived Organoid	X,Y	13,16	15,16	11,14	14,16	12,13	20,21	10,13	16	18,24	11,12	9
SNU-2337A	Cell Line	X,Y	16	15,16	11,14	14,16	13	20,21	10,13	16	18,24	11,12	9
SNU-2337B-CO	Cell Line Derived Organoid	X,Y	13,16	15,16	11,14	14,16	12,13	20,21	10,13	16	18,24	11,12	9
SNU-2337B	Cell Line	X,Y	13,16	15,16	11,14	14,17	12,13	20,21	10,13	16	18,24	11,12	9
SNU-2536B	Cell Line	X	16				8,11		11	10,14	18,23	10,11	
SNU-2536B-CO	Cell Line Derived Organoid	X	16	16	11,14	13,15	11	11,15	11	14	18,23	10,11	11,12
SNU-2536C	Cell Line	X	16	16	11,14	13	8,11	11,15	11	14	18,23	10,11	11,12
SNU-2536C-CO	Cell Line Derived Organoid	X,Y	16	16	11,14	13,15	8,11	11,15	11	14	18,23	10,11	11,12
SNU-2600A-CO	Cell Line Derived Organoid	X	14,16	15,18	11,13	14,17	8,11	5,17	9	18	19,23	10,11	9,12
SNU-2600A	Cell Line	X	14,16	15,18	11,13	14,17	8,11	5,17	9	18	19,23	10,11	9,12
SNU-2600B-CO	Cell Line Derived Organoid	X,Y	14,16	15,18	11,13	14,17	8,11	5,17	9	18	19,23	10,11	9,12
SNU-2600B	Cell Line	X,Y	14,16	15,18	11,13	14,17	8,11	5,17	9	18	19,23	10,11	9,12

Continued

Sample Name	Sample Type	Amelogenin	D3S1358	D1S1656	D2S441	D10S1248	D13S317	Penta E	D16S539	D18S51	D2S1338	CSF1PO	Penta D
SNU-2600C-CO	Cell Line Derived Organoid	X,Y	14,16	15,18	11,13	14,17	8,11	5,17	9	18	19,23	10,11	9,12
SNU-2600C	Cell Line	X,Y	14,16	15,18	11,13	14,17	8,11	5,17	9	18	19,23	10,11	9,12
SNU-3546A	Cell Line	X	16				8,13		9,12	16	21,25	11	
SNU-3546A-CO	Cell Line Derived Organoid	X	16	14,18.3	9.1,10	13	8,13	12,13	9,12	16	21,25	11	10,11,12
SNU-3546B	Cell Line	X	16				8,13		9,12	16	21,25	11	
SNU-3546B-CO	Cell Line Derived Organoid	X	16	14,18.3	9.1,10	13	8,13	12,13	9,12	16	21,25	11	10,12
SNU-5455-TO	Organoid	X,Y	17				9,12		9,12	16	25	10,12	
SNU-5455_TilTissue	Til Tissue	X,Y	17				9,12		9,12	15,16	25	10,12	
SNU-5455_NormalTissue	Tissue	X,Y	17				9,12		9,12	15,16	25	10,12	
SNU-5455	Cell Line	X,Y	17	13,15	11,14	13	9,12	10	9,12	16	25	10,12	10,11
SNU-5455L-TO	Organoid	X,Y	17				9,12		9,12	16	25	10,12	
SNU-5455LT	Cell Line	X	17	13,15	11,14	13	12	10	9,12	16	25	10,12	10,11

Continued

Sample Name	TH01	Vwa	D21S11	D7S820	D5S818	TPOX	D8S1179	D12S931	D19S433	D61043	D22S1045	DYS391	FGA	DYS576	DYS570
SNU-2337A-CO	9,10	17,18	29	12	11,12	8,9	11,13	15	14.2	12	17	10	20,24	18	19
SNU-2337A	9,10	17,18	29	12	11,12	8,9	11,13	15	14.2	12	17	10	20,24	18	19
SNU-2337B-CO	9,10	17	29	12	11,12	8,9	11,13	15	14,14.2	12	17	10	20,24	18	19
SNU-2337B	9,10	17	29	12	11,12	8,9	11,13	15	14,14.2	12	17	10	20,24	18	19
SNU-2536B	6,9	16,18	29,30	8,11	10,13	8,10			13,14.2				22		
SNU-2536B-CO	6,9	16,18	29,30	8,11	10,13	8,10	15	21	13,14.2	11,19	15		22		
SNU-2536C	6,9	16,18	29,30	8,11	10,13	8,10	15	21	13,14.2	11,19	15		22		
SNU-2536C-CO	6,9	16,18	29,30	8,11	13	8,10	15	21	13,14.2	19	15	10	22	17	19
SNU-2600A-CO	7,9	14,17	30,32.2	12,13	13	8,11	15,17	17,19	13,14	12,19	11,15	-	23	-	-
SNU-2600A	7,9	17	30,32.2	12,13	13	8,11	15,17	19	13,14	12,19	11,15		23		
SNU-2600B-CO	7,9	14,17	30,32.2	12,13	13	8,11	15,17	17,19	13,14	12,19	11,15	10	23	17	19
SNU-2600B	7,9	17	30,32.2	12,13	13	8,11	15,17	19	13,14	12,19	11,15	10	23	17	19

Continued

Sample Name	TH01	Vwa	D21S11	D7S820	D5S818	TPOX	D8S1179	D12S931	D19S433	D61043	D22S1045	DYS391	FGA	DYS576	DYS570
SNU-2600C-CO	7,9	14,17	30,32.2	12,13	13	8,11	15,17	17,19	13,14	12,19	11,15	10	23	17	19
SNU-2600C	7,9	14,17	30,32.2	12,13	13	8,11	15,17	17,19	13,14	12,19	11,15	10	23	17	19
SNU-3546A	6	16	28,29	11	11	8,11			13				22,24		
SNU-3546A-CO	6	16	28,29	11	11	8,11	11,13	18,20	13	12,16	16,17	-	22,24	-	-
SNU-3546B	6	16	28,29	11	11	8,11			13				22,24		
SNU-3546B-CO	6	16	28,29	11	11	8,11	11,13	18,19,20	13	12,16	16,17	-	22,24	-	-
SNU-5455-TO	7,9	17,18	31	8,12	9,12	10,11			13,14				24		
SNU-5455_TilTissue	7,9	17,18	30,31	8,12	9,12	10,11			13,14				22,24		
SNU-5455_NormalTissue	7,9	17,18	30,31	8,12	9,12	10,11			13,14				22,24		
SNU-5455	7,9	17,18	31	8,12	9,12	10,11	10,15	21,23	13,14	11,12,18	17	10	24	18	18
SNU-5455L-TO	7,9	17,18	31	8,12	9,12	10,11			13,14				24		
SNU-5455LT	7,9	17	31	8	12	10,11	10,15	21,23	13,14	11,18	17		24		

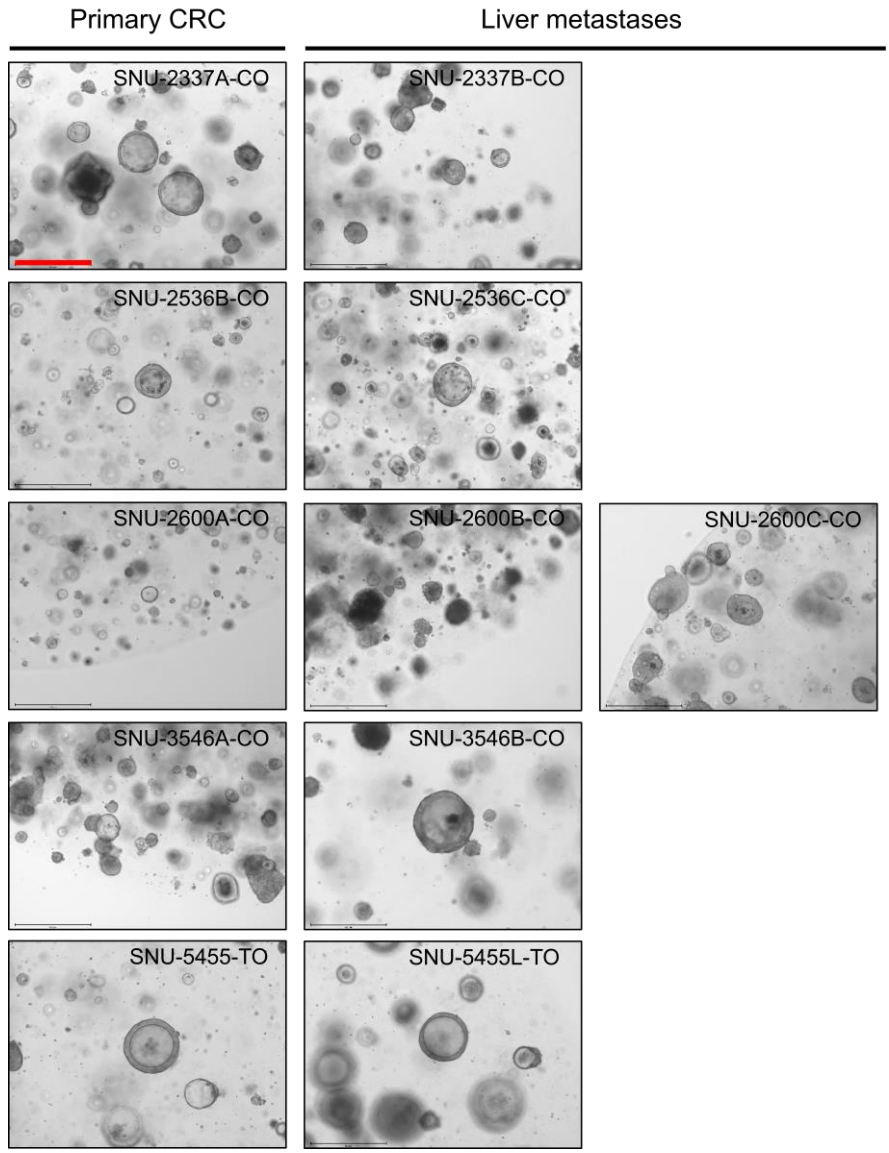

Figure 3. The morphology of CLM organoids. The CLM PDOs had a heterogenous morphologic characteristic. Organoids formed lumens, cystic structure and cell aggregates are mixed together in each sample. Organoids had similar morphology regardless of tumor type (pCRC or LM). Scale bar (red bar) = 200μm. Magnification = 10X.

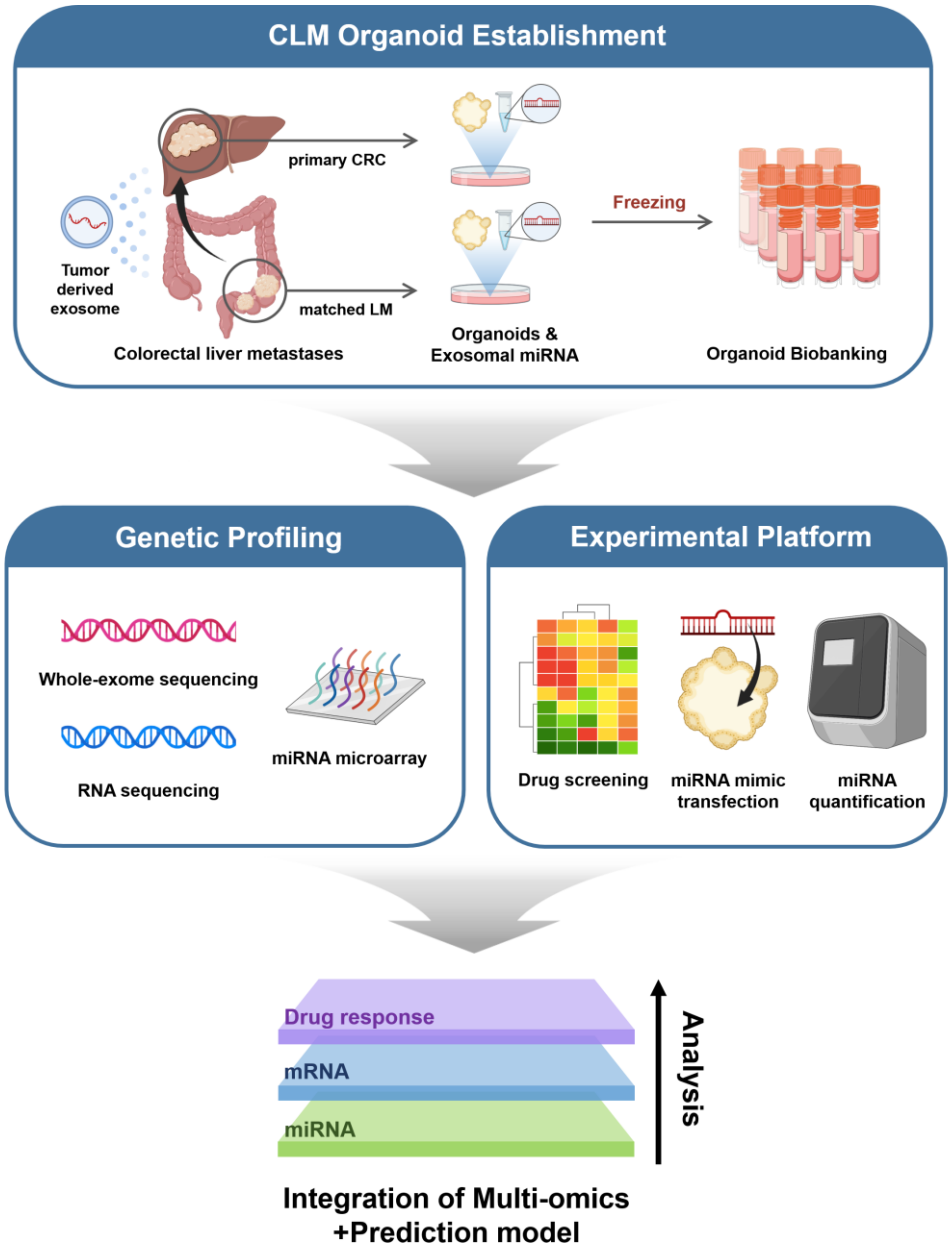

Figure 4. Graphical scheme of study. Eleven PDOs and PDCOs were successfully established, and exosomes were extracted from culture media. The CLM organoids and miRNAs were subjected to whole exome sequencing (WES), RNA sequencing (RNA-seq), miRNA

microarray, and drug screening. Subsequently, multi-omics analyses and prediction model construction were performed.

Liver metastatic organoids had identical mutational characteristics of primary CRC organoids

WES was conducted to identify the mutational characteristics of CLM organoids. The normal mucosa was acquired together during surgery and used as a criterion to sort somatic mutations for the SNU-5455 case. In the case of other samples, somatic mutations were identified based on an in-house normal colorectal organoid panel. All samples showed only known pathogenic mutations from Clinvar. Several studies report that *APC*, *KRAS*, and *TP53* mutations are major genes in CRC tumorigenesis [30]. In my cohort, mutations of CRC driver genes including *APC*, *KRAS*, and *TP53* were observed in the organoids of SNU-2337, SNU-2600, SNU-3546, and SNU-5455 (Figure 5). Exceptionally, SNU-2536 only harbored *SACS*, *FBN1*, and *SLC4A1* mutations. A recent study reported that *SACS* and *FBN1* have oncogenic functions in CRC [31, 32]. Intriguingly, I found that mutations in primary cancer organoids were equal to mutation in metastatic organoids in all PDOs except for the *RFX5* c.353+2T>G splice sites mutation in SNU-2337A-CO (Figure 5; Table 4). Furthermore, I could not observe the excessive VAF value variation in the process of metastasis (Table 4).

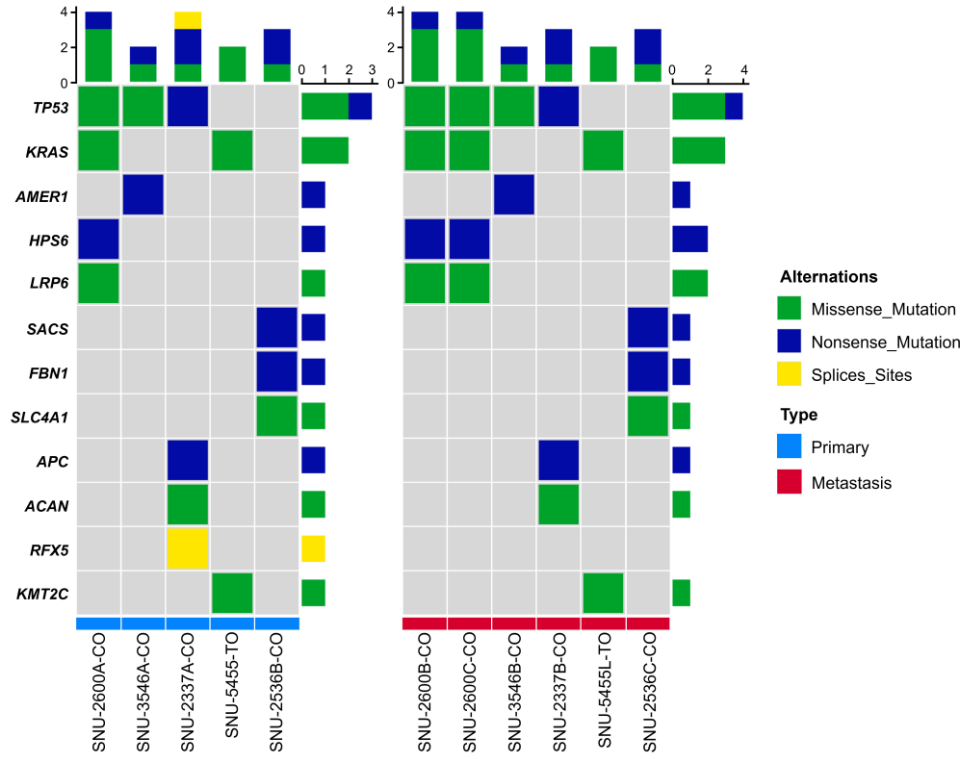

Figure 5. Mutational landscape of somatic mutations in CLM cohort. Adjacent normal mucosa (for SNU-5455) and 50 normal colorectal organoid panel (for SNU-2337; SNU-2536; SNU-2600; SNU-3546) were used for eliminating germline mutations. LM organoids had almost identical somatic mutations as pCRC organoids.

Table 4. Mutational profiles of 11 CLM organoids.

Sample Name	Gene Symbol	Chromosome	HGVS.c	HGVS.p	DP	AD	VAF	Variant
SNU-2337A-CO	RFX5	chr1	c.353+2T>G	.	10	2	0.2	splice_donor_variant&intron_variant
SNU-2337A-CO	APC	chr5	c.4012C>T	p.Gln1338*	70	70	1	stop_gained
SNU-2337A-CO	ACAN	chr15	c.7204C>T	p.Arg2402Cys	19	19	1	missense_variant
SNU-2337A-CO	TP53	chr17	c.637C>T	p.Arg213*	69	69	1	stop_gained
SNU-2337B-CO	APC	chr5	c.4012C>T	p.Gln1338*	37	37	1	stop_gained
SNU-2337B-CO	ACAN	chr15	c.7204C>T	p.Arg2402Cys	14	14	1	missense_variant
SNU-2337B-CO	TP53	chr17	c.637C>T	p.Arg213*	48	48	1	stop_gained
SNU-2536B-CO	SACS	chr13	c.10855G>T	p.Glu3619*	148	39	0.263513514	stop_gained
SNU-2536B-CO	FBN1	chr15	c.4615C>T	p.Arg1539*	35	6	0.171428571	stop_gained
SNU-2536B-CO	SLC4A1	chr17	c.1331C>A	p.Thr444Asn	45	45	1	missense_variant
SNU-2536C-CO	SACS	chr13	c.10855G>T	p.Glu3619*	132	32	0.242424242	stop_gained
SNU-2536C-CO	FBN1	chr15	c.4615C>T	p.Arg1539*	41	21	0.512195122	stop_gained
SNU-2536C-CO	SLC4A1	chr17	c.1331C>A	p.Thr444Asn	29	29	1	missense_variant
SNU-2600A-CO	HPS6	chr10	c.2038C>T	p.Gln680*	171	96	0.561403509	stop_gained
SNU-2600A-CO	LRP6	chr12	c.1418G>A	p.Arg473Gln	186	92	0.494623656	missense_variant
SNU-2600A-CO	KRAS	chr12	c.35G>T	p.Gly12Val	92	42	0.456521739	missense_variant
SNU-2600A-CO	TP53	chr17	c.817C>T	p.Arg273Cys	36	34	0.944444444	missense_variant

Continued

Sample Name	Gene Symbol	Chromosome	HGVS.c	HGVS.p	DP	AD	VAF	Variant
SNU-2600B-CO	HPS6	chr10	c.2038C>T	p.Gln680*	122	65	0.532786885	stop_gained
SNU-2600B-CO	LRP6	chr12	c.1418G>A	p.Arg473Gln	86	54	0.627906977	missense_variant
SNU-2600B-CO	KRAS	chr12	c.35G>T	p.Gly12Val	49	27	0.551020408	missense_variant
SNU-2600B-CO	TP53	chr17	c.817C>T	p.Arg273Cys	28	28	1	missense_variant
SNU-2600C-CO	HPS6	chr10	c.2038C>T	p.Gln680*	134	67	0.5	stop_gained
SNU-2600C-CO	LRP6	chr12	c.1418G>A	p.Arg473Gln	153	58	0.379084967	missense_variant
SNU-2600C-CO	KRAS	chr12	c.35G>T	p.Gly12Val	64	15	0.234375	missense_variant
SNU-2600C-CO	TP53	chr17	c.817C>T	p.Arg273Cys	34	34	1	missense_variant
SNU-3546A-CO	TP53	chr17	c.659A>G	p.Tyr220Cys	66	65	0.984848485	missense_variant
SNU-3546A-CO	AMER1	chrX	c.1489C>T	p.Arg497*	40	40	1	stop_gained
SNU-3546B-CO	TP53	chr17	c.659A>G	p.Tyr220Cys	38	38	1	missense_variant
SNU-3546B-CO	AMER1	chrX	c.1489C>T	p.Arg497*	39	39	1	stop_gained
SNU-5455-TO	KMT2C	chr7	c.925C>T	p.Pro309Ser	296	58	0.195945946	missense_variant
SNU-5455-TO	KRAS	chr12	c.35G>A	p.Gly12Asp	92	47	0.510869565	missense_variant
SNU-5455L-TO	KMT2C	chr7	c.925C>T	p.Pro309Ser	284	47	0.165492958	missense_variant
SNU-5455L-TO	KRAS	chr12	c.35G>A	p.Gly12Asp	94	49	0.521276596	missense_variant

In order to investigate the impact of mutations on the metastatic process, MutationalPatterns packages in R was utilized to examine the mutation patterns in CLM organoids. Initially, I observed that C>T transitions occurred most frequently in our cohort (Figure 6). Comparing the mutational patterns of cancer driver genes between pCRC and LM, I found that they were nearly identical. This implies that there were no newly discovered mutation associated with liver metastasis.

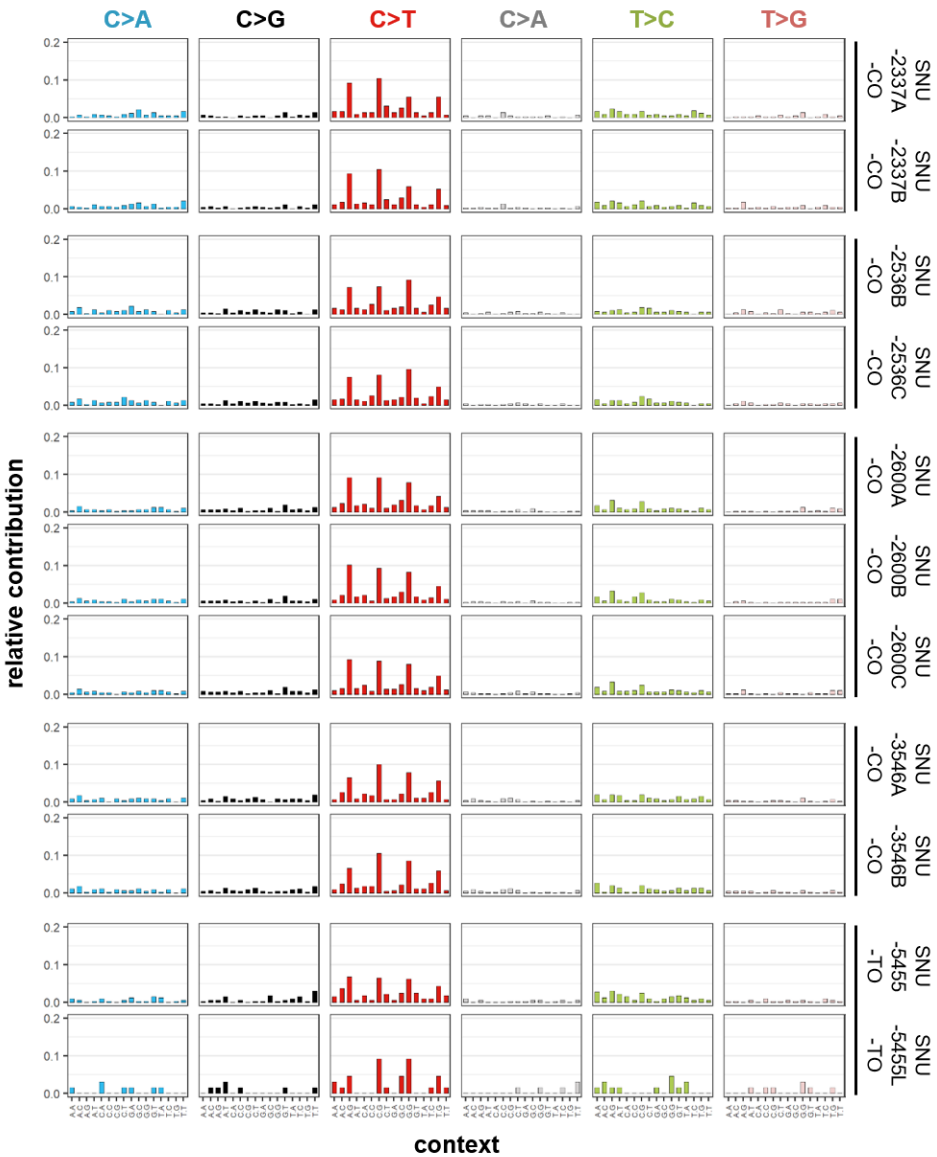

Figure 6. Mutation patterns of cancer driver genes in CLM organoid. Y axis means proportion of mutation and x axis means trinucleotides (sequence) context of mutation. Overally, CLM organoids showed similar mutation patterns among organoids derived from the same patient.

Transcriptomic analysis elucidates high heterogeneity between patient in CLM organoids

I plotted the multi-dimensional scaling (MDS) based on the total 35,994 normalized mRNA expressions to understand the overall transcriptome feature in my cohort. Transcriptomic features were similar between initial and metastatic PDOs derived from the same patient compared to inter-patient comparison (Figure 7). Continuously, using the DESeq2 in the R package, differentially expressed genes (DEGs) analysis was conducted to compare dysregulated gene expression pattern between pCRC and LM organoids. Normalized read-count values of CLM organoids were analyzed by paired t-test A comparison result showed twenty-one genes were highly expressed in LM organoids ($|\log_2$ Fold Change (FC)$| \geq 1$, p-value ≤ 0.05) (Figure 8). Among the DEGs, eleven protein-coding genes were identified including *CYSTM1* (Cysteine Rich Transmembrane Module Containing 1), *CCDC7* (Coiled-Coil Domain Containing 7), *DRAM2* (DNA Damage Regulated Autophagy Modulator 2), *CD300LD* (CD300 Molecule Like Family Member D), *ARL5B* (ADP Ribosylation Factor Like GTPase 5B), *CXorf58* (Chromosome X Open Reading Frame 58), *CASP1* (Caspase 1), *CENPI* (Centomere Protein I), *GCNT1* (Glucosaminyl Transferase 1),

GIMAP1 (GTPase, IMAP Family Member 1), *GOPC* (Golgi Associated PDZ And Coiled-Coil Motif Containing), and *ALX3* (ALX Homeobox 3). There was no low-expression gene satisfying statistical significance in LM compared to pCRC (Figure 8).

Eleven up-regulated genes in LM were validated using the GEO database (GEO accession: GDS4393). The gene expression levels of fifty-four CLM patients were extracted and used for comparative analysis. The paired t-test was applied by dividing cluster into thirty-three pCRC and twenty-one LM. I was able to confirm the expression patterns of eight genes except for *CYSTM1*, *CCDC7*, and *CD300LD*. In GEO database, *CXorf58*, *CENPI*, *GIMAP1*, and *GOPC* were shown to be up-regulated in LM compared to pCRC (Figure 9). Among them, *CENPI* was identified as a statistically significant gene, with *p-value* of 0.05 or less. *CENPI* is a member of centromere protein family, forming efficient and correct microtubule attachment. It is reported that up-regulated of *CENPI* increased CRC cell migration and invasion. My study also confirmed that CENPI, as an oncogene, is involved in metastasis of CRC [33].

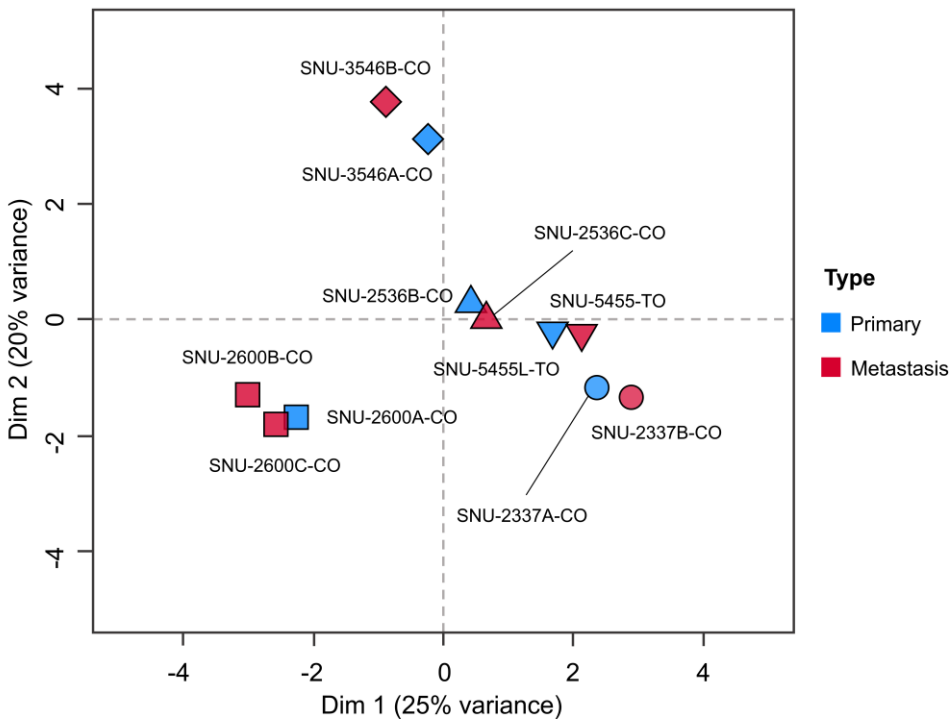

Figure 7. Multidimensional scaling (MDS) of CLM organoids based on 35,994 normalized gene expression. Color represents tumor types (blue, pCRC; red, LM). The sample origin was classified into shapes. Organoids derived from the same patients are located close to each other on the MDS plot.

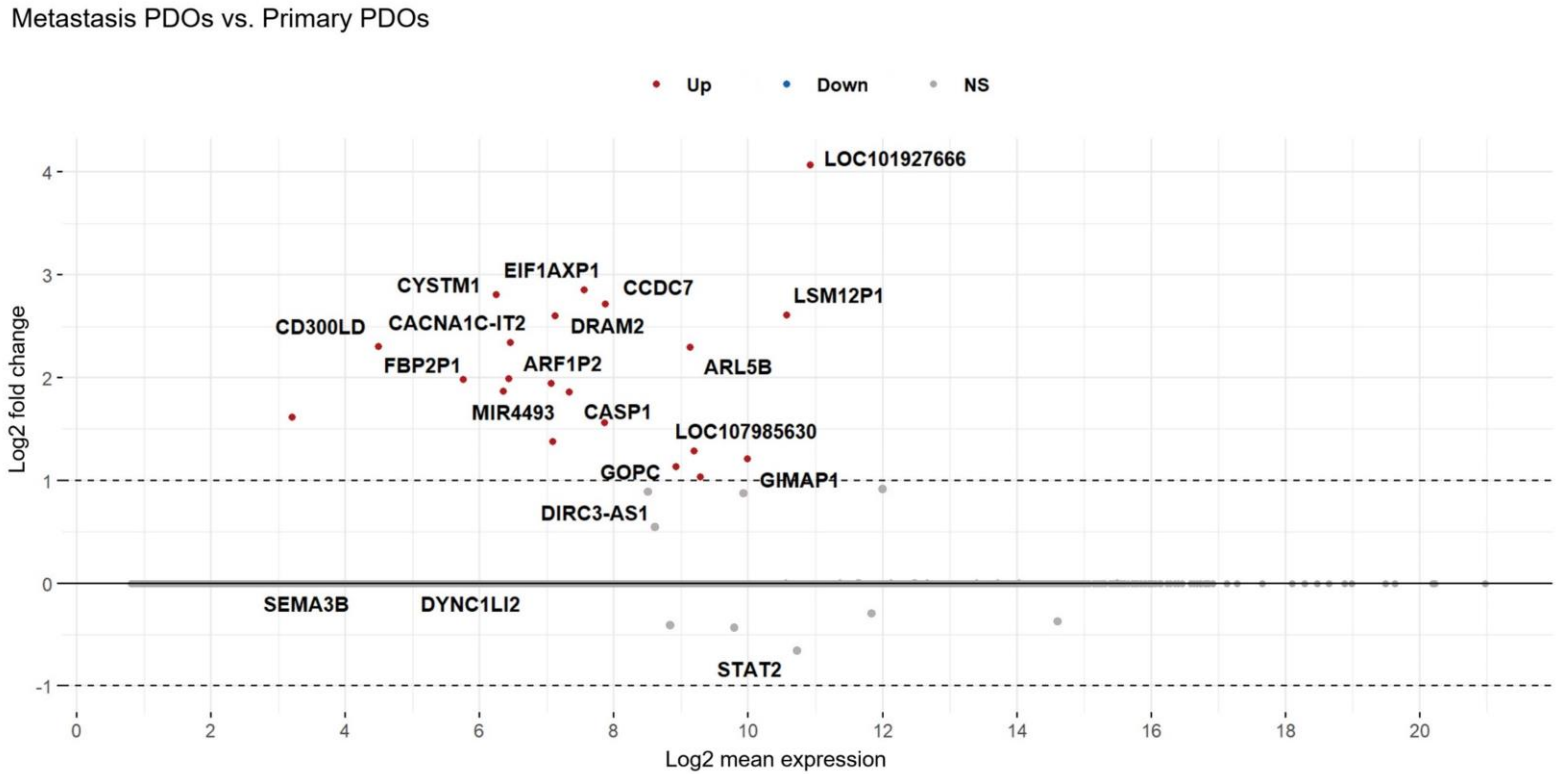

Figure 8. MA-plot depicting differentially expressed genes (DEGs) of CLM organoids.

x axis = Log_2 mean expression, y axis = Log_2FC. Up-regulated genes, down regulated genes, and non-significantly expressed genes were differently represented (red, up-regulated; blue, down-regulated; gray, non-significant). The normalized read-count values of pCRC and LM organoids were analyzed by paired t-test to compare gene expression between them.

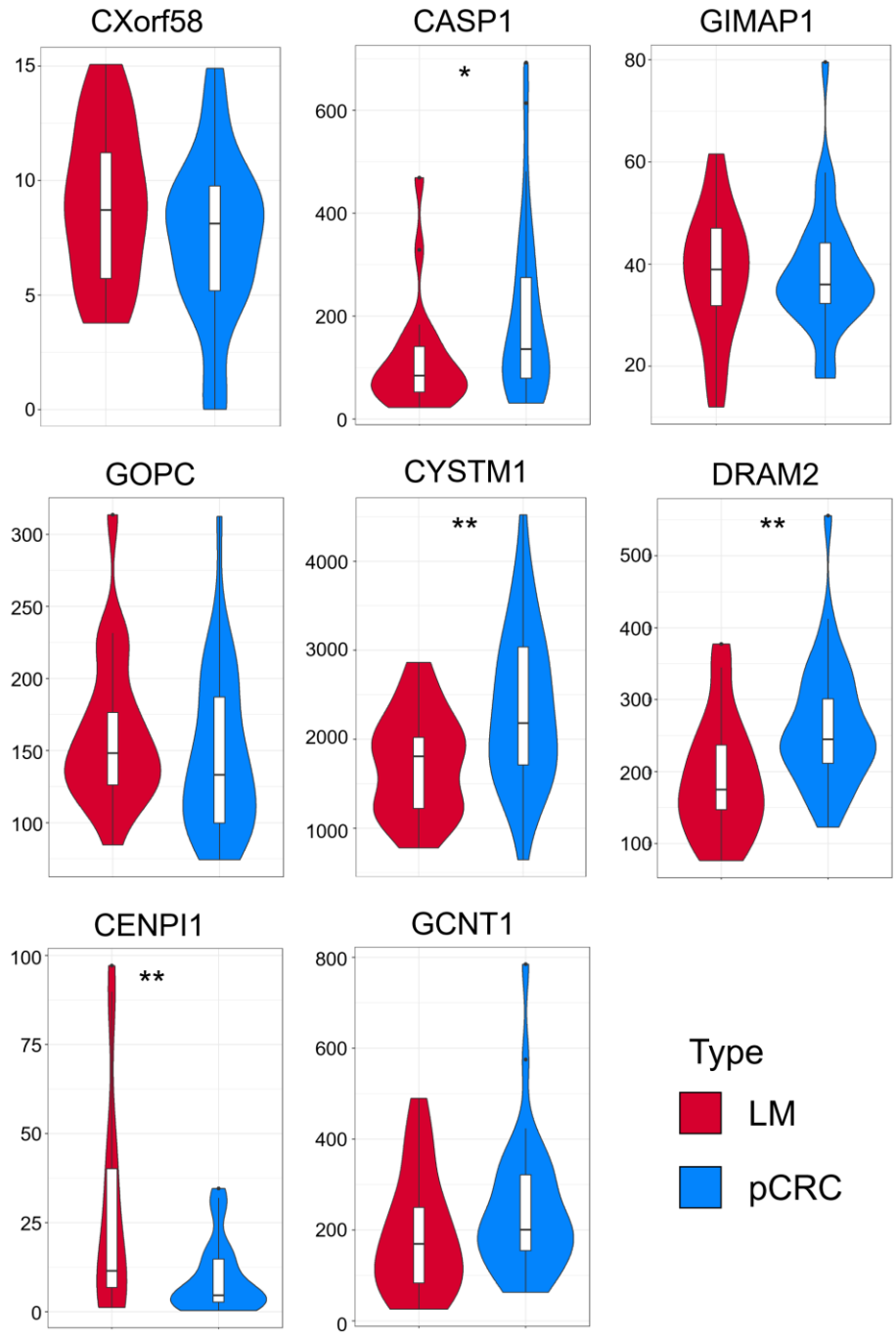

Figure 9. GEO database analysis of up-regulated genes in LM organoids. The following analysis was performed for further

validation. The expression data in CLM were extracted from GEO expression data (GEO accession: GDS4393). The paired t-test was applied between thirty-three pCRCs and twenty-one LM. If a *p-value* is less than 0.05, it is marked with a single star (*). If a *p-value* is less than 0.01, it is marked with two stars (**).

A gene set variation analysis was performed to validate the enriched pathways in each PDO. Pathway estimates were assigned to each organoid using the GSVA in the R package. The entire transcriptome read count (n = 35,994) was utilized for the GSVA analysis. In general, pathways associated with the development process exhibited a high enrichment score, while those related to the immune process showed a low enrichment score (Figure 10; Table 5). Notably, the hedgehog pathway, which plays a role in cell signaling, was highly enriched in CLM organoids. Additionally, the epithelial mesenchymal transition (EMT) pathway and angiogenesis, both crucial processes for metastasis, were examined. The EMT pathway was highly enriched in 83% (5 out of 6) of pCRC organoids, while the angiogenesis pathway was highly enriched in 66% (4 out of 6) of pCRC organoids and in 2 out of 6 LM organoids. It was expected that the enrichment patterns would be similar between primary and metastatic tumors. However, there was no significant consistency observed between the two types (Table 5).

Table 5. Hallmark pathway enrichment score of CLM organoids.

Hallmark pathway	SNU-2337A-CO	SNU-2337B-CO	SNU-2536B-CO	SNU-2536C-CO	SNU-2600A-CO	SNU-2600B-CO	SNU-2600C-CO	SNU-3546A-CO	SNU-3546B-CO	SNU-5455-TO	SNU-5455L-TO
HALLMARK_ADIPOGENESIS	−0.07	−0.01	−0.22	−0.09	−0.04	0.02	0.02	−0.08	−0.13	−0.07	−0.15
HALLMARK_ALLOGRAFT_REJECTION	−0.41	−0.46	−0.31	−0.46	−0.48	−0.47	−0.45	−0.42	−0.39	−0.43	−0.44
HALLMARK_ANDROGEN_RESPONSE	0	0.06	−0.01	−0.06	−0.13	0.01	0.16	−0.07	−0.11	−0.17	−0.1
HALLMARK_ANGIOGENESIS	0.21	−0.22	−0.05	−0.24	0.1	−0.03	−0.12	0.14	−0.01	0.07	0.35
HALLMARK_APICAL_JUNCTION	−0.21	−0.16	−0.17	−0.14	−0.02	−0.17	−0.17	0	−0.03	−0.16	−0.2
HALLMARK_APICAL_SURFACE	−0.25	−0.14	−0.18	0.05	−0.18	0.03	−0.07	0.39	0.12	−0.2	−0.21
HALLMARK_APOPTOSIS	−0.06	−0.04	0.02	−0.01	−0.16	−0.19	−0.08	−0.11	−0.21	−0.22	−0.2
HALLMARK_BILE_ACID_METABOLISM	−0.07	−0.04	−0.07	0.28	0.06	−0.07	−0.01	−0.14	0	0.01	0.1
HALLMARK_CHOLESTEROL_HOMEOSTASIS	−0.05	−0.17	−0.21	−0.15	−0.12	−0.09	−0.04	0.02	0.11	−0.09	−0.02
HALLMARK_COAGULATION	−0.21	−0.12	−0.23	−0.3	−0.1	−0.16	0.02	−0.19	−0.06	−0.24	−0.15
HALLMARK_COMPLEMENT	−0.16	−0.15	−0.22	−0.17	−0.18	−0.19	−0.03	−0.19	−0.04	−0.15	−0.08
HALLMARK_DNA_REPAIR	−0.13	−0.15	−0.06	−0.07	−0.09	−0.19	−0.1	−0.13	−0.07	−0.01	0.11

Continued

Hallmark pathway	SNU-2337A-CO	SNU-2337B-CO	SNU-2536B-CO	SNU-2536C-CO	SNU-2600A-CO	SNU-2600B-CO	SNU-2600C-CO	SNU-3546A-CO	SNU-3546B-CO	SNU-5455-TO	SNU-5455L-TO
HALLMARK_E2F_TARGETS	−0.12	−0.06	0	−0.06	−0.01	0.07	0	0.06	0.19	−0.08	−0.13
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	−0.07	−0.1	−0.02	−0.17	0.14	0.03	0.22	−0.11	−0.14	0.02	−0.07
HALLMARK_ESTROGEN_RESPONSE_EARLY	−0.13	−0.04	−0.1	−0.12	−0.08	−0.16	−0.13	−0.23	−0.01	−0.16	−0.08
HALLMARK_ESTROGEN_RESPONSE_LATE	−0.15	−0.07	−0.1	−0.04	−0.13	−0.21	−0.06	−0.17	−0.1	−0.1	−0.15
HALLMARK_FATTY_ACID_METABOLISM	−0.12	−0.04	−0.06	−0.07	−0.03	−0.04	−0.01	−0.03	−0.07	0	−0.04
HALLMARK_G2M_CHECKPOINT	−0.04	−0.06	0.03	−0.06	0.01	−0.02	−0.11	0.09	0.15	−0.08	−0.15
HALLMARK_GLYCOLYSIS	0	0.02	−0.04	−0.13	0.04	−0.03	−0.01	−0.13	−0.12	−0.12	−0.05
HALLMARK_HEDGEHOG_SIGNALING	0.26	0.26	0.14	0.11	0.01	0.21	0.12	−0.01	0.08	−0.05	−0.06
HALLMARK_HEME_METABOLISM	−0.05	0.12	−0.15	−0.14	0	−0.02	0.06	−0.11	0.02	−0.21	−0.13
HALLMARK_HYPOXIA	0.1	0.05	−0.04	−0.04	0.01	−0.14	0.07	−0.15	−0.16	−0.19	−0.26
HALLMARK_IL2_STAT5_SIGNALING	−0.16	−0.07	−0.09	−0.02	−0.14	−0.1	−0.24	−0.01	−0.01	0.09	0.1
HALLMARK_IL6_JAK_STAT3_SIGNALING	−0.27	−0.25	−0.16	−0.32	0.19	−0.12	0.03	−0.14	−0.05	−0.17	−0.14

Continued

Hallmark pathway	SNU-2337A-CO	SNU-2337B-CO	SNU-2536B-CO	SNU-2536C-CO	SNU-2600A-CO	SNU-2600B-CO	SNU-2600C-CO	SNU-3546A-CO	SNU-3546B-CO	SNU-5455-TO	SNU-5455L-TO
HALLMARK_INFLAMMATORY_RESPONSE	−0.07	−0.09	0.05	−0.23	0.01	−0.18	0.02	−0.19	−0.13	−0.13	−0.11
HALLMARK_INTERFERON_ALPHA_RESPONSE	−0.28	−0.38	−0.33	−0.34	−0.26	−0.37	−0.33	−0.13	−0.15	−0.35	−0.32
HALLMARK_INTERFERON_GAMMA_RESPONSE	−0.32	−0.33	−0.32	−0.3	−0.31	−0.44	−0.38	−0.18	−0.21	−0.23	−0.25
HALLMARK_KRAS_SIGNALING_DN	−0.07	−0.13	0.02	−0.15	−0.09	−0.25	−0.17	−0.05	−0.17	−0.02	−0.09
HALLMARK_KRAS_SIGNALING_UP	−0.26	−0.1	0.03	−0.05	−0.08	−0.14	0.01	−0.09	−0.05	−0.13	−0.16
HALLMARK_MITOTIC_SPINDLE	−0.06	−0.01	−0.04	−0.01	0	−0.06	−0.14	−0.1	−0.16	−0.04	−0.14
HALLMARK_MTORC1_SIGNALING	−0.15	−0.19	−0.1	0.15	−0.07	−0.03	−0.03	0.1	0.03	−0.13	−0.14
HALLMARK_MYC_TARGETS_V1	−0.27	−0.12	−0.18	−0.17	−0.15	−0.1	−0.09	−0.04	−0.22	−0.1	−0.15
HALLMARK_MYC_TARGETS_V2	−0.22	−0.15	−0.2	−0.16	−0.12	0.11	0.06	0.24	−0.12	0.1	−0.12
HALLMARK_MYOGENESIS	0.11	−0.19	−0.1	−0.06	−0.1	−0.23	0.11	−0.13	0.03	−0.07	−0.14
HALLMARK_NOTCH_SIGNALING	0.11	0.18	−0.02	0.05	−0.41	−0.13	−0.12	0.12	−0.14	−0.23	−0.14
HALLMARK_OXIDATIVE_PHOSPHORYLATION	−0.05	−0.07	−0.09	−0.15	−0.18	−0.17	−0.1	−0.06	−0.14	−0.07	−0.07
HALLMARK_P53_PATHWAY	−0.14	−0.11	−0.01	0.1	−0.15	−0.22	−0.17	−0.09	−0.12	−0.02	−0.15

Continued

Hallmark pathway	SNU-2337A-CO	SNU-2337B-CO	SNU-2536B-CO	SNU-2536C-CO	SNU-2600A-CO	SNU-2600B-CO	SNU-2600C-CO	SNU-3546A-CO	SNU-3546B-CO	SNU-5455-TO	SNU-5455L-TO
HALLMARK_PANCREAS_BETA_CELLS	0.05	0.05	−0.37	−0.02	−0.36	0	0.17	−0.26	−0.09	−0.24	−0.3
HALLMARK_PEROXISOME	−0.29	−0.36	0.06	0.21	0.14	0.04	0.14	−0.19	−0.12	−0.12	−0.03
HALLMARK_PI3K_AKT_MTOR_SIGNALING	−0.16	−0.06	−0.15	−0.21	−0.23	−0.17	−0.25	−0.22	−0.06	−0.1	−0.06
HALLMARK_PROTEIN_SECRETION	0.09	0.16	−0.01	−0.05	−0.17	−0.23	0.15	0.02	−0.15	0.1	−0.05
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY	0.02	0.01	−0.13	−0.2	−0.13	−0.18	−0.2	−0.09	−0.13	−0.42	−0.38
HALLMARK_SPERMATOGENESIS	−0.12	−0.12	0.18	0.13	0.05	−0.07	−0.11	−0.06	0	−0.17	−0.01
HALLMARK_TGF_BETA_SIGNALING	−0.24	0.02	−0.32	0.02	−0.25	−0.18	0.05	−0.09	−0.03	0.22	0.25
HALLMARK_TNFA_SIGNALING_VIA_NFKB	−0.04	−0.05	−0.1	−0.12	−0.08	−0.2	−0.16	−0.21	−0.23	−0.15	−0.11
HALLMARK_UNFOLDED_PROTEIN_RESPONSE	−0.02	−0.01	−0.06	−0.05	−0.1	−0.13	−0.11	0.03	−0.02	−0.1	0.04
HALLMARK_UV_RESPONSE_DN	−0.18	−0.05	−0.16	−0.23	−0.02	0.01	−0.01	−0.15	−0.03	0.03	−0.03
HALLMARK_UV_RESPONSE_UP	−0.29	−0.28	−0.07	−0.2	−0.02	−0.12	−0.1	−0.35	−0.21	−0.26	−0.11
HALLMARK_WNT_BETA_CATENIN_SIGNALING	−0.19	−0.27	0.03	−0.02	−0.19	−0.08	−0.07	0.03	0.06	−0.32	0.01
HALLMARK_XENOBIOTIC_METABOLISM	−0.15	−0.16	−0.25	−0.1	−0.08	−0.11	0.01	−0.09	−0.16	−0.1	−0.24

Figure 10. Heatmap of normalized enrichment score (NES) by gene set variation analysis (GSVA).

Pathways were categorized and marked by each representative color. The range of enrichment scores is −4.8 ~ 3.9 (purple, enriched; green, depleted).

Exosomal miRNA Profiling and functional characterization of exosomal miRNA in CLM

Differentially expressed miRNAs (DEMs) analysis was conducted on 11 CLM organoids using miRNA microarray data. The exosomal miRNA was extracted from culture media, and it was confirmed that there was no cell lysate contamination by assessing the presence of CD63 and calnexin. CD63, an essential component of exosomal membranes from the tetraspanin family, was utilized as a positive marker for validating exosomes. The expression of CD63 was observed in the media of all organoid media (Figure 11), indicating successful extraction of exosomes from the organoid culture media. Calnexin, a protein expressed in the endoplasmic reticulum, was not detected in any of the exosome media, confirming the absence of protein or vesicle contamination from other compartments [34].

A total of 6,609 Homo sapiens–specific miRNAs were detected from the microarray. Among the captured probes, miRNAs detected in both pCRC and LM were selected for further analysis in each group. The fold change was calculated as $\log_2$[Normalized expression of LM PDOs/Normalized expression of pCRC PDOs]. The volume represents the detection intensity and is defined as the geometric mean of expression values between the two groups (Table 6). In

cases where the fold change is the same, a probe with higher volume is considered more reliable than a probe with lower volume. The top 20 miRNAs (|FC| ≥ 3, top 20 volume) in each group were identified as significant miRNAs that potentially influence the metastasis process (Figure 12A-E; Table 6).

To gain further insights into DEMs, the target genes of the top 20 DEMs were predicted using miRWalk 3.0. Subsequently, the target genes were cross-validated from the actual RNA-seq data (Figure 1).

Additionally, gene ontology (GO) analysis was performed on the common genes identified from each set. The Q-value cutoff was set to 0.05. The resulting biological processes associated with the common genes in each set are depicted in the cnet plot (Figure 13A-E). In the case of SNU-2337B-CO, the common genes were mainly related to lipid metabolism and localization (Figure 13A). In SNU-2536C-CO, the common genes were highly associated with the ERK1 and ERK2 cascades and cell morphogenesis processes (Figure 13B). The common genes identified in the liver metastatic type of SNU-2600 were mainly involved in several organ development processes (Figure 13C). As for SNU-3546B-CO, the common genes were primarily associated with mesenchymal cell differentiation, cell

junction assembly, and skin development (Figure 13D). In particular, processes such as mesenchymal cell differentiation and cell junction assembly are involved in epithelial–mesenchymal transition (EMT) or mesenchymal–epithelial transition (MET), which facilitate the movement and colonization of primary cancer cells in a new environment during metastasis (Figure 13D) [35–39]. However, in the case of SNU–5455, no biological process met the Q–value criterion due to an insufficient number of target genes.

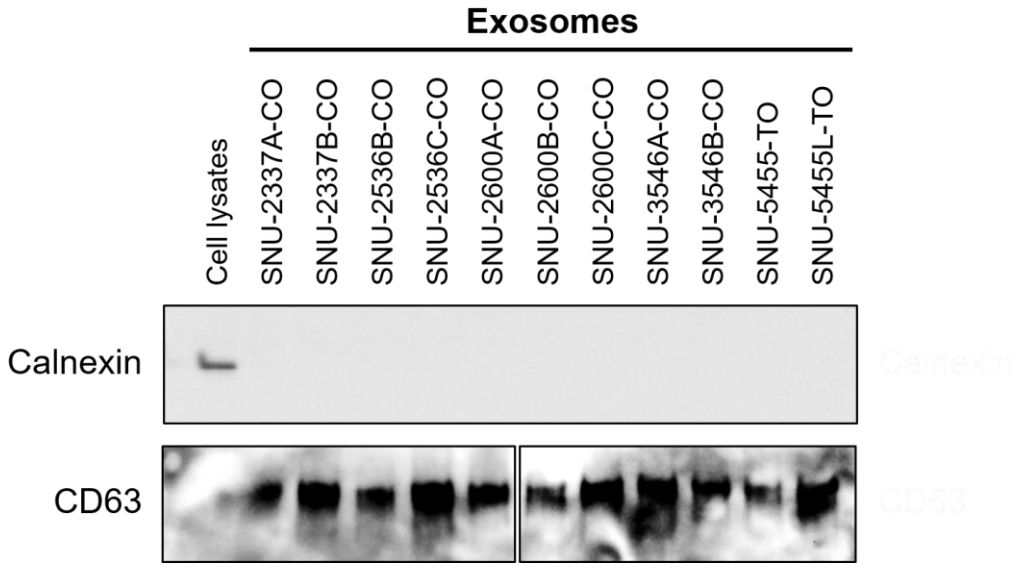

Figure 11. Expression of exosomes markers identified by western blot. CD63 was used as a positive marker for the exosome, and calnexin (an endoplasmic reticulum marker) was used as a negative marker. To confirm the contamination of cell lysate, cell lysate was used as an internal control.

Table 6. Top 20 DEMs of CLM organoids.

SNU-2337 set			SNU-2536 set			SNU-2600 set		
mirBase ID	Log2FC	Volume	mirBase ID	Log2FC	Volume	mirBase ID	Log2FC	Volume
hsa-miR-6090	-1.99	10.65	hsa-miR-320a	2.11	8.22	hsa-miR-4440	-2.49	8.72
hsa-miR-6089	-1.89	10.49	hsa-miR-320c	2.01	8.11	hsa-miR-6732-5p	-1.93	8.64
hsa-miR-4516	-1.69	9.4	hsa-miR-320b	2.16	8.11	hsa-miR-4743-5p	-2.32	8.52
hsa-miR-4743-5p	2.12	7.8	hsa-miR-320d	1.67	6.95	hsa-miR-1246	-3.37	8.34
hsa-miR-6732-5p	1.75	7.66	hsa-miR-4484	-1.66	6.84	hsa-miR-4793-3p	-2.15	8.3
hsa-miR-4440	2.07	6.94	hsa-miR-1290	1.95	6.2	hsa-miR-619-5p	-3.14	7.84
hsa-miR-4793-3p	1.68	6.63	hsa-miR-22-3p	2.43	5.82	hsa-miR-1273g-3p	-2.28	7.54
hsa-miR-619-5p	2.09	6.28	hsa-miR-23a-3p	1.93	4.99	hsa-miR-8075	-2.11	7.36
hsa-miR-7110-5p	-2.05	6.26	hsa-miR-361-5p	2.38	4.76	hsa-miR-3620-5p	-2.27	6.85
hsa-mir-320e	1.66	5.84	hsa-miR-4429	1.82	4.02	hsa-miR-1228-5p	-1.75	6.7
hsa-miR-4443	-1.84	5.47	hsa-miR-103a-3p	-2	3.75	hsa-miR-297	-2.93	6.51
hsa-miR-3663-3p	-1.65	5.39	hsa-miR-24-3p	2.39	3.48	hsa-mir-6869	-2.66	6.51
hsa-miR-1290	-2.21	4.93	hsa-miR-106a-5p	-1.84	3.45	hsa-miR-5093	-2.9	6.19
hsa-miR-4486	1.78	4.86	hsa-miR-107	-1.88	3.1	hsa-miR-1910-5p	-1.65	6.11
hsa-miR-6749-5p	-1.73	4.65	hsa-miR-320e	2.7	2.58	hsa-miR-6869-5p	1.92	6.08
hsa-miR-5093	2.99	4.53	hsa-miR-1273h-5p	1.79	2.52	hsa-miR-4486	-2.31	6.03
hsa-miR-4433b-3p	-2.16	4.28	hsa-miR-6879-5p	-2	2.45	hsa-miR-4507	-2.2	5.99
hsa-miR-4530	-2.29	4.12	hsa-miR-1273f	1.7	2.33	hsa-miR-92a-3p	-1.91	5.88
hsa-miR-4689	-2.03	3.94	hsa-miR-338-5p	1.69	2.32	hsa-miR-6756-5p	-1.59	5.82
hsa-miR-92b-5p	1.71	3.82				hsa-miR-1915-3p	2.13	5.69

Continued

SNU-3546 set			SNU-5455 set		
mirBase ID	Log2FC	Volume	mirBase ID	Log2FC	Volume
hsa-miR-4516	1.8	8.66	hsa-miR-7110-5p	2.03	5.66
hsa-miR-320a	-2.29	8.62	hsa-miR-6756-5p	1.69	5.47
hsa-miR-320c	-1.9	8.57	hsa-miR-6848-5p	1.59	5.36
hsa-miR-320b	-2.08	8.54	hsa-miR-6789-5p	2.02	5.15
hsa-miR-7110-5p	1.8	7.8	hsa-miR-2861	1.79	4.95
hsa-miR-6786-5p	1.9	7.48	hsa-miR-619-5p	2.98	4.87
hsa-miR-1273g-3p	-2.56	7.07	hsa-miR-320b	-2.36	4.46
hsa-miR-663a	2.36	7.01	hsa-miR-4707-5p	-2.18	4.32
hsa-miR-4440	-3.2	6.84	hsa-miR-4687-3p	-1.73	3.96
hsa-miR-4281	3.12	6.4	hsa-miR-4440	2.9	3.78
hsa-miR-6816-5p	1.9	6.23	hsa-miR-4529-3p	2.03	3.54
hsa-miR-4793-3p	-3.64	6.18	hsa-miR-6743-5p	1.99	3.49
hsa-miR-6724-5p	2.16	6.07	hsa-miR-548ac	-1.61	3.44
hsa-miR-1915-3p	1.96	5.93	hsa-miR-6722-3p	1.68	3.4
hsa-miR-619-5p	-2.34	5.86	hsa-miR-7150	-1.82	3.13
hsa-mir-4539	-1.95	5.81	hsa-miR-320d	-3.07	3
hsa-miR-4687-3p	3.02	5.56	hsa-miR-668-5p	-2.67	2.87
hsa-miR-1910-5p	-1.92	5.44	hsa-miR-4535	1.79	2.8
hsa-miR-4508	2.31	5.35			
hsa-miR-6749-5p	3.06	5.19			

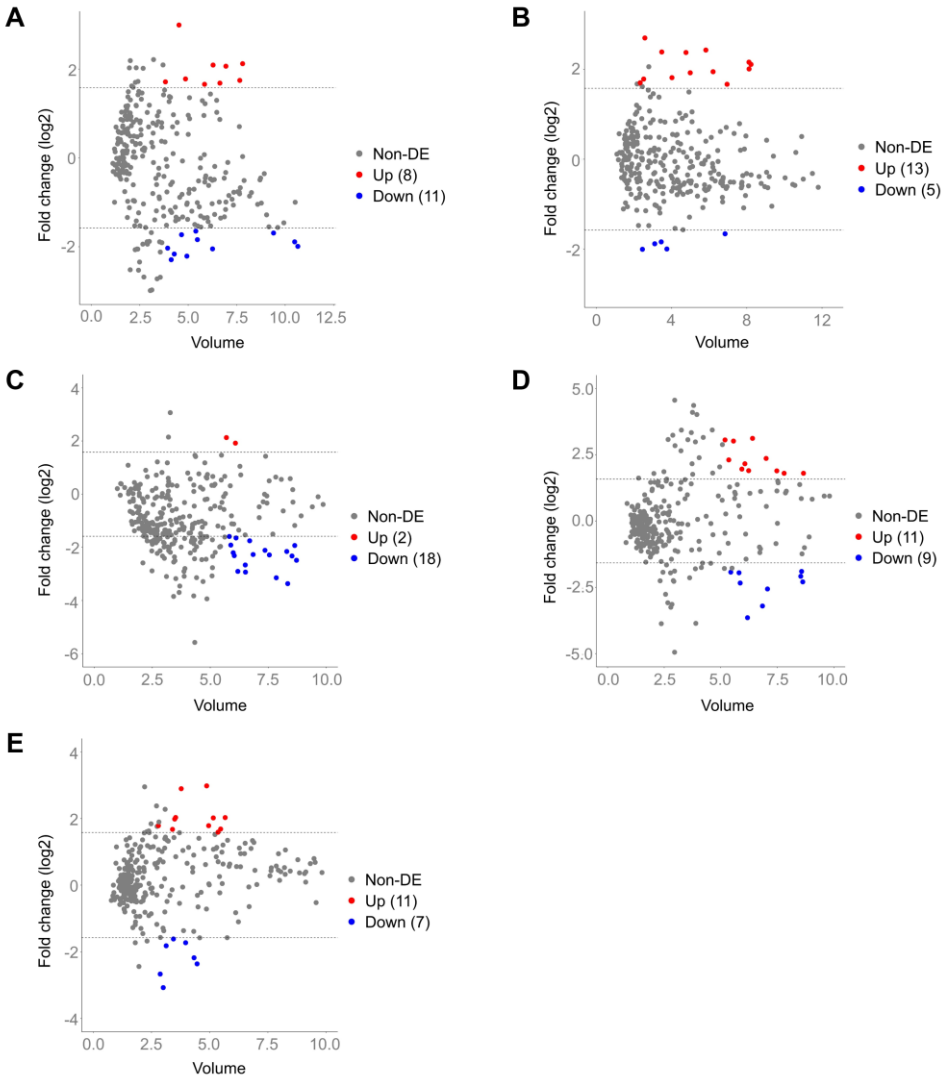

Figure 12A–E. Volume plots visualized differentially expressed miRNAs (DEMs) in LM organoids compared to pCRC organoids. x axis = volume. y axis = $\log_2$FC. While satisfying the cut-off (|FC| >3), the top 20 ranking probes in the order of the highest volume were marked with colored dots (red, up-regulated; blue, down-regulated; gray, non-DEM). High volume means high intensity in

microarray results.

A Volume plot of the SNU-2337 group. **B** Volume plot of the SNU-2536 group. **C** Volume plot of the SNU-2600 group. **D** Volume plot of the SNU-3546 group. **E** Volume plot of the SNU-5455 group

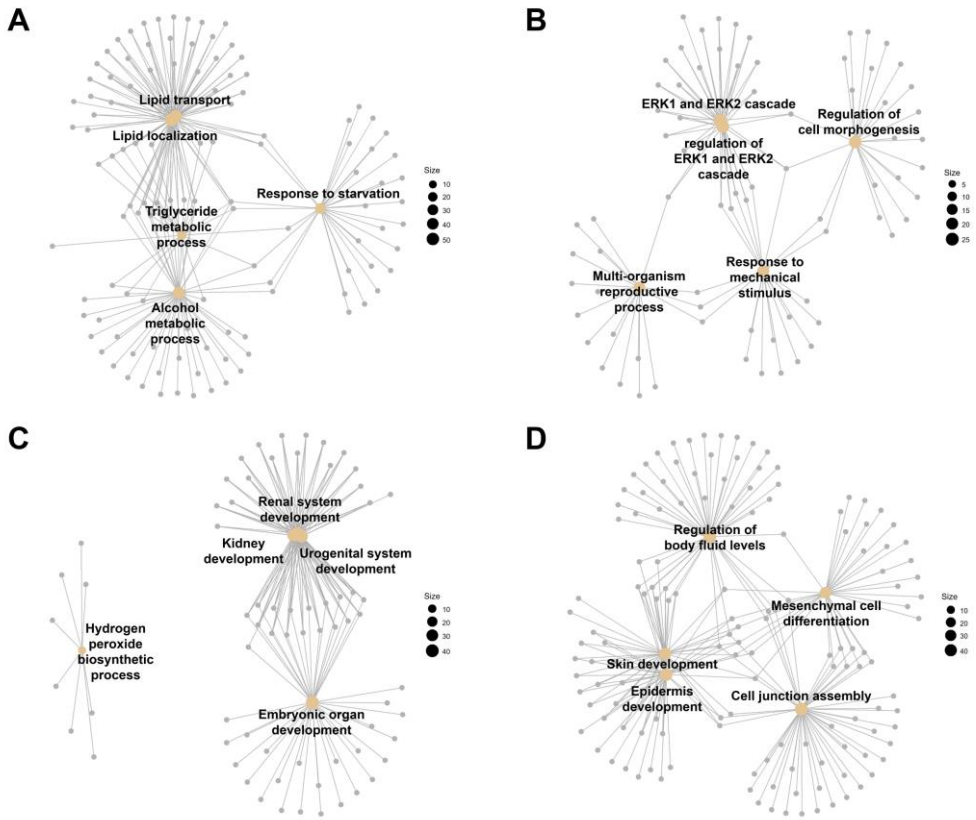

Figure 13A–D. Gene ontology (GO) analysis of common genes (miRNA target gene – DEGs). In the cnet plot, biological processes and genes related to each other are linked to the edges. The size of the GO term (circle) represents the associated genes. Cut-off = qvalue, 0.05.

A GO of the SNU-2337 group. **B** GO of the SNU-2536 group. **C** GO of the SNU-2600 group. **D** GO of the SNU-3546 group.

Integrated analysis of transcriptomic and pharmaco-omics data derived LM specific therapeutic targets

Twenty-one compounds library was designed for drug screening of the PDOs. Each drug was selected based on their typical usage in clinical practice as determined by the National Cancer Institute (NCI). The drug screening process was conducted at least three times in different passages. Drug sensitivity was assessed using the area under the curve (AUC). In my drug library, there are certain drugs that exert cytotoxic effects on cells such as TAS-102, Capecitabine, 5-FU, and Oxaliplatin. Therefore, it is challenging to represent the responsiveness of these drugs using IC50, which indicates the maximum concentration at which the activity of cells is reduced by half upon drug administration. Consequently, we graphically depicted the drug responsiveness as a function of drug concentration at six points and calculated the AUC beneath this curve. Through this approach, we compared the pharmacological efficacy among the drugs. The final drug AUC was calculated as the average of the technical replicates. A high AUC indicates resistance to drugs, and a low AUC means sensitivity to drugs. To categorize the drugs, a clustering analysis using the elbow and silhouette methods was performed (Figures 14A–C). Cluster 1 demonstrated a sensitive

response in most PDOs (Figure 14C; Table 7). Cluster 2 exhibited a heterogeneous response across the different PDOs and included the drugs Irinotecan, AZD-2014, MK-5108, SAHA, ICG-001, and Everolimus. The final cluster, cluster 3, generally demonstrated resistance in most PDOs.

Table 7. Drug AUC values of the 11 CLM organoids on 21 compounds.

Drug	Max Conc. (uM)	SNU-2337A-CO	SNU-2337B-CO	SNU-2536B-CO	SNU-2536C-CO	SNU-2600A-CO
Tas-102	100	1.59	1.74	1.39	1.85	1.83
Regorafenib	100	1.55	1.91	1.4	1.92	1.49
Capecitabine	1000	1.69	1.93	1.33	1.83	1.49
Apitolisib	50	0.72	0.87	0.8	0.95	0.7
Belinostat	100	0.58	0.78	0.47	0.55	0.69
Trametinib	50	0.96	0.58	0.43	0.56	0.68
Cyclopamine	50	1.84	1.87	1.52	2	1.61
ICG-001	100	1.54	1.07	0.95	1.04	1.01
Buparlisib	100	0.79	0.97	0.47	0.99	1.02
SAHA	50	0.71	1.41	1.02	1.5	1.31
Afatinib	50	0.49	0.67	0.26	1.2	1.07
AZD-2014	5	1.67	1.23	0.99	1.63	1.45
MK-5108	1000	1.54	1.37	1.06	1.5	1.23
Olaparib	50	1.77	1.48	1.45	1.57	1.73
Irinotecan	100	1.56	1.27	1.17	1.53	1.24
Fluorouracil	20000	0.59	0.79	0.65	0.62	0.87
Oxaliplatin	100	1.34	1.47	1.41	1.76	1.7
Dabrafenib	10	1.9	1.77	1.25	1.89	1.53
MK-2206	4	1.74	1.64	1.45	1.56	1.44
Cabozantinib	10	1.66	1.55	1.45	1.77	1.52
Everolimus	10	1.38	1.65	0.81	0.62	0.78

Continued

Drug	Max Conc. (uM)	SNU-2600B-CO	SNU-2600C-CO	SNU-3546A-CO	SNU-3546B-CO	SNU-5455-TO	SNU-5455L-TO
Tas-102	100	1.7	1.82	1.35	1.72	1.55	1.67
Regorafenib	100	1.43	1.34	1.41	1.81	1.56	1.32
Capecitabine	1000	1.64	1.53	1.5	1.56	1.43	1.56
Apitolisib	50	0.82	0.72	0.97	0.99	0.66	0.78
Belinostat	100	0.68	0.53	0.7	0.6	0.4	0.3
Trametinib	50	0.62	0.58	0.53	0.78	0.43	0.36
Cyclopamine	50	1.79	2.08	2.2	1.82	1.77	1.71
ICG-001	100	1.09	1.27	1.08	0.92	1.1	1.39
Buparlisib	100	0.89	0.77	1	0.93	0.73	1.03
SAHA	50	1.78	1.28	1.34	1.07	1.07	1.1
Afatinib	50	1.13	0.58	0.57	0.94	1.01	1.04
AZD-2014	5	1.59	1.56	1.36	1.43	1.4	1.41
MK-5108	1000	1.3	1.32	1.52	1.4	1.3	1.38
Olaparib	50	1.57	1.5	1.72	1.78	2.02	1.9
Irinotecan	100	1.43	1.51	1.44	1.71	1.58	1.16
Fluorouracil	20000	0.93	0.57	0.75	0.85	0.75	0.53
Oxaliplatin	100	2	1.55	1.84	1.43	1.31	1.66
Dabrafenib	10	1.63	1.37	1.91	1.51	1.41	1.4
MK-2206	4	1.8	1.38	1.64	1.57	1.16	1.08
Cabozantinib	10	1.8	1.57	1.7	1.48	1.12	1.48
Everolimus	10	1.06	1.33	1.12	0.99	1.1	0.88

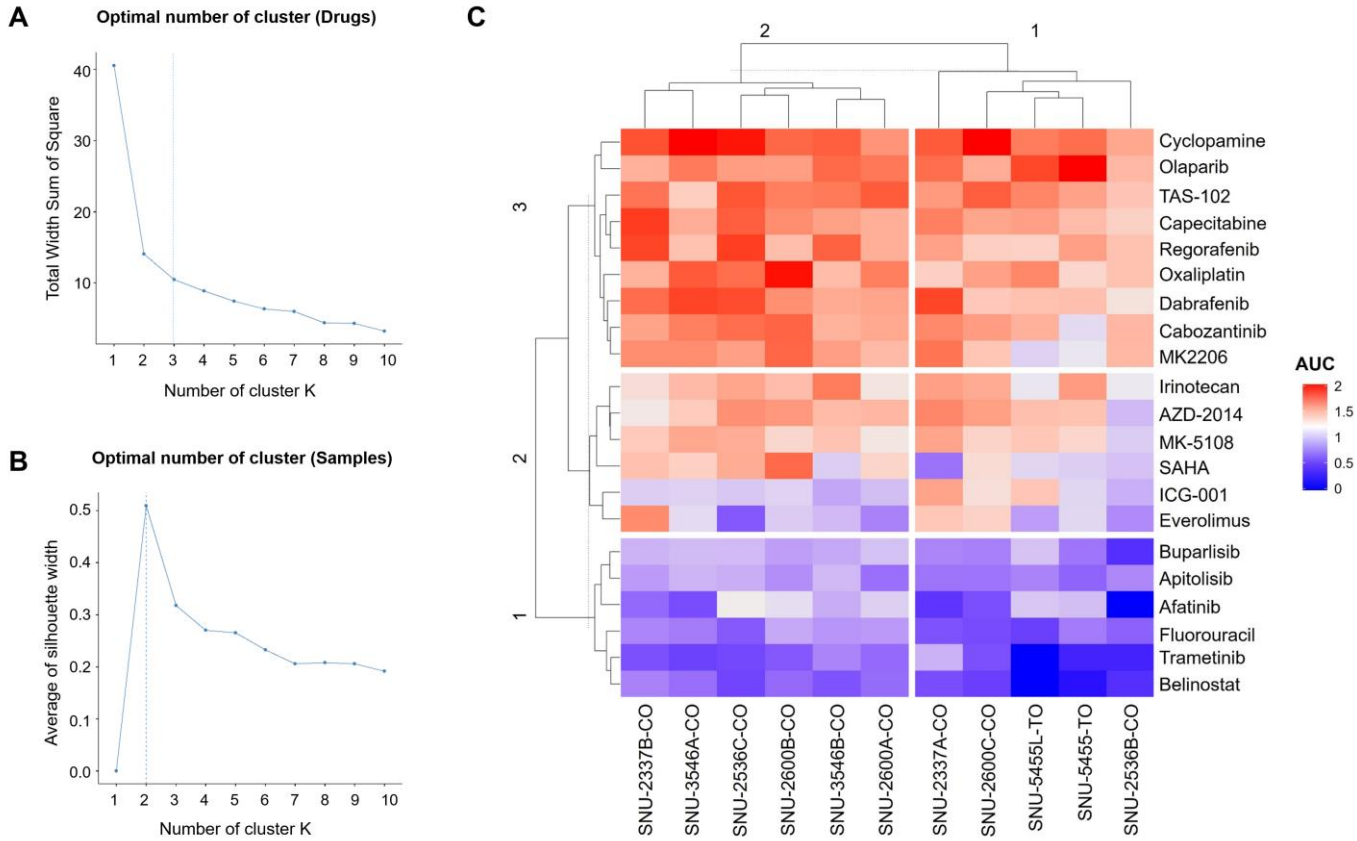

Figure 14A–C. Heatmap of area under the curve (AUC) values represented the sensitivity of 21 compounds on CLM organoids.

A Elbow plot. B Silhouette plot. C Drugs and samples were clustered using the elbow method and the silhouette method. The range of AUC is 0.26 ~ 2.20 and the drug response was distinguished by different colors (blue, sensitive; red, resistant).

To further investigate the correlation among miRNA expression, mRNA expression, and drug response, DIABLO method of mixOmics package in R was used for data integration. Since DIABLO is a feature extraction method based on supervised learning, I designated pCRC cancer as dependent variable 1 and LM cancer as dependent variable 2. From each normalized dataset, 319 mRNA expressions, 296 miRNA expressions, and twenty-one drug AUC values were selected and used as input data for DIABLO.

To ensure distinct correlations, R-squared value was set to 0.825. Using the circos plot, I confirmed significant correlations within the specified cut-off range, including miR-3613-5p - *eEF2K* - TAS-102 and miR-663b - *HTD2* - Apitolisib (Figure 15). The AUC value of TAS-102 exhibited a negative correlation with both *eEF2K* gene expression and miR-3613-5p. Additionally, *eEF2K* showed a positive correlation with miR-3613-5p. Furthermore, *HTD2* and miR-663b demonstrated a negative correlation in Apitolisib sensitivity, while *HTD2* and has-miR-663b exhibited a positive correlation.

To elucidate the association among miR-3613-5p, *eEF2K*, and TAS-102, the target genes of miR-3613-5p were predicted using the mirWalk database. Among the target genes, *EIF4B*, *POLDIP3*, and

RPL37 were found to interact with *eEF2K* in a protein-protein interaction (PPI) network (Figure 16).

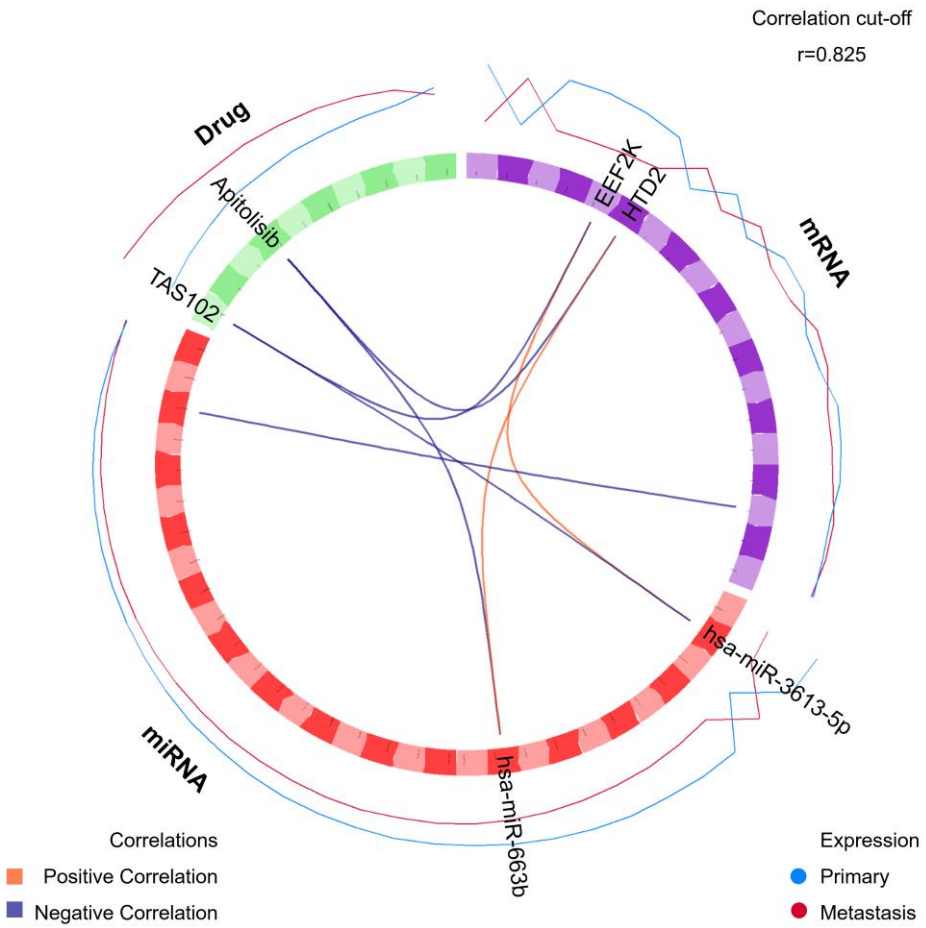

Figure 15. Circos plot of DIABLO analysis. Normalized mRNA expression, miRNA expression, and drug AUC values were used as input data for N-integration (purple, mRNA; red, miRNA; light green, drug). The cutoff was set at r = 0.825 for obtaining clear correlations. The correlation between each factor was connected by lines (orange, positive correlation; deep blue, negative correlation).

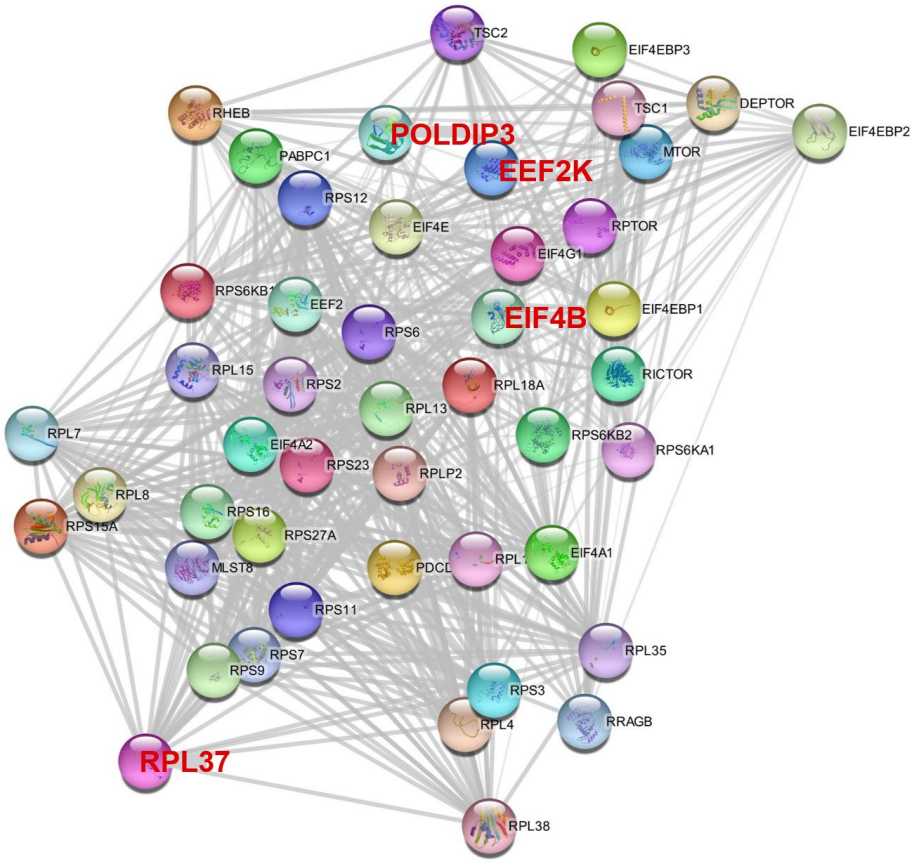

Figure 16. Protein–protein interaction (PPI) network between *eEF2K* and target genes of miR–3613–5p. Each circle represents a gene. The genes interact with each other and are connected by edges. Significant genes (*eEF2K* and target genes of miR–3613–5p) are highlighted in bold font.

Despite encountering challenges due to the inter patient heterogeneity, I was able to identify bio-signatures that effectively clustered the variables of pCRC and LM within my cohort. Using fifteen factors, the eleven CLM PDOs were categorized into pCRC and LM groups (Figure 17). It was observed that TAS-102, Buparilisib, SAHA, Apitolisib, and Capecitabine exhibited relative resistance in LM. Additionally, significant differences in the expression patterns of miR-541-5p, miR-3613-5p, miR-6800-5p, miR-6789-5p, miR-221-3p, *DEGS2*, *ADAMTS16*, *EBF4*, *KIF25*, and *eEF2K* were observed between pCRC and LM.

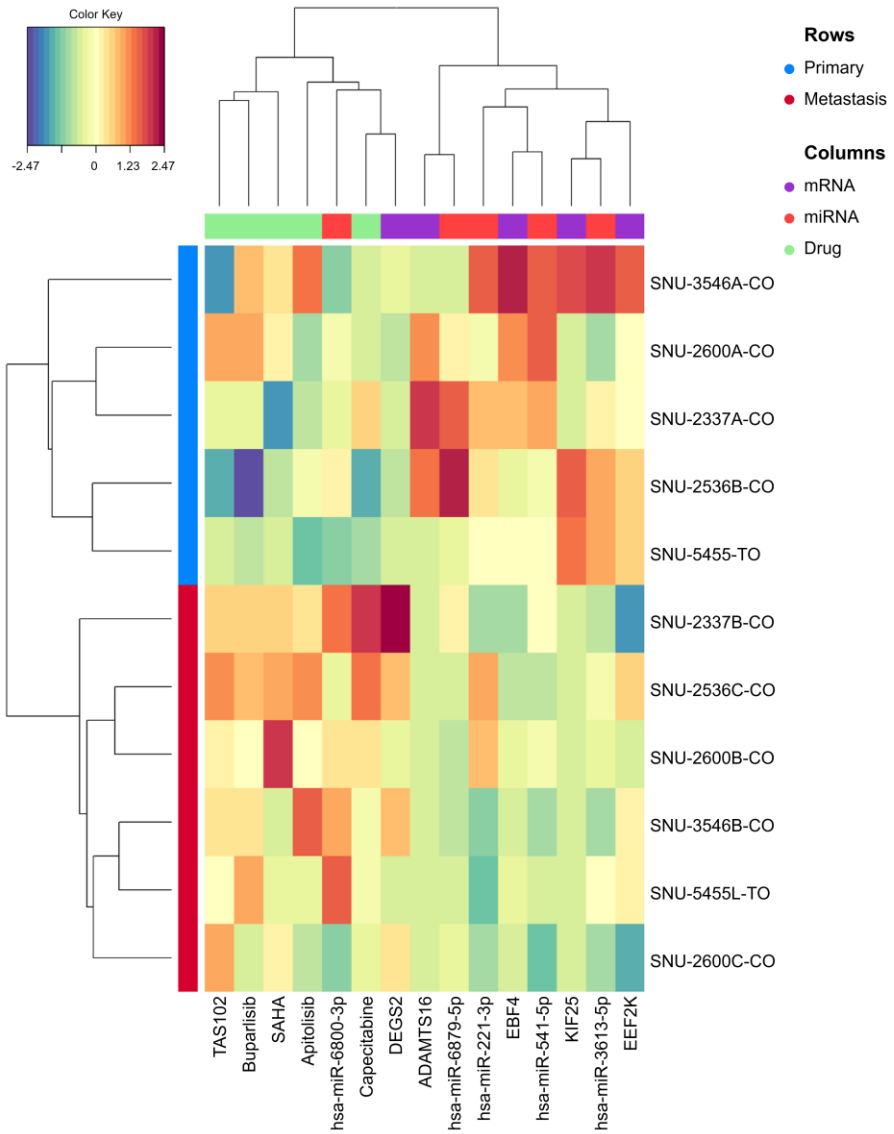

Figure 17. Clustered image map (CIM) of DIABLO analysis. Among the input values (319 mRNA expression, 296 miRNA expression, and 21 drug AUC values), 15 factors that were assigned high weight to explain the dependent variable were represented on CIM. Cancer types were well clustered by those variables.

Building regression models indicates actionable molecular targets connecting KRAS signaling to Regorefenib and Apitolisib

It is significant to apply the dataset to predicting the metastatic potency of the pCRC. I built a decision tree model utilizing a multi-omics dataset along with the responses of twenty-one clinically relevant drugs. The Decision Tree package in the R program was employed to standardize the patterns of multiple numeric values, incorporating one hundred and two predictors comprising drug response in terms of AUC, pathway enrichment scores, and miRNA expression.

KRAS signaling down pathway and the response to Regorafenib were the most statistically significant factors for predicting LM in pCRC (Figure 18). Down-regulation of the KRAS signaling down pathway (< -0.112) indicated the highest possibility of LM. Here, the HALLMARK KRAS signaling down pathway refers to a gene set that is down-regulated upon KRAS activation. Generally, a low NES for this pathway was associated with LM (4 out of 5 cases). This implies that there is less expression of down-regulated genes when *KRAS* is activated in LM compared to pCRC. On the other hand, when the NES of HALLMARK KRAS signaling down pathway was up-regulated (≥ -0.112), the response to regorafenib became a

determining factor for the potency of LM.

Furthermore, I identified the top 15 variables for LM prediction using lares in the R package. Least Angle Regression (LARS) is an automated machine learning (autoML) model that utilizes the LARS algorithm for optimal variable selection, model building, and handling of multi-collinearity. One hundred and two factors used in the decision tree analysis were equally selected as input values. Among these variables, TGF beta signaling emerged as the most influential factor in discriminating LM within my cohort (Figure 19). Previous studies have already highlighted the role of late stage TGF beta signaling in promoting tumor invasion or metastasis [40]. This finding further validated the efficacy of my model in representing CLM. KRAS signaling down, which was selected as a splitter for predicting type in the decision tree, appeared as the second most relevant variable, accounting for 17% of autoML respondents (Figures 18 and 19). Regorafenib played a role in determining the difference between pCRC and LM, which are not classified by KRAS signaling down. Regorafenib also had a high probability of predicting LM. Apitolisib and miR-3613-5p, which exhibited significant correlations in DIABLO analysis, ranked as the fourth and fifteenth most relevant variables, respectively (Figures 15 and 19).

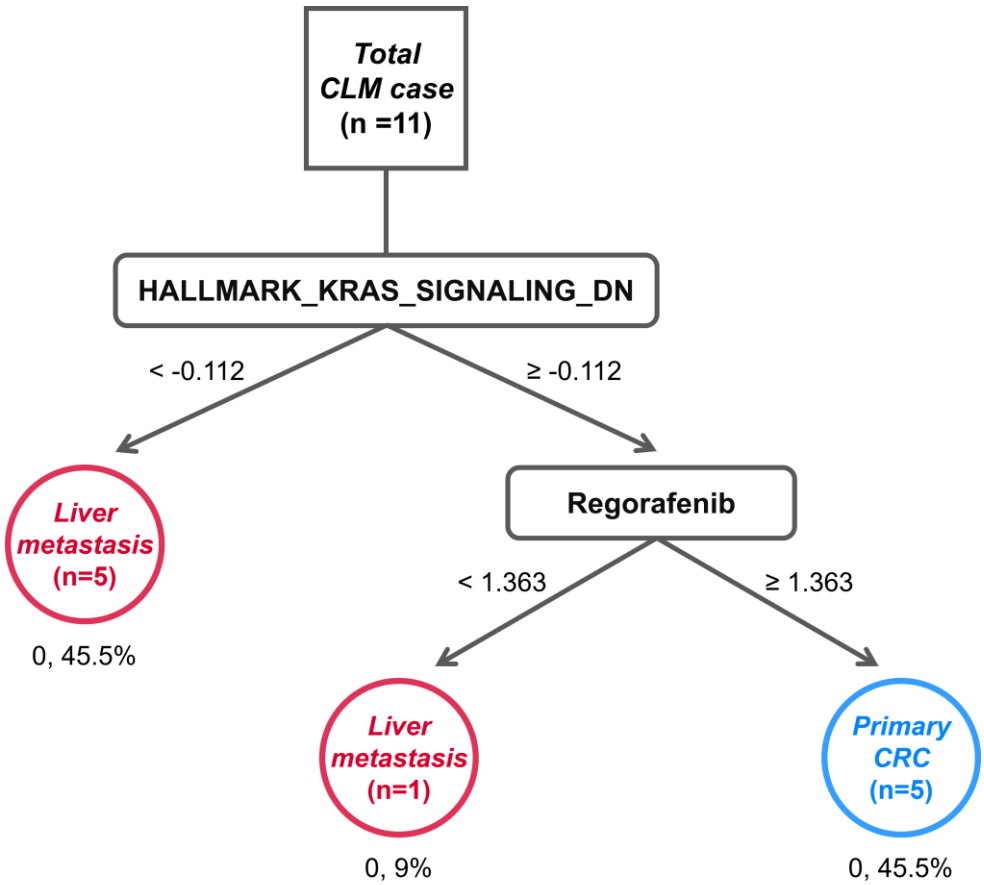

Figure 18. Decision tree using eleven CLM organoids. HALLMARK_KRAS_SIGNALING_DN pathway and Regorafenib were estimators capable of distinguishing LM from pCRC. The NES of the KRAS signaling down (−0.112) and the AUC value of Regorafenib (1.363) were the criteria for classification. Numerical values under each circle (nodes) represent impurities and percentages of the total.

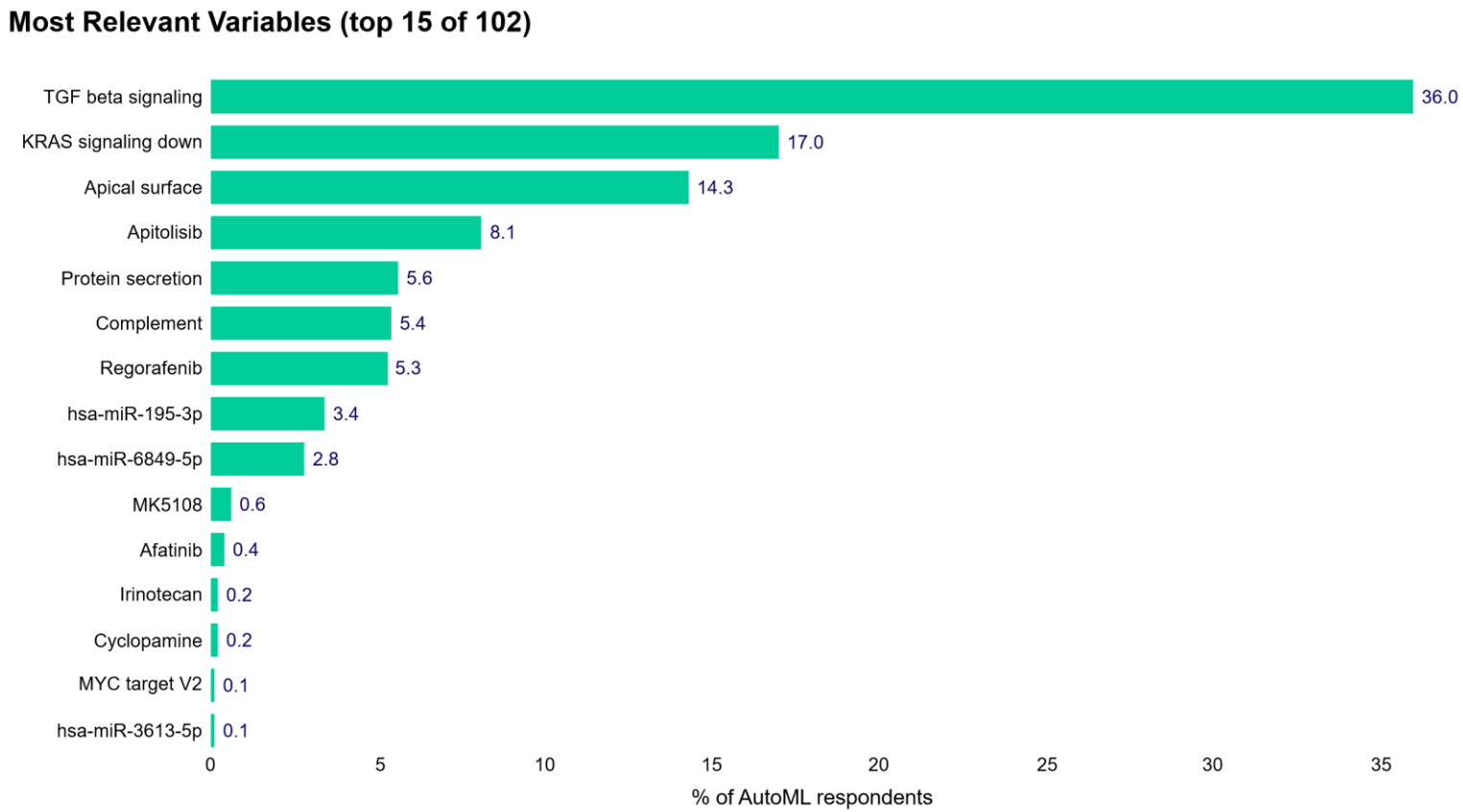

Figure 19. Top 15 most LM-relevant variables. The top 15 variables out of 102 variables were shown. TGF beta signaling pathway accounted for 36% of AutoML respondents, making it the most relevant variable.

Discussion

Despite extensive considerations regarding chemotherapy and precision medicine for CRC, metastasis remains a significant challenge in treatment. Due to the liver's role as a major blood filter in the body, cancer cells that separate from the primary tumor often travel through the bloodstream and settle in the liver, making it the most common site of metastasis for colon cancer [4]. Consequently, anatomical factors such as blood flow play a crucial role in the occurrence of distant metastases, particularly liver metastases. Exosomes present in the bloodstream can transport various molecules that contribute to the tumor microenvironment and regulate transcription. Exosomal miRNAs are believed to facilitate the process of primary cancer cells settling or proliferating in the liver through the bloodstream. Therefore, I noted exosomal miRNAs as potential drivers of CLM.

However, there was a lack of 3D models available to study the mechanism of CLM. To overcome these limitations, I developed the CLM PDCO by combining in-house 3D culture method with pre-existing cell lines. By utilizing these *in-vitro* models, I was able to overcome the time constraints associated with establishing pCRC and

matched simultaneously, while still benefiting from organoids' ability to better simulate the complexity of living organism. The cell lines used in organoid production were selected at early passages to ensure they retained the genetic characteristics of the tumor tissues as closely as possible.

Herein, I successfully established 11 CLM PDOs and PDCOs by constructing a biobank. I performed comprehensively profiling of CLM organoids characteristics using NGS and a HTS strategy. Then, I focused on a multi-omics approach, examining transcriptome, miRNA expression, and drug response.

There was a notable correlation between miR-3613-5p (miRNA expression), *eEF2K* (mRNA expression), and TAS-102 (drug response). Eukaryotic elongation factor 2 (*eEF2*) kinase (*eEF2K*), also known as calmodulin-dependent protein kinase III (*CAMKIII*), is a gene that regulates *eEF2* protein synthesis by inducing phosphorylation. The *eEF2* gene facilitates codon-codon translocation at the ribosome during protein synthesis, allowing transcription to be continued [41]. Previous studies has indicated that elevated levels of *eEF2K* expression enhance invasion and migration, leading to metastasis [42]. I observed that *eEF2K* was highly expressed in pCRC compared to LM (Figure 11). This

suggests that elevated *eEF2K* expression may trigger the metastatic potential of primary cancer.

EIF4B, *POLDIP3*, and *RPL37* were identified as target genes of miR-3613-5p that interact with *eEF2K* (Figure 12). Previous studies have implicated mammalian *POLDIP3* and *EIF4B* in the regulation of the mTOR signaling pathway[43, 44]. *RPL37* is a ribosomal component involved in mRNA translation like *eEF2K*. there is evidence suggesting that eEF2K activity can be modulated by the mTOR signaling pathway [45]. Consequently, it is inferred that miR-3613-5p is linked to *eEF2K* and participates in protein synthesis and the mTOR cascade.

TAS-102 is an oral drug composed of trifluridine (TFD) and tipiracil hydrochloride (TPI). TFD is the key compound of TAS-102 acting as a binding agent to DNA rather than a thymidine-based nucleoside to show an anti-tumor effect. TPI is an auxiliary substance that maximizes TPI efficacy while inhibiting the TFI degradation enzyme [46]. Although there is currently no research specifically investigating the resistance of TAS-102 related to *eEF2K* or miR-3613-5p. It is anticipated that *eEF2K* and miR-3613-5p may impact cancer-related pathways, leading to resistance against TAS-102 in LM.

Preclinical studies have demonstrated significant anti-tumor activity of TAS-102 in mCRC groups resistant to 5-FU, making it a promising option for mCRC treatment. [47, 48]. However, in my study, TAS-102 exhibited resistance in LM compared to pCRC. Further research appears to be necessary to understand why TAS-102 exhibits resistance in LM organoids, aiming to find promising anti-cancer drugs for 5-FU resistant CLM. Additionally, in-silico analysis alone may have limitations in studying miRNAs that function as transcription factors. Therefore, additional functional studies are needed to explore the relationship between miR-3613-5p, *eEF2K*, and TAS-102.

Through the prediction model and N-integration analysis, I identified significant variables that differentiate pCRC from LM. KRAS signaling down, Apitolisib, and Regorafenib emerged as important factors. In LM, Hallmark KRAS signaling down was found to be suppressed compared to pCRC (Table 7). Regorafenib, an FDA-approved multi-kinase inhibitor targeting *RAF*, acts as a substrate for various molecules, thereby exerting an anti-tumor effect [49]. It can be inferred that the up-regulated KRAS affects downstream proteins of the mTOR and MAPK pathways, leading to resistance to Regorafenib and Apitolisib resistant (Figure 20). In addition, I note that *KRAS* and

the gene set which is downregulated when KRAS signaling is activated can be the putative molecular target of CLM (Table 8). Moreover, KRAS wild-type (WT) patients have shown a significant extension in overall survival when treated with Regorafenib. However, in the case of most colorectal cancer patients with KRAS mutations, relatively minimal benefits have been reported [50]. There have been reports suggesting that the activation of KRAS influences the response to Regorafenib. Therefore, considering the influence of KRAS is essential when targeting LM with Regorafenib.

Understanding the mechanisms underlying changes in drug response is important for the treatment development. My study holds importance in utilizing the organoid model to efficiently screen various drugs and establish correlations between gene expression, pathway, and miRNA expression. Moreover, I derived a potential prediction system from these findings.

Indeed, due to the lack of cohort numbers, previous statistical analysis could not be convinced that there was a clear correlation between the tumor type and molecular features. Because my study has the specificity of focusing on CLM, there was insufficient public data for result validation. Nevertheless, I employed supervised learning methods to enhance the impact of the study and identify drug

targets specific to CLM. This approach could provide a basis for effective drug prescription or chemotherapy development for unresectable CLM patient. However, it is crucial to validate my results accurately through further research.

In summary, liver metastasis remains a significant therapeutic challenge in colorectal cancer, and there has been a scarcity of *in-vitro* models for studying CLM. My study provides a comprehensive understanding of the molecular features and pharmacological responses of LM compared to pCRC, achieved through the establishment of PDOs. As a result, PDOs have proven to be valuable preclinical models for investigating anti-cancer targets and building prediction system. Furthermore, it is meaningful that exosomes were integrated into organoids to expand the scope of organoids' use as preclinical model.

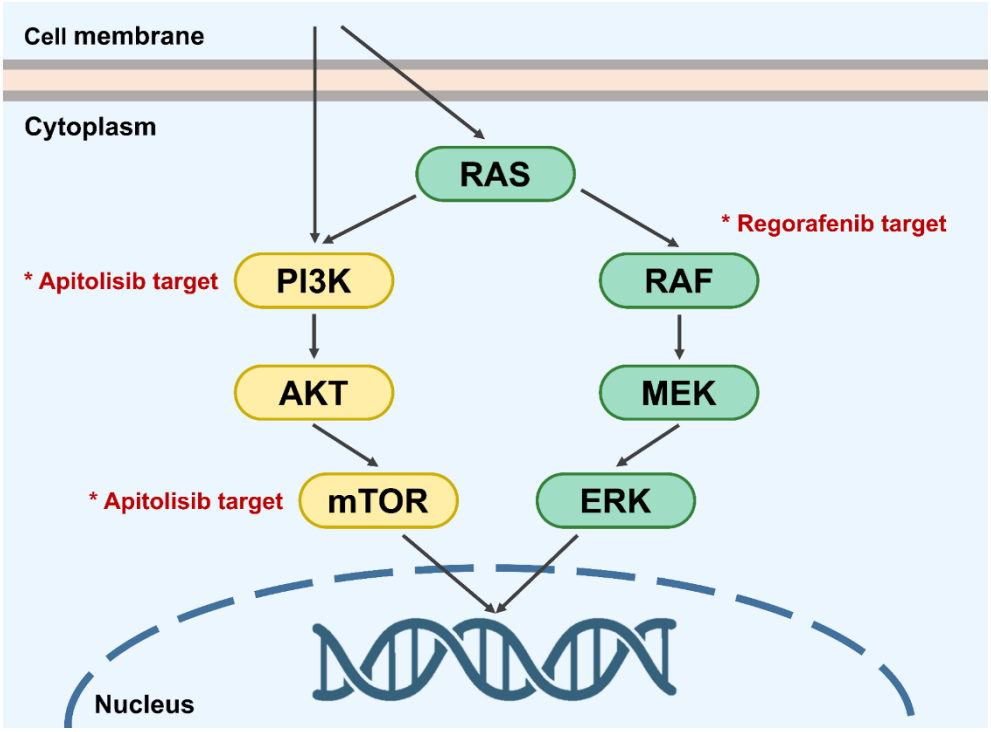

Figure 20. Molecular targets of Regorafenib and Apitolisib. The sharp arrow (→) means stimulation while blunt arrow (⊥) indicates inhibition.

Table 8. Gene set of HALLMARK_KRAS_signaling_dn pathway.

Human Gene Set: HALLMARK_KRAS_SIGNALING_DN									
ABCB11	CALML5	COL2A1	FGFR3	IL12B	LFNG	NRIP2	RGS11	SLC5A5	TFCP2L1
ABCG4	CAMK1D	COPZ2	FGGY	IL5	LGALS7	NTF3	RIBC2	SLC6A14	TFF2
ACTC1	CAPN9	COQ8A	FSHB	INSL5	LYPD3	NUDT11	RSAD2	SLC6A3	TG
ADRA2C	CCDC106	CPA2	GAMT	IRS4	MACROH2A2	OXT	RYR1	SMPX	TGFB2
AKR1B10	CCNA1	CPB1	GDNF	ITGB1BP2	MAGIX	P2RX6	RYR2	SNCB	TGM1
ALOX12B	CCR8	CPEB3	GP1BA	ITIH3	MAST3	P2RY4	SCGB1A1	SNN	THNSL2
AMBN	CD207	CYP11B2	GP2	KCND1	MEFV	PAX3	SCN10A	SOX10	THRB
ARHGDIG	CD40LG	CYP39A1	GPR19	KCNE2	MFSD6	PAX4	SELENOP	SPHK2	TLX1
ARPP21	CD80	DCC	GPR3	KCNMB1	MSH5	PCDHB1	SERPINA10	SPRR3	TNNI3
ASB7	CDH16	DLK2	GPRC5C	KCNN1	MTHFR	PDCD1	SERPINB2	SPTBN2	TSHB
ATP4A	CDKAL1	DTNB	GRID2	KCNQ2	MX1	PDE6B	SGK1	SSTR4	UGT2B17
ATP6V1B1	CELSR2	EDAR	GTF3C5	KLHDC8A	MYH7	PDK2	SHOX2	STAG3	UPK3B
BARD1	CHRNG	EDN1	HNF1A	KLK7	MYO15A	PKP1	SIDT1	SYNPO	VPREB1
BMPR1B	CHST2	EDN2	HSD11B2	KLK8	MYOT	PLAG1	SKIL	TAS2R4	VPS50
BRDT	CKM	EFHD1	HTR1B	KMT2D	NGB	PNMT	SLC12A3	TCF7L1	WNT16
BTG2	CLDN16	EGF	HTR1D	KRT1	NOS1	PRKN	SLC16A7	TCL1A	YBX2
C5	CLDN8	ENTPD7	IDUA	KRT13	NPHS1	PRODH	SLC25A23	TENM2	YPEL1
CACNA1F	CLPS	EPHA5	IFI44L	KRT15	NPY4R	PROP1	SLC29A3	TENT5C	ZBTB16
CACNG1	CLSTN3	FGF16	IFNG	KRT4	NR4A2	PTGFR	SLC30A3	TEX15	ZC2HC1C
CALCB	CNTFR	FGF22	IGFBP2	KRT5	NR6A1	PTPRJ	SLC38A3	TFAP2B	ZNF112

Reference

1. Jung, K.W., et al., *Prediction of Cancer Incidence and Mortality in Korea, 2022.* Cancer Res Treat, 2022. **54**(2): p. 345-351.

2. Ma, J., et al., *Bioinformatic analysis reveals an exosomal miRNA-mRNA network in colorectal cancer.* BMC Med Genomics, 2021. **14**(1): p. 60.

3. Keum, N. and E. Giovannucci, *Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies.* Nat Rev Gastroenterol Hepatol, 2019. **16**(12): p. 713-732.

4. Valderrama-Treviño, A.I., et al., *Hepatic Metastasis from Colorectal Cancer.* Euroasian J Hepatogastroenterol, 2017. **7**(2): p. 166-175.

5. Diener, M.K. and S. Fichtner-Feigl, *Biomarkers in colorectal liver metastases: Rising complexity and unknown clinical significance?* Ann Gastroenterol Surg, 2021. **5**(4): p. 477-483.

6. Liu, J., et al., *Epigenetic Alternations of MicroRNAs and DNA Methylation Contribute to Liver Metastasis of Colorectal Cancer.* Dig Dis Sci, 2019. **64**(6): p. 1523-1534.

7. Misiakos, E.P., N.P. Karidis, and G. Kouraklis, *Current treatment for*

colorectal liver metastases. World J Gastroenterol, 2011. **17**(36): p. 4067-75.

8. Adam, R., et al., *Rescue surgery for unresectable colorectal liver metastases downstaged by chemotherapy: a model to predict long-term survival.* Ann Surg, 2004. **240**(4): p. 644-57; discussion 657-8.

9. Lares, B., *lares: Analytics & Machine Learning Sidekick*. 2023.

10. Buzzelli, J.N., et al., *Colorectal cancer liver metastases organoids retain characteristics of original tumor and acquire chemotherapy resistance.* Stem Cell Res, 2018. **27**: p. 109-120.

11. Li, H., et al., *Modeling tumor development and metastasis using paired organoids derived from patients with colorectal cancer liver metastases.* J Hematol Oncol, 2020. **13**(1): p. 119.

12. Martin, J., et al., *Colorectal liver metastases: Current management and future perspectives.* World J Clin Oncol, 2020. **11**(10): p. 761-808.

13. Zhou, Z., L. Cong, and X. Cong, *Patient-Derived Organoids in Precision Medicine: Drug Screening, Organoid-on-a-Chip and Living Organoid Biobank.* Front Oncol, 2021. **11**: p. 762184.

14. Danac, J.M.C., A.G.G. Uy, and R.L. Garcia, *Exosomal microRNAs in colorectal*

cancer: Overcoming barriers of the metastatic cascade (Review). Int J Mol Med, 2021. **47**(6).

15. Hu, G., K.M. Drescher, and X.M. Chen, *Exosomal miRNAs: Biological Properties and Therapeutic Potential.* Front Genet, 2012. **3**: p. 56.

16. Dai, J., et al., *Exosomes: key players in cancer and potential therapeutic strategy.* Signal Transduct Target Ther, 2020. **5**(1): p. 145.

17. Cho, W.C., et al., *Exosomal miR-193a and let-7g accelerate cancer progression on primary colorectal cancer and paired peritoneal metastatic cancer.* Transl Oncol, 2021. **14**(2): p. 101000.

18. Kalogianni, D.P., et al., *Advances in microRNA analysis.* Anal Bioanal Chem, 2018. **410**(3): p. 695-713.

19. Galasso, M., S.K. Sandhu, and S. Volinia, *MicroRNA expression signatures in solid malignancies.* Cancer J, 2012. **18**(3): p. 238-43.

20. Singh, A., et al., *DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays.* Bioinformatics, 2019. **35**(17): p. 3055-3062.

21. Team, R.C., *R: A Language and Environment for Statistical Computing*. 2021: Vienna, Austria.

22. Terry Therneau, B.A., Brian Ripley, Beth Atkinson. *rpart: Recursive Partitioning and Regression Trees*. 2022; 4.1.19:[Available from: https://cran.r-project.org/web/packages/rpart/index.html.

23. Hothorn T, Z.A., *partykit: A Modular Toolkit for Recursive Partytioning in R* Journal of Machine Learning Research, 2015. **16**: p. 3905-3909.

24. Klaus Schliep, K.H., Antoine Lizee, *kknn: Weighted k-Nearest Neighbors*. 2016.

25. Wickham, M.K.a.H., *Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles*. 2020.

26. Emil Hvitfeldt, M.K., Posit Software, PBC *discrim: Model Wrappers for Discriminant Analysis*. 2023.

27. Ripley, W.N.V.a.B.D., *Modern Applied Statistics with S*. 2002, Springer.

28. Wiener, A.L.a.M., *Classification and Regression by randomForest.* R News, 2002. **2**(3): p. 18-22.

29. Hadley Wickham, P.S., PBC, *modelr: Modelling Functions that Work with the Pipe*. 2023.

30. Kawaguchi, Y., et al., *Mutation Status of RAS, TP53, and SMAD4 is Superior to Mutation Status of RAS Alone for Predicting Prognosis after Resection*

of Colorectal Liver Metastases. Clin Cancer Res, 2019. **25**(19): p. 5843-5851.

31. Salehitabar, E., et al., *Identification of genes with high heterogeneity of expression as a predictor of different prognosis and therapeutic responses in colorectal cancer: a challenge and a strategy.* Cancer Cell Int, 2022. **22**(1): p. 276.

32. Lv, L., et al., *New Studies of the Aberrant Alterations in Fibrillin-1 Methylation During Colorectal Cancer Development.* Front Oncol, 2022. **12**: p. 862887.

33. Ding, N., et al., *CENPI is overexpressed in colorectal cancer and regulates cell migration and invasion.* Gene, 2018. **674**: p. 80-86.

34. Lu, K., et al., *Exosomes as potential alternatives to stem cell therapy for intervertebral disc degeneration: in-vitro study on exosomes in interaction of nucleus pulposus cells and bone marrow mesenchymal stem cells.* Stem Cell Res Ther, 2017. **8**(1): p. 108.

35. Chang, A.I., et al., *Involvement of mesenchymal stem cells in cancer progression and metastases.* Curr Cancer Drug Targets, 2015. **15**(2): p. 88-98.

36. Kalluri, R. and R.A. Weinberg, *The basics of epithelial-mesenchymal transition.* J Clin Invest, 2009. **119**(6): p. 1420-8.

37. Le Bras, G.F., K.J. Taubenslag, and C.D. Andl, *The regulation of cell-cell adhesion during epithelial-mesenchymal transition, motility and tumor progression.* Cell Adh Migr, 2012. **6**(4): p. 365-73.

38. Ribatti, D., R. Tamma, and T. Annese, *Epithelial-Mesenchymal Transition in Cancer: A Historical Overview.* Transl Oncol, 2020. **13**(6): p. 100773.

39. Pei, D., et al., *Mesenchymal-epithelial transition in development and reprogramming.* Nat Cell Biol, 2019. **21**(1): p. 44-53.

40. Xie, F., et al., *TGF-β signaling in cancer metastasis.* Acta Biochim Biophys Sin (Shanghai), 2018. **50**(1): p. 121-132.

41. Kaul, G., G. Pattan, and T. Rafeequi, *Eukaryotic elongation factor-2 (eEF2): its regulation and peptide chain elongation.* Cell Biochem Funct, 2011. **29**(3): p. 227-34.

42. Xie, J., et al., *Eukaryotic elongation factor 2 kinase upregulates the expression of proteins implicated in cell migration and cancer cell metastasis.* Int J Cancer, 2018. **142**(9): p. 1865-1877.

43. Shahbazian, D., et al., *The mTOR/PI3K and MAPK pathways converge on*

eIF4B to control its phosphorylation and activity. Embo j, 2006. **25**(12): p. 2781-91.

44. Shiga, A., et al., *Alteration of POLDIP3 splicing associated with loss of function of TDP-43 in tissues affected with ALS.* PLoS One, 2012. **7**(8): p. e43120.

45. Wang, X., et al., *Eukaryotic elongation factor 2 kinase activity is controlled by multiple inputs from oncogenic signaling.* Mol Cell Biol, 2014. **34**(22): p. 4088-103.

46. Miyamoto, Y., H.J. Lenz, and H. Baba, *A novel antimetabolite: TAS-102 for metastatic colorectal cancer.* Expert Rev Clin Pharmacol, 2016. **9**(3): p. 355-65.

47. Cicero, G., et al., *TAS-102 in metastatic colorectal cancer (mCRC): efficacy, tolerability, and quality of life in heavily pretreated elderly patients: a real-life study.* Drugs Context, 2020. **9**.

48. Mayer, R.J., et al., *Randomized trial of TAS-102 for refractory metastatic colorectal cancer.* N Engl J Med, 2015. **372**(20): p. 1909-19.

49. Castellano, E. and J. Downward, *RAS Interaction with PI3K: More Than Just Another Effector Pathway.* Genes Cancer, 2011. **2**(3): p. 261-74.

50. Camaj, P., et al., *KRAS exon 2 mutations influence activity of regorafenib in an SW48-based disease model of colorectal cancer.* Future Oncol, 2015. **11**(13): p. 1919-29.

국문 초록

대장암 환자에게 전이는 여전히 치료의 장애물이다. 간은 대장과 비교적 멀리 떨어져 있는 기관임에도 불구하고, 간 문맥에 의하여 연결되어 있기 때문에 대장암 전이가 가장 호발하는 부위이다. 전이의 다양한 원인 중 엑소좀 유래 마이크로 RNA는 상피-간엽 이행, 면역 억제, 종양 미세환경 구성 등의 생물학적 과정에 영향을 주어 원격 전이를 유발한다. 하지만 대장암 간 전이를 연구할 수 있는 실험 모델은 여전히 부족한 상황이다.

본 연구에서는 동일 환자 유래 대장 원발암과 간 전이암으로 구성된 11개의 환자 유래 오가노이드 (PDO) 및 환자 유래 세포주 오가노이드(PDCO)를 수립하여 대장암의 간 전이 연구를 가능하게 하는 생체 외 모델을 구축하였다. 수립한 대장암 간 전이 (CLM) 모델은 게놈 및 전사체, 약리학적 분석에 사용하였다.

본 연구는 전사체 및 마이크로 RNA, 약물 반응성 수치를 통합하는 다중 오믹스 분석을 통해 대장 원발암과 간 전이암 사이의 약물 반응성에 영향을 미치는 마이크로 RNA 바이오마커를 밝혔다. 또한 요인들 간의 상관관계 분석을 통해 약물 반응성에 영향을 미치는 간 전이암 특이적인 분자 간의 상호 작용을 서술하였다. 뿐만 아니라 기계 학습 모델을 적용하여 대장 원발암과 간 전이암의 차이를 예측할 수 있는 회귀 모델을 도출하였고, 더 나아가 KRAS 신호 조절이 간 전이암의 항암제 반응과 높

은 연관성이 있음을 제시하였다.

주요어: 대장암 간 전이, 오가노이드, 엑소좀 유래 마이크로 RNA, 다중오믹스, 항암제 스크리닝, 예측 모델

학번: 2021-28472