이학석사 학위논문

Identification of heterogeneity of colon cancer stem cells by single cell analysis

단일세포 유전체 분석을 통한 대장암 줄기세포의 이질성 규명

2018년 7월

서울대학교 융합과학기술대학원

분자의학 및 바이오제약학과

민 동 욱

Abstract

Identification of heterogeneity of colon cancer stem cells by single cell analysis

Dong Wook Min

Department of Molecular Medicine

and Biopharmaceutical Sciences

World Class University Graduate School

of Convergence Science and Technology

Seoul National University

Tumor heterogeneity is one of the ongoing huddles in the field of colon cancer therapy. It is evident that there are countless clones which exhibit different phenotypes and therefore, single cell analysis is inevitable. Cancer stem cells (CSCs) are rare cell population within tumor which is widely known to function in cancer metastasis and recurrence. Although there have been trials to prove intra-tumoral heterogeneity using single cell sequencing, that of CSCs has not been clearly elucidated. Through sequencing of CSCs, one could identify clones in their population with different functions. Here, we articulate the presence of heterogeneity within cancer stem cells through single cell sequencing. As a proof of principle, we performed phenotype-based high-throughput laser isolation and single cell sequencing (PHLI-seq) of 8 CD133 positive cells in a frozen tumor tissue obtained from a patient with colorectal cancer. The result proved that CD133 positive cells were shown to be heterogeneous both in copy number and mutational profiles. . Single cancer stem cell specific mutations such as RNF144A, PAK2, PARP4, ADAM21, HYDIN, KRT38 and CELSR1 could be also detected in metastatic tumor of the same patient. Collectively, these data suggest that single cell analysis used to identify heterogeneity within rare population, will lead to new strategies to tackle colon cancer metastasis.

Keywords: Cancer stem cells, Single cell sequencing, CD133, colon cancer

Student number: 2016 – 23387

Table of Contents

List of Figures

Introduction

The definition of intra-tumoral heterogeneity (ITH) is the genetic and phenotypic variation within cells or clones in individual tumors [1]. It could be a difference in the pattern of copy number alteration, single nucleotide variants and gene expression etc. This phenomenon is known to cause cancer to maintain its oncogenic potential and survive under drug therapy [2]. Numerous researches have been done to unravel the intra-tumoral heterogeneity and identified a rare population called cancer stem cells [3, 4].

Cancer stem cells were first treated as a controversial concept and it states that this rare subpopulation within cancer which only takes a small percentage causes therapeutic resistance and plays an important role in cancer relapse [5, 6]. Through their high expression of ATP binding cassette (ABC) transporter, which is known to be responsible for drug efflux, they could survive from general chemotherapy [7]. Therefore, this hinders the complete eradication of cancer through chemotherapy. Additionally, this subpopulation has high tumor initiating ability. This means that this subpopulation could cause cancer metastasis [8]. Cancer stem cells therefore are also inferred as “tumor initiating cells” (TICs), due to their ability to initiate tumor as the “seed” [9]. This is one of the many characteristics that cancer stem cells possess, including its ability to form spheres in 3D environment and high

expression of EMT, stem cell markers [10]. As cancer stem cells rarely reside in the cancer population, the phenotypes of these cells are easily cancelled out in bulk analysis.

Single cell sequencing is arising as a technique to successfully bring the study of genotype and phenotypes to the cellular level and it is widely used to verify rare subpopulation of cancer such as cancer stem cells [11]. There are series of steps in single cell sequencing and it is detrimental to use the right technique in each of those steps; single cell isolation, whole genome amplification and sequencing itself. Numerous isolation techniques had been developed such as laser capture microdissection, microfluidics and FACS [12]. However, there was a lack of technique to safely isolate single cells and at the same time to visually and topographically identify the rare subpopulation.

Here, our group demonstrates a novel single cell sequencing technique, called phenotype-based laser isolation single cell sequencing (PHLI-seq) [13, 14]. The phenotype of cancer stem cells we used to distinguish them was expression of CD133, a colorectal cancer stem cell marker [15]. CD133, a five-transmembrane glycoprotein, is widely used as a marker to isolate and identify colorectal cancer stem cells [16]. From patient colorectal tumor tissue, single cancer stem cells were isolated via CD133 and were sequenced for verification of intra-tumoral heterogeneity within cancer stem cells.

Material and Methods

Patient information and sample preparation

Human frozen colon cancer tissues were obtained from the Department of Surgery, Seoul National University Hospital and analyzed under the approved Institutional Review Board (IRB) protocol (IRB No. 1607-110-777). Frozen colon cancer tissues were stored at -80℃ until they were sliced into 10μm thick sections. Tissue sections were thaw-mounted onto ITO-coated glass slides. Glass slides were stored at -20℃ until analysis. The tissue used in the experiment was obtained from a 71-year-old man who underwent surgery in February 2011. The metastatic liver tumor tissue for bulk metastatic tissue exome sequencing was resected in May 2012 from the same patient.

Isolation of single CD133-positive cancer cell and Whole Genome Amplification

Single CD133-positive cancer cell and single or tens of CD133-negative cancer cells are isolated based on Phenotype based high throughput laser isolation and sequencing (PHLI-seq) method [13]. To briefly explain, immunofluorescence stained colon cancer tissue slide is first whole-slide imaged using automated microscopy (Inverted Microscope Eclipse Ti-E, Nikon Instruments Inc., Melville,

NY). Cells are classified into CD133-positive or CD133-negative based on immunofluorescence and DAPI signal, and the annotated CD133-positive/-negative cells are isolated by pulsed IR laser. We used slide glass coated with 100nm thick ITO layer for tissue section mount. Thus, irradiation of IR layer result in vaporization of ITO layer and discharging of cells of interest. The discharged cancer cells are received by 8-strip PCR tube caps, followed by lysis of the cells using proteinase K (cat no. P4850-1ML, Sigma Aldrich). After cell lysis, MDA (Multiple Displacement Amplification) is carried out using GE Illustra Genomiphi V2 DNA amplification kit (cat no. 25-6600-30) to obtain sufficient amount of gDNA from single cells. We used ^{TOPQ}XSEP MagBead (XB6050, Celemics, Seoul, Korea) for subsequent DNA purification.

Sequencing of the amplified genome

The whole genome amplified products or genomic DNA extracts from patient tissue were first fragmented using an EpiSonic Multi-Functional Bioprocessor 1100 (Epigentek) to obtain a 150~250-bp fragment size distribution. NGS sequencing library is prepared using Celemics NGS Library Preparation Kit (LI1096, Celemics, Seoul, Korea) for whole genome sequencing, and SureSelectXT Human All Exon V4 (Agilent, CA, US) for whole exome sequencing. DNA libraries were amplified using the KAPA Library Amplification Kit (KAPA Biosystems, KK2602) and

sequencing library were quatified by TapeStation 2200 (Agilent, CA, US) and QuBit fluorometer (ThermoFisher Scientific, MA, US). We generated 1Gbp/sample for whole genome sequencing and 5Gbp/sample for whole exome sequencing using HiSeq 2500 system (Illumina) and 150PE sequencing.

Sequence alignment and data preprocessing

Sequence reads in FASTQ format are first aligned to human assembly US National Center for Biotechnology Information (NCBI) build 37 (hg19) using the BWA-MEM [17] (BWA version 0.7.15) with default parameter, followed by sorting by chromosomal coordinates and PCR duplicates removal using Picard (version 2.9.1). Afterwards, reads having mapping quality score less than 30 or that have a supplementary alignment were removed for subsequent analysis.

Copy Number Alteration estimation

Copy Number Alterations (CNAs) are estimated based on the NGS read density using the variable binning method [18]. To briefly explain, human genome is divided into 5,000 variable-sized bins (median bin size = 555 kbp) whose size is adjusted to have equal expected number of uniquely mapped reads. Read depth of each bin is first LOWESS normalized based on the GC contents of each bin. And the resulting GC-normalized read density is divided by median of read depth to

convert read depth into chromosomal copy-number scale, assuming the near-diploidy of the tumor. Finally, Circular Binary Segmentation (CBS) [19] is used to detect copy-number alteration site, and mean copy number of bins in one segment is regarded as true copy number of the segment.

Single Nucleotide Variation Detection.

First, GATK (v3.7-0) IndelRealigner and BaseRecalibrator were used to locally realign reads around the Indel and to recalibrate the base quality [20]. Variant calling is performed by combining the result from 3 different caller [14]. We used GATK UnifiedGenotyper, Varscan2 [21] (ver 2.3.9), and MuTect [22] (ver 1.1.4). Briefly, variant sites detected by at least two variant caller is listed up to obtain intra-sample double called sites, which would reduce false-positive variant detection derived from NGS error [23]. Among the intra-sample double called sites, variant sites found in at least two samples (either CD133+/- cell or tumor population) were considered to be confident. This is for suppressing false variant detection caused by MDA amplification error. Finally, a variant in the sample is considered to be true if one of three variant caller called the variant, or allele count of the variant base is significantly larger than that of non-reference bases (Fisher's exact test, $p < 10^{-3}$).

Cell culture

NCI-H508 cell line was obtained from the Korea Cell Line Bank. The cell line was grown in RPMI-1640 with 10% fetal bovine serum (FBS;Welgene) and gentamicin (10μg/ml) at 37℃ in a humidified 5% CO_2 atmosphere.

FACS

Cell suspensions were washed with PBS with 0.1% BSA. 1 x 10^5 cells were incubated with primary antibody (CD133-FITC conjugated, Miltenyi biotech, Bergisch Gladbach, Germany) for 20 minutes in the dark, 4℃. Cells were further washed with PBS with 0.1% BSA. CD133 positive cells were assessed by counting the number of cells that stained positive for CD133-FITC. BD FACS Aria II was used for cell sorting.

Immunofluorescence

The slide sections were covered in Methacarn solution (60% absolute methanol, 30% Chloroform, 10% Glacial acetic acid) and fixed for 2 hours in RT. Slides were washed with absolute ethanol for 3 times followed by 20 minutes incubation in 0.2% Triton X-100. Slides were blocked in 3% BSA for 60 minutes. Blocking buffer was removed and slides were probed overnight at 4℃ with primary antibody (CD133-FITC conjugated, Miltenyi biotech, Bergisch Gladbach, Germany)

Western blot

Cultured cells were washed with ice-cold phosphate-buffered saline (PBS) and lysed with lysis buffer (50mM Tris-HCl, pH 7.5, 1% NP-40, 0.1% sodium deoxycholate, 150mM NaCl, 50mM NaF, 1mM sodium pyrophosphate, 1mM EDTA and protease/phosphatase inhibitors). Lysates were cleared by centrifugation at 13,000 rpm for 20 minutes. Protein concentrations were quantified with a Bicinchoninic Acid Protein Assay Reagent (Pierce, Rockford, IL), according to manufacturer's instructions. Samples containing equal quantities of total proteins were resolved on SDS-polyacrylamide denaturing gel, transferred to nitrocellulose membranes. The membranes were incubated in blocking solution containing 1% skim milk and 1% bovine serum albumin for an hour at room temperature and probed overnight at 4℃ with primary antibodies.

Reagents

Oxaliplatin, 5-FU and irinotecan were purchased from Selleck. Stock solutions were prepared in dimethyl sulfide (DMSO), stored at -80℃, and diluted in fresh media before each experiment. Antibodies were purchased from Miltenyi Biotech.

Cell viability assay

Cells were seeded 2500 cells each in 96 well plate at the day of seeding and after

the treatment of reagents at the following day, the viability of cells was assessed at 72 hours using MTT (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide) assay (Sigma-Aldrich).

In vitro invasion assay

After 24 hours of serum deprivation, cells were seeded on a Matrigel-coated membrane matrix (BD) in the insert of a 24-well culture plate. 20% fetal bovine serum was added to the lower chamber as a chemoattractant. After 48 hours, the non-invading cells were gently removed with a cotton swab. Migrating cells were fixed with 3.7% formaldehyde then were stained with the coomassie blue. Migrating cells were imaged under a microscope.

Sphere formation assay

CD133 ^{high} cells and CD133 ^{low} cells were sorted through FACS and incubated in DMEM/F12 with 10% B27 Supplement (Gibco), 20ng/ml EGF, 10ng/ml bFGF and 1% Penicillin/Streptomycin for indicated days. Sphere formation was observed through microscopy.

Statistics

Survival differences between cells treated with different drugs were derived from 6

experimental replicates and graphs were drawn using GraphPad Prism version 7.00 (Graphpad Software, La Jolla, CA, USA). Clustering Analysis and heatmap were drawn using R software version 3.4.1 (R project, Vienna, Austria). Non-synonymous mutations used in heatmap were supported by at least two single cells and the variants with support from only one cell were filtered out, assuming that they are false positive errors.

Results

Overall workflow of phenotype based single cell laser isolation using PHLI-seq

Frozen colorectal tumor tissue from a patient in stage III cancer was sectioned on to specialized slides and they were H&E stained. Tumor region was identified histologically by a specialist. DAPI stain was used to precisely locate the cells with CD133 expression. DAPI and CD133 stain was performed on separate tissue sections to minimize the DNA damage which could be caused by UV light while detecting DAPI signal (Figure 1A). Single cells were isolated through a novel single cell isolation technique called PHLI-seq, which was introduced in previous studies of our group [13, 14]. This is a modified version of laser capture microdissection (LCM) and it is known to cause less damage to the cell during the isolation process. Unlike the conventional laser microdissection technique where the laser, which is coupled a microscope, cuts out the single cell from the surrounding tissue, PHLI-seq, uses laser to expand the single cell volume for a short period of time and successfully 'pushes' the cell into a tube below. Through this technique, DNA within a single cell could be less damaged which leads to better quality of single cell sequencing data. Also, this technique provides topographical information from where the cells were dissociated (Figure 1B). The

isolated single cells were then brought into whole genome amplification process as only small amount of DNA is present in a single cell and this is insufficient to carry out single cell sequencing. Among series of whole genome amplification techniques which are used widely in recent single cell research field, Multiple Displacement Amplification (MDA) method was used in our study (Figure 1C). MDA uses a random hexamer as a primer and Φ29 DNA polymerase, which has strong strand displacement activity, extends the random primers and produce branched structures [11]. This is known to offer high genome coverage but is shown to cause over-amplification or under-amplification of certain regions. Therefore, amplified single cell DNA samples were thoroughly quality checked before actual single cell sequencing. Both whole genome sequencing and whole exome sequencing were carried out in order to investigate single cell heterogeneity of CD133-positive cancer stem cells.

CD133-positive cells exhibit cancer stem cell properties

Before single cell isolation from patient tissue, the cancer stem cell phenotype marker, CD133 had to be validated. CD133, also called Prominin-1, is a pentaspan membrane glycoprotein which had been used for isolation of stem-like cells from a variety of normal and pathological tissues [15]. Recently, it has become a cancer stem cell marker in different cancer types including colon cancer [16]. In order to prove that this particular marker can isolate cells with cancer stem cell phenotype, series of validations were carried out. Cancer stem cell model of NCI-H508, a colorectal cancer cell line, was used for CD133 validation. NCI-H508 was grown in cancer stem cell media for 5 days and sphere formation was detected. CD133-positive cells were isolated using FACS along with CD133 negative proportion of parental cell line (Figure 2A). Both of the cells were cultured in cancer stem cell media for three days and far more number of spheres was formed by CD133 positive cells (Data not shown). CD133 protein expression in CSC was tested by western blot which was higher in cancer stem cell (Figure 2B). As mentioned in the introduction, cancer stem cells tend to show resistance to chemotherapeutic drugs [7]. Three different types of chemotherapeutic drugs which are widely used in the clinical field were treated to parental cell line and cancer stem cells. While parental cells showed decrease in cell viability in dose-dependent manner, cancer stem cells exhibited relative resistance to drugs such as fluorouracil, oxaliplatin and

irinotecan (Figure 2C).

These results imply that CD133 can serve as colorectal cancer stem cell marker and this was shown by sphere formation assay, protein expression and cell viability assay against general chemotherapeutic drugs.

Copy Number, mutational profiling identify heterogeneity within single CD133-positive cancer stem cells

Frozen tumor tissue was from a 71 years old male patient in stage III cancer. After primary tumor resection, this patient acquired liver metastasis and the resection of liver metastatic tumor was carried out in the following year. The patient had no detectable KRAS mutation and was microsatellite-stable. Microsatellite-stable patient was chosen as there was a need to eliminate the chance of sporadic mutational events which might conceal the development of metastatic tumor due to cancer stem cells. CD133-positive 30 single cells were isolated from 5 consecutive tissue sections of a colorectal cancer patient. CD133 negative region which consists of 10 to 15 cells were also sampled from the same frozen tissue slide to verify that variants found in single cancer stem cells are CSC population specific. The slides were sectioned into a thickness of 10μm and were stained right after. After single cell isolation process, the samples have gone through whole genome sequencing first in order to check for the amplification bias and also to investigate intra-tumoral heterogeneity in copy number profile. Whole genome sequencing of normal tissue, tumor tissue, CD133-negative region and 8 single CD133-positive cells was conducted in a low depth of 0.1x. Copy number variation pattern was drawn for individual samples and checked for missing chromosomes or over-amplified regions. Through this, 8 CD133-positive single cells were remained

ready for single cell exome-sequencing. Copy number variation of samples showed similar and different patterns. As previously known, tumor tissue showed higher copy number variation compared to normal tissue. In case of CD133-negative cells, they showed different copy number alteration pattern compared to single CD133-positive cells for instance in chromosome 5, 7 and 11 etc. In addition, single CD133-positive cells exhibited similar CNV patterns in chromosome 5, 8, 20 etc. while showed difference in chromosome 1, 7, 12, 13 and 16 etc (Figure 3A). Single cell exome-sequencing was conducted in a depth of 100x. Some had twice as much mutation rate per 10^6 bases, but in general they had low mutation rates of <10 per 10^6 bases which further indicates that this patient was microsatellite stable (Figure 3B). CSC1 and 3 possessed higher number of mutations compared to bulk tumor tissue and other single CD133-positive cells. The mutation burden of the single CD133-positive cells and CD133-negative region were all dominated by C:G to T:A and C:G to A:T transition (Figure 3C). In general, they all show previously studied mutation signature of colorectal cancer. However, CSC1 and 3 exhibited less proportion of C:G to T:A and C:G to A:T transition but slightly higher transition of T:A to C:G. In addition, CD133 negative cells showed larger difference when compared to single CD133-positive cells which is twice as higher proportion of C:G to A:T transition. The variants with support from only one cell were filtered out, assuming that they are false positive errors [24]. In total number

of 107 non-synonymous mutations were collected and the cells were clustered by clustering analyses package in R. Bulk tumor tissue and CD133-negative region were clustered together showing CD133-positive cells are rare population which reside in bulk tissue. PCA analysis was carried out using the non-synonymous mutations in order to identify clusters of samples which possess similar mutations. In result, there were 3 clusters of samples which were identified; Cluster 1 with CSC1 and 3, Cluster 2 with CSC8, 13, 14 and 15 and finally Cluster 3 with CSC 7,10 and CD133 negative cells and bulk tumor tissue (Figure 3D). Mutation correlation analysis also supported the clustering analysis and showed CSC1 and CSC3 are scarcely related to other single cancer stem cells (Figure 3E). Heatmap was drawn using non-synonymous mutation calls in single CD133-positive cells and bulk tumor tissue (Figure 3F). Heatmap of non-synonymous mutations identified single CD133-positive cell specific mutations which are absent in primary bulk tissue for instance mutations in KCNJ13, HECW1, RELN and OR51D1. CSC1 and 3 which were clustered together both in PCA and mutation correlation analysis commonly shared strong driver mutations such as APC, TP53 and TLE4. Although there were mutations which are found commonly throughout single CD133-positive cells, there were mutations which their presence varied throughout the samples. This implies that single cell sequencing had identified mutations which could not be detected in bulk tissue sequencing.

These results indicate that intra-tumoral heterogeneity of copy number alteration pattern and somatic mutational profile could be spotted not only in between CD133-positive and negative cells but also in between single CD133-positive cells.

Mutations in single cancer stem cells are also identified in metastatic tumor tissue

One of the characteristics of cancer stem cells is their ability to initiate tumor and to trigger metastasis, as mentioned previously. It is known in numerous previous studies that CD133, the marker our group used to sort out cancer stem cell from the tissue, is expressed in colon cancer cells with metastatic potential, leading to poor prognosis of patients [8]. In order to prove that this could be implemented in our case, same cell sample was used from Figure 2 where cells with high CD133 expression were assessed for their stem cell properties. Cell invasion ability was compared between CD133 ^{low} cells and CD133 ^{high} cells. They were seeded in transwell chamber for 48 hours and cells passed through the pores were stained with crystal violet. CD133 high cells exhibited higher ability to invade compared to CD133 low cells (Figure 4A). Furthermore, the difference in CD133 expression was investigated by immunofluorescence, in primary and metastatic whole tissue section from PHLI-seq. Both primary and metastatic tissue was stained with CD133 antibody and expression of CD133 was assessed visually through confocal microscopy. Metastatic tissue sections exhibited larger amount of CD133 in general (Data not shown). Invasion assay and immunofluorescene support that CD133 expressing cells have the ability to initiate tumor in the secondary site of tumor. Then our group investigated whether the mutations in metastatic tumor tissue are

originated from single CD133-positive cells in primary tumor tissue. First, mutational signature was investigated and strikingly similar base transition pattern was identified in between primary and metastatic tumor tissue. They were mostly dominated by C:G to T:A transition. However, single CD133-positive cells showed different pattern of signature and they exhibited higher proportion of C:G to T:A and C:G to A:T transition (Figure 4B). 230 non-synonymous mutations were identified in bulk metastatic tumor tissue and our group investigated the presence of these mutations in the exome sequencing raw data of primary tumor tissue and single cancer stem cells. 7 mutations (RNF144A, PAK2, PARP4, ADAM21, HYDIN, KRT38 and CELSR1) were commonly identified in single CD133-positive cells and metastatic tumor tissue but not in bulk primary tumor tissue (Figure 4C). Well-known driver mutation genes such as PIK3CA were included but in different site of mutation and these may have roles in metastasis and tumor initiation. In order to investigate how mutational genes specifically detected in both single cancer stem cells and bulk metastatic tumor can contribute to liver metastasis, gene set enrichment analysis was carried out using the common mutational genes (Figure 4D). In result, signaling pathway of NOTCH and FGFR1 was found to be mainly induced. It could be supported by the fact that there have been numerous studies on how NOTCH and FGFR1 pathway induces colon cancer metastasis to liver [25, 26].

In conclusion, our group had shown that rare population in primary tumor tissue, which is challenging to be detected by bulk sequencing, is responsible for liver metastasis in colorectal cancer. In our case, this was CD133-positive cells.

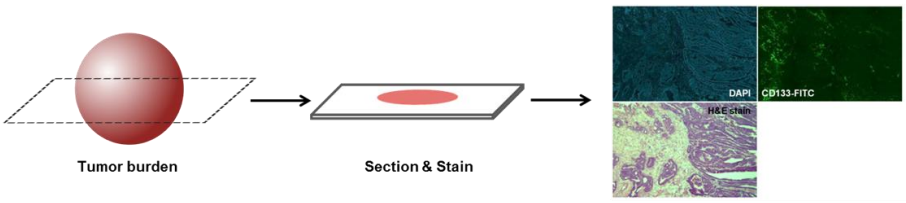
Figure 1. Overall workflow of phenotype based single cell laser isolation and Sequencing (PHLI-seq)

(A) Firstly, frozen tissue was sectioned and stained with CD133 antibody, DAPI and H&E

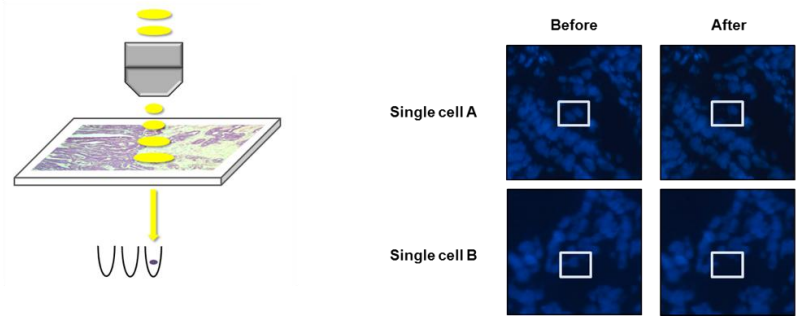
(B) Based on the whole tissue image, CD133-positive cells were isolated via phenotype based single cell laser isolation system

(C) DNA was extracted from isolated single cells and single cell exome-seq was conducted. Finally, the data were analyzed and validated.

(A)


Tumor burden
Section & Stain
DAPI
CD133-FITC
H&E stain
Whole Tissue stained image & targeting

(B)


Before
After
Single cell A
Single cell B

(C)

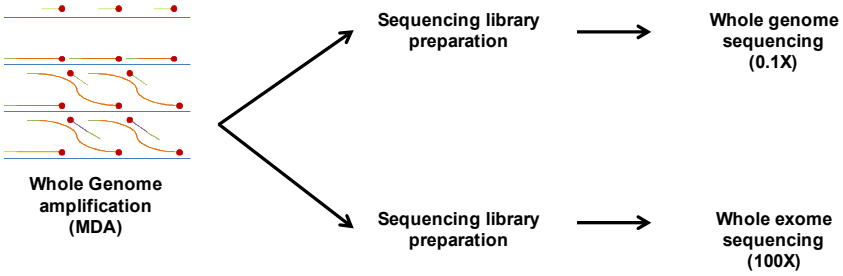

Whole Genome amplification (MDA)
Sequencing library preparation
Whole genome sequencing (0.1X)
Sequencing library preparation
Whole exome sequencing (100X)

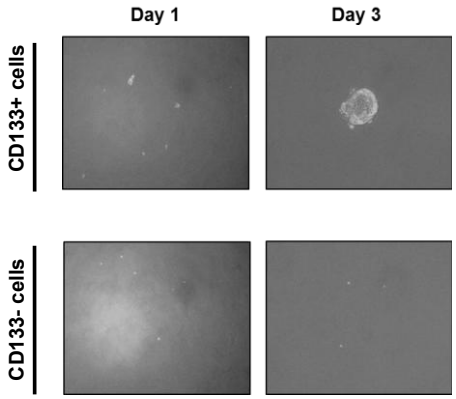
Figure 2. CD133-positive cells exhibit cancer stem cell properties

(A) NCI-H508 colorectal cancer cell line was stained with CD133 antibody and sorted into two groups; CD133-positive/negative. Cells were then cultured in cancer stem cell media for indicated days.

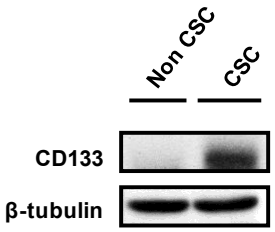
(B) Western blot analysis of CD133 expression in parental NCI-H508 cell line and cancer stem cells, which were derived from it.

(C) Drug sensitivity assay against chemotherapeutic drugs; 5-FU, oxaliplatin and SN-38. Parental NCI-H508 cells and cancer stem cells were seeded into 96 well culture plates and viability was assessed after 72 hours, post-treatment. Blue indicates the graph of parental cell line, red shows that of cancer stem cells. Error bars indicate SD.

(A)


Day 1
Day 3
CD133+ cells
CD133- cells

(B)


Non CSC
CSC
CD133
β-tubulin

(C)

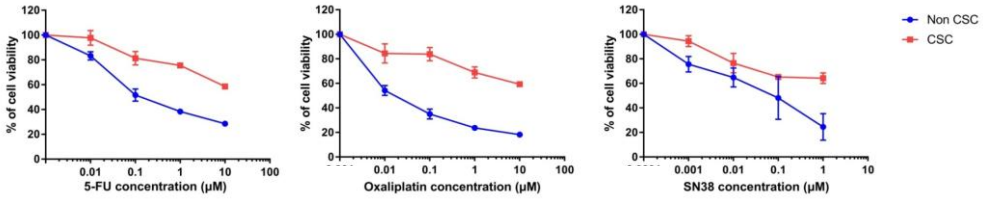

% of cell viability
5-FU concentration (μM)
Oxaliplatin concentration (μM)
SN38 concentration (μM)
Non CSC
CSC

Figure 3. Copy Number, mutational profiling identify heterogeneity within single CD133-positive cancer stem cells

(A) Copy number variation patterns of normal, tumor tissue, CD133-negative region and 8 single CD133-positive cells.

(B) The numbers of silent/non-silent mutations per 10^6 bases in each single cancer stem cells, primary tumor tissue and CD133-negative region.

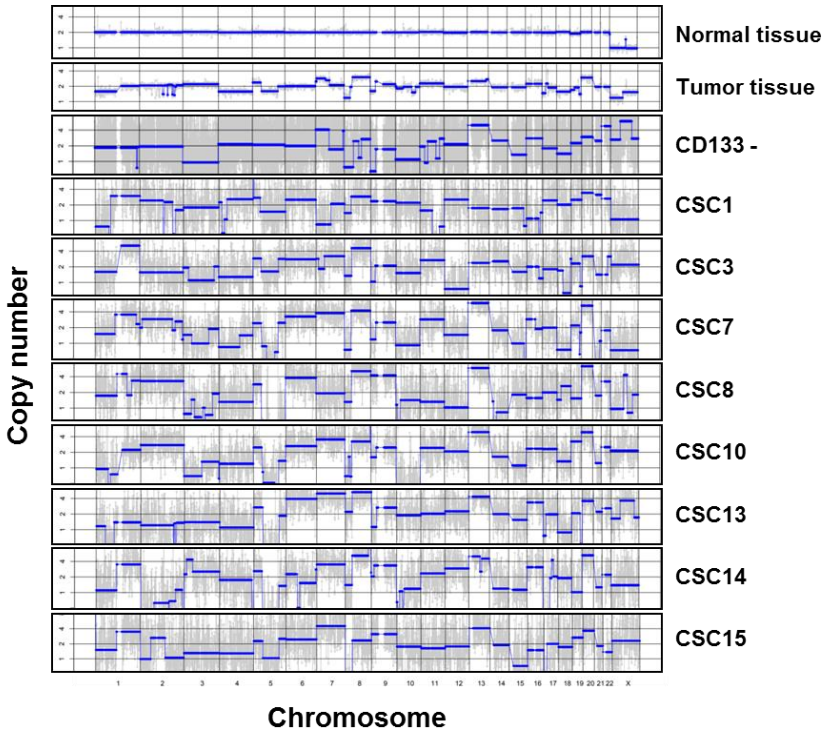
(C) Mutational signatures of the isolated single CD133-positive cells and CD133-negative region.

(D) Principal components analysis (PCA) analysis of cancer stem cells based on the SNVs from tumor tissue, CD133-negative region and 8 single CD133-positive cells.

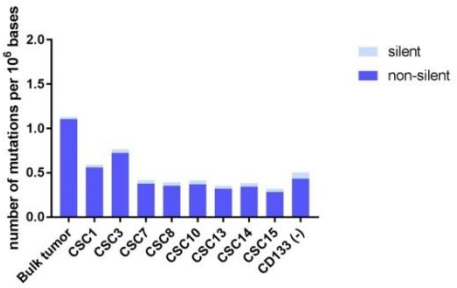
(E) Mutation correlation analysis of tumor tissue, CD133-negative region and 8 single CD133-positive cells.

(F) Heatmap of non-synonymous SNVs from tumor tissue, CD133-negative region and 8 single CD133-positive cells. Red indicates the absence of variants and Blue indicates the presence.

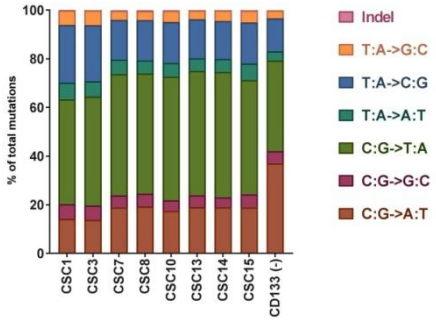
(A)


Normal tissue
Tumor tissue
CD133 -
CSC1
CSC3
CSC7
CSC8
CSC10
CSC13
CSC14
CSC15
Copy number
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
Chromosome

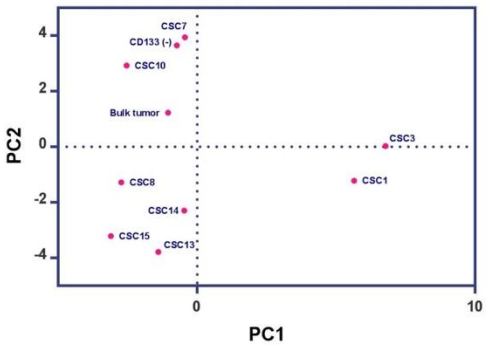
(B)


number of mutations per 10^6 bases
2.0
1.5
1.0
0.5
0.0
silent
non-silent
Bulk tumor
CSC1
CSC3
CSC7
CSC8
CSC10
CSC13
CSC14
CSC15
CD133 (-)

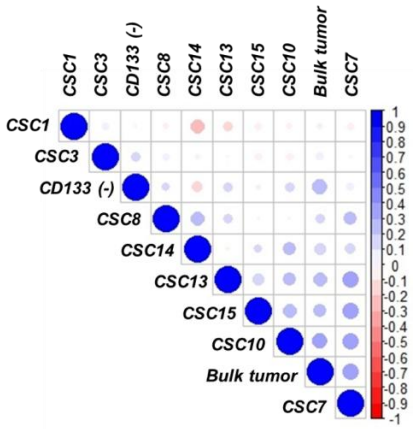
(C)


% of total mutations
100
80
60
40
20
0
Indel
T:A->G:C
T:A->C:G
T:A->A:T
C:G->T:A
C:G->G:C
C:G->A:T
CSC1
CSC3
CSC7
CSC8
CSC10
CSC13
CSC14
CSC15
CD133 (-)

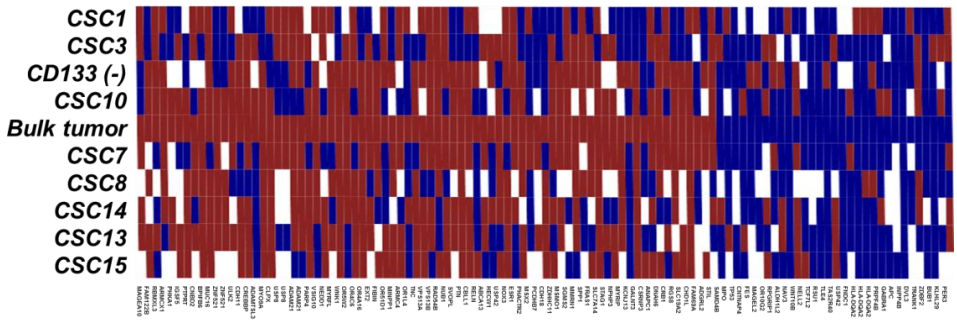
(D)



(E)



(F)



Color Key

: With Variant

: Without Variant

Figure 4. Mutations in single cancer stem cells are also identified in metastatic tumor tissue

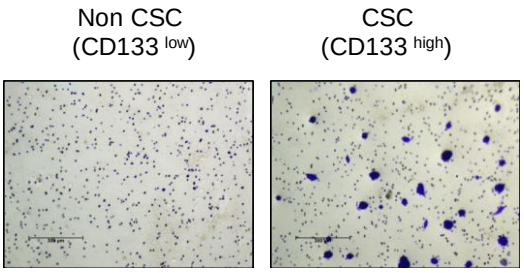
(A) Cell invasion analysis of CD133 $^{low/high}$ cells were serum-starved for 24 hours and cultured for 48 hours in transwell chamber. Images were captured at a magnification of 50x.

(B) Mutational signatures of the isolated single cancer stem cells, bulk primary tumor and bulk metastatic tumor.

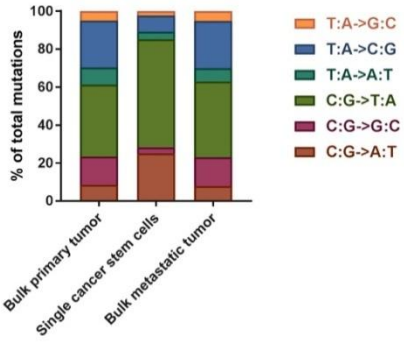
(C) A schematic summary of the mutations which are commonly found in bulk metastatic tumor tissue and single cell CSCs from primary tissue but not in bulk primary tumor tissue.

(D) Gene set enrichment analysis of mutational genes commonly detected in single cancer stem cells and bulk metastatic tumor

(A)



(B)

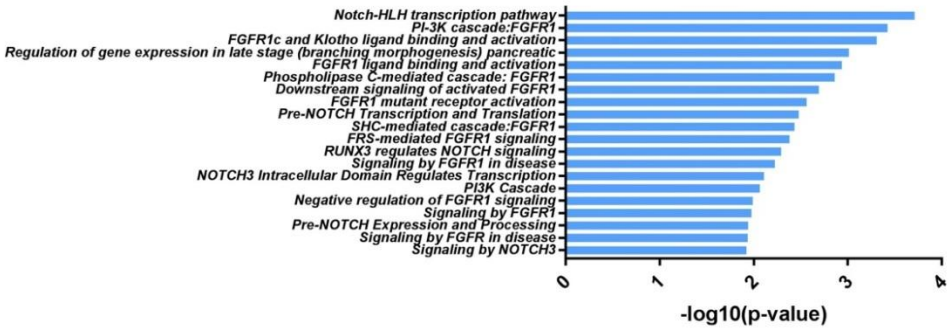


(C)

	Bulk tumor	Single CSC	Bulk metastatic tumor
TP53_p.R175H			
APC_p.R216*			
PIK3CA_p.E976*			
PIK3CA_p.Q546K			
RNF144A_p.T4A			
PAK2_p.Q101H			
PARP4_p.H490Q			
ADAM21_p.P40L			
HYDIN_p.E2994G			
KRT38_p.Q235*			
CELSR1_p.T2234M			

: Missense mutation
: Splice site
: Nonsense mutation

(D)


Notch-HLH transcription pathway
PI-3K cascade:FGFR1
FGFR1c and Klotho ligand binding and activation
Regulation of gene expression in late stage (branching morphogenesis) pancreatic
FGFR1 ligand binding and activation
Phospholipase C-mediated cascade: FGFR1
Downstream signaling of activated FGFR1
FGFR1 mutant receptor activation
Pre-NOTCH Transcription and Translation
SHC-mediated cascade:FGFR1
FRS-mediated FGFR1 signaling
RUNX3 regulates NOTCH signaling
Signaling by FGFR1 in disease
NOTCH3 Intracellular Domain Regulates Transcription
PI3K Cascade
Negative regulation of FGFR1 signaling
Signaling by FGFR1
Pre-NOTCH Expression and Processing
Signaling by FGFR in disease
Signaling by NOTCH3
0
1
2
3
4
-log10(p-value)

Discussion

In this study, our group intended to prove intra-tumoral heterogeneity which is known to be present in a rare population of cancer burden, called cancer stem cells. This was verified by novel phenotype based high throughput laser single cell isolation and sequencing (PHLI-seq) technique. This study serves as a case report of a patient with colon cancer who further acquired liver metastasis.

The concept of cancer stem cells (CSCs) was proposed decades ago, supporting the idea that tumor growth is fueled by small numbers of tumorigenic stem cells [10]. Recent lineage tracing techniques could prove the actual existence of CSCs. The discovery of colorectal cancer stem cells further supported the idea of intra-tumoral heterogeneity by revealing the existence of various markers and their ability to self-renew, resist chemotherapy and seed secondary tumors [6]. There are numerous colon cancer stem cell markers such as CD44, LGR5, CD166 and CD133. CD133 is especially known to be expressed in a rare fraction of cancer stem cells which fuel colon cancer [27]. It is also known that CD133-positive colorectal cancer stem cells grow exponentially in vitro as undifferentiated spheres under serum-free conditions and generate tumors in vivo with properties similar to their tumor of origin and also cause liver metastasis of colon cancer [28]. Therefore, CD133 was a promising option for PHLI-seq to effectively isolate colon cancer stem cells from frozen tissue.

As the proportion of cancer stem cells within tumor is low, there was a need for single cell sequencing. Single cell sequencing had enabled the cancer researchers to unravel the genetic information of rare populations in tumor such as cancer stem cells to prove intra-tumoral heterogeneity. By conducting sequencing of cancer stem cells, one can increase not only the understanding of cancer stem cells in tumor heterogeneity, but also the knowledge to identify drugs to target CSCs [29]. Recently, as the technology of single cell sequencing develops, there have been numerous trials to unravel the genetic, transcriptomic information of cancer stem cells [30, 31]. However, there was lack of trials to phenotypically isolate colorectal cancer stem cells and to prove intra-tumoral heterogeneity. This study not only unraveled the intra-tumoral heterogeneity, but also showed how the rare population of cancer stem cells in primary site can be related to liver metastatic tumor in the same patient. Single cell sequencing can be applied to investigate stages of cancer progression, in this case, from primary tumor development to metastasis [32]. Here, we verified that mutations that were found specifically in single CD133-positive cancer stem cells could be detected in bulk sequencing of liver metastatic tumor tissue in the same patient.

PHLI-seq method that was used in this study provides numerous advantages. Firstly, as PHLI-seq uses an infrared laser to expand the volume of cell of interest to ‘push’ the cell down without damaging the cell. This is unlikely to happen in

conventional laser capture microdissection methods where laser is used to ‘cut’ the single cell out from adjacent tissue which is likely to charge damage to the single cell. Secondly, PHLI-seq not only can isolate the cells via their phenotypes, but also can acquire spatial information of where the single cell was located before isolation. This could provide insight of the microenvironment surrounding the single cell of interest.

There were still remaining improvements that could have been made in this study. First of all, there was a lack of sample number due to the rarity of CD133-positive cancer stem cells on the tissue section. In order to support the idea of intra-tumoral heterogeneity in cancer stem cells, more number of samples should be needed. The whole study could also be done in another patient with metastatic tumor burden.

Various types of bias could be occurred as the initial amount of DNA is minute. Small error in the initial DNA can be thus critical and there were numerous steps where the amplification could go wrong; first, DNA could be damaged during single cell isolation process and thus, some chromosomes might be missing at the first place. Moreover, it could be due to a frequently occurring error in Multiple Displacement Amplification which is over-amplification of a random region. As the amplification is done isothermally, some part of the chromosomal region can be over-expressed in the final sample. Increase in the number of samples would lessen the amount of bias which arises during single cell sequencing.

This study is meaningful as it suggested that single cell sequencing could identify mutations in rare population which participate in liver metastasis. This means, through single cell sequencing of cancer stem cells, one can predict for metastasis. We detected strong driver mutations such as PIK3CA both in single cancer stem cells and bulk metastatic tumor tissue, which implies that mutations that are responsible for metastasis were already present in primary tumor tissue.

In summary, our study is about unraveling intra-tumoral heterogeneity of CD133-positive cancer stem cells and it suggests cancer stem cells in primary colorectal tumor are responsible for liver metastasis.

References

1. Burrell, R.A., et al., *The causes and consequences of genetic heterogeneity in cancer evolution.* Nature, 2013. **501**(7467): p. 338-45.
2. Gerashchenko, T.S., et al., *Intratumor heterogeneity: nature and biological significance.* Biochemistry (Mosc), 2013. **78**(11): p. 1201-15.
3. Zheng, H., et al., *Single cell analysis reveals cancer stem cell heterogeneity in hepatocellular carcinoma*. Hepatology, 2018.
4. Chen, Y.C., et al., *High-Throughput Single-Cell Derived Sphere Formation for Cancer Stem-Like Cell Identification and Analysis.* Sci Rep, 2016. **6**: p. 27301.
5. De Angelis, M.L., et al., *Cancer Stem Cell-Based Models of Colorectal Cancer Reveal Molecular Determinants of Therapy Resistance.* Stem Cells Transl Med, 2016. **5**(4): p. 511-23.
6. Zeuner, A., et al., *Colorectal cancer stem cells: from the crypt to the clinic.* Cell Stem Cell, 2014. **15**(6): p. 692-705.
7. Begicevic, R.R. and M. Falasca, *ABC Transporters in Cancer Stem Cells: Beyond Chemoresistance.* Int J Mol Sci, 2017. **18**(11).
8. Shiozawa, Y., et al., *Cancer stem cells and their role in metastasis.* Pharmacol Ther, 2013. **138**(2): p. 285-93.
9. Paget, S., *THE DISTRIBUTION OF SECONDARY GROWTHS IN CANCER OF THE BREAST.* The Lancet. **133**(3421): p. 571-573.

10. Batlle, E. and H. Clevers, *Cancer stem cells revisited.* Nat Med, 2017. **23**(10): p. 1124-1134.

11. Gawad, C., W. Koh, and S.R. Quake, *Single-cell genome sequencing: current state of the science.* Nat Rev Genet, 2016. **17**(3): p. 175-88.

12. Grun, D. and A. van Oudenaarden, *Design and Analysis of Single-Cell Sequencing Experiments.* Cell, 2015. **163**(4): p. 799-810.

13. Kim, S., et al., *Constructing and visualizing cancer genomic maps in 3D spatial context by phenotype-based high-throughput laser-aided isolation and sequencing (PHLI-seq).* bioRxiv, 2018.

14. Kim, S., et al., *Evaluating Tumor Evolution via Genomic Profiling of Individual Tumor Spheroids in a Malignant Ascites from a Patient with Ovarian Cancer Using a Laser-aided Cell Isolation Technique.* bioRxiv, 2018.

15. Li, Z., *CD133: a stem cell biomarker and beyond.* Exp Hematol Oncol, 2013. **2**(1): p. 17.

16. Ren, F., W.Q. Sheng, and X. Du, *CD133: a cancer stem cells marker, is used in colorectal cancers.* World J Gastroenterol, 2013. **19**(17): p. 2603-11.

17. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

18. Baslan, T., et al., *Genome-wide copy number analysis of single cells.* Nat Protoc, 2012. **7**(6): p. 1024-41.

19. Venkatraman, E.S. and A.B. Olshen, *A faster circular binary segmentation algorithm for the analysis of array CGH data.* Bioinformatics, 2007. **23**(6): p. 657-63.

20. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data*. Genome Res, 2010. **20**(9): p. 1297-303.

21. Koboldt, D.C., et al., *VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing*. Genome Res, 2012. **22**(3): p. 568-76.

22. Cibulskis, K., et al., *Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples.* Nat Biotechnol, 2013. **31**(3): p. 213-9.

23. Lodato, M.A., et al., *Somatic mutation in single human neurons tracks developmental and transcriptional history.* Science, 2015. **350**(6256): p. 94-98.

24. Zafar, H., et al., *Monovar: single-nucleotide variant detection in single cells.* Nat Methods, 2016. **13**(6): p. 505-7.

25. Candy, P.A., et al., *Notch-induced transcription factors are predictive of survival and 5-fluorouracil response in colorectal cancer patients.* British Journal Of Cancer, 2013. **109**: p. 1023.

26. Sato, T., et al., *Overexpression of the fibroblast growth factor receptor-1 gene correlates with liver metastasis in colorectal cancer.* Oncol Rep, 2009. **21**(1): p. 211-6.

27. Francipane, M.G., et al., *Crucial role of interleukin-4 in the survival of colon cancer stem cells.* Cancer Res, 2008. **68**(11): p. 4022-5.

28. Klonisch, T., et al., *Cancer stem cell markers in common cancers – therapeutic implications.* Trends in Molecular Medicine. **14**(10): p. 450-460.

29. Klevebring, D., et al., *Sequencing of breast cancer stem cell populations indicates a dynamic conversion between differentiation states in vivo.* Breast Cancer Res, 2014. **16**(4): p. R72.

30. Giessler, K.M., et al., *Genetic subclone architecture of tumor clone-initiating cells in colorectal cancer.* J Exp Med, 2017. **214**(7): p. 2073-2088.

31. Giustacchini, A., et al., *Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia.* Nat Med, 2017. **23**(6): p. 692-702.

32. Tsoucas, D. and G.C. Yuan, *Recent progress in single-cell cancer genomics.* Curr Opin Genet Dev, 2017. **42**: p. 22-32.

국문 초록

종양 내 이질성은 대장 암 치료 분야에서 지속되는 난관 중 하나이다. 다른 표현형을 나타내는 수많은 클론이 존재하므로 단일 세포 분석이 불가피하다는 것이 분명하다. 암 줄기 세포 (Cancer stem cell, CSCs)는 종양 내에서 드물게 나타나는 세포 집단으로 암전이 및 재발에 깊숙이 관여하는 것으로 알려져 있다. 단일 세포 염기 서열 분석을 사용하여 종양 내 이질성을 입증하는 여러 시도가 있었지만, 암 줄기 세포에서는 분명히 밝혀지지 않았다. 암 줄기 세포의 염기 서열 분석을 통해, 다른 기능을 가진 클론들을 동정 할 수 있다.

이 논문에서는 단일 세포 염기 서열 분석을 통해 암 줄기 세포 내에서 이질성의 존재를 분명히 한다. 우리는 대장암 환자에게서 얻은 냉동 종양 조직에서 표현형 기반 고 처리량 레이저 분리 및 8 개의 CD133 양성 세포의 단일 셀 시퀀싱 (PHLI-seq)을 수행하였다. 그 결과, CD133 양성 세포는 복제 수 및 돌연변이 프로파일 모두에서 이질적으로 나타났다. 또한 전이성 암의 돌연변이 분석이 동일한 환자에서 수행되었으며, 원발성 암에는 발견되지 않았지만 RNF144A, PAK2, PARP4, ADAM21, HYDIN, KRT38 and CELSR1 와 같은 돌연변이가 전이 암과 일부 단일 CSC 에서 모두 검출 될 수 있음이 밝혀졌다.

종합적으로, 이 데이터는 희귀 개체군 내의 이질성을 확인하는 데 사용되는 단일 세포 분석이 결장암 전이에 대처하는 새로운 전략으로 이어질 것이라고 제안한다.

주 요 어: 암줄기세포, 단일세포 분석법, CD133, 대장암

학 번: 2016-23387