**Material and Methods**

**Samples and patients collection**

The study was approved by the Guangxi Medical University Ethics Review Board and was performed in agreement with the Helsinki Declaration. All participants provided informed consent. The study was performed at Guangxi Clinical Research Center for Colorectal Cancer (Nanning, China) from March 2018 and May 2019. A total of 70 histologically confirmed mCRC and 30 CRC (total = 100) cases that had matched pairs of tissue and peripheral blood samples for circulating tumor cells (CTCs) analysis, and 12 mCRC cases for which matched pairs of primary tumor tissues, adjacent mucosa tissues and liver metastases tissues were evaluated for proteomics analysis.

**CTCs isolation and identification**

Canpatrol® system (SurExam Bio-Tech, China) be uesd for rapid, size-based capture of CTCs from peripheral blood (PB). CTCs separation, enrichment and classification identification of two parts, CTC enrichment and multiple RNA in situ analysis and detection.

**CTCs enrichment**

According to the CTC enrichment technology of ISET, using a calibrated membrane with 8uM diameter pores and multiplex mRNA in situ hybridization (ISH) assay to identify and classify CTCs, the optimized enrichment technique is more effective for CTC isolation and characterization. 5ml peripheral blood of patients were collected in EDTA anticoagulated tube, then erythrocytes were removed by a red blood cell lysis buffer (154 mM NH4Cl, 10 mM KHCO3 and 0.1 mM EDTA). And the cells were transferred to the filtration tube after resuspended in PBS containing 4% formaldehyde for 5 minutes. Then the pump valve was switched on to reach at least 0.08MPa and the manifold vacuum plate valve was then switched on to fulfill filtration.

**Multiple RNA in situ analysis and detection**

To help us distinguish epithelial, mesenchymal and hybrid CTCs. On the membrane of the 24-well plate, cells were treated with a protease before hybridization with indicated capture probe specific for EpCAM, CK8/18/19, vimentin, twist, and CD45(Table 1). After incubation at 42°C for 2 h, cells were washed with buffer to remove the unbound probes. Then cells were incubated with preamplifier solution [30% horse serum, 1.5% sodium dodecyl sulfate, 3mM Tris-HCl (pH 8.0) and 0.5 fmol of preamplifier; the sequences are shown in Table 2] at 42°C for 2 h for the purpose of signal amplification. The membranes were washed with 1000 μl of wash buffer (0.1 × SSC), and then incubated with 100 μl of amplifier solution [30% horse serum, 1.5% sodium dodecyl sulfate, 3mM Tris-HCl (pH 8.0) and 1 fmol of amplifier; the sequences are shown in Table 2]. Fluorescently labeled probes, which had been conjugated with fluorescent dyes Alexa Fluor 594 (for the epithelial biomarkers EpCAM and CK8/18/19), Alexa Fluor 488 (for the mesenchymal biomarkers vimentin and twist), Alexa Fluor 750(for CD45), Alexa Fluor 647 (CD133), were added and incubated at 42°C for 2 min. After staining with DAPI, cells were analyzed with a fluorescence microscope (Olympus BX53, Tokyo, Japan). Red and green fluorescence signal points represent the expression of epithelial and interstitial genes on CTC（Figures.1）.

****

Fig.1 CTCs isolation and identification

**Data independent acquisition (DIA)**



Fig.2 DIA experimental procedure

**Sample preparation**

Frozen samples were transferred into low protein binding tubes (1.5ml Eppendorf) and lysed with 300 μL lysis buffer supplemented with1mM PMSF. Then samples were further lysed with sonication. The parameters were set as 1s/1s intervals, 3min time and 80W power. After sonication, the samples were centrifuged at 15000 g for 15 min to remove insoluble particles, repeat once to further exclude precipitation. Protein concentration was determined by BCA assay and aliquoted to store at -80°C。

**Protein Concentration Determination**

According to the instructions of BCA (Bicinchonininc acid) kit, the required volume of BCA working reagent is allocated by buffer A: Buffer B = 50:1 (v/v). Take out part of the protein solution to be tested and dilute it with ultra-pure water (to prevent excessive concentration from exceeding the working range of the standard curve). Prepare a clean 96-well plate and add BSA standard protein solution of gradient as follows: 0, 1, 2, 4, 8, 12, 16, and 20 μL. Then add the corresponding volume of ultra-pure water to each hole to supplement the volume to 20 μL. Two microliter protein solution was added to 96-well plate with three multiple holes for each sample, and the volume of the protein solution was added to 20 μL. Then 200 μL pre-configured working reagent was added into each hole, and the mixture incubate for 30 minutes at the temperature of 37. The absorbance value (wave length 562 nm) was determined by enzyme labeling instrument. Then calculating the standard curve according to the known concentration and absorbance value of the standard protein solution, and substituting the absorbance value of the sample to be measured, the protein concentration value can be calculated.

**SDS-PAGE electrophoresis**

The 10 ug proteins of each sample were acquired and separated by 12% SDS-PAGE gel. Then separation gel was stained by CBB according to Candiano’s protocol: Firstly, the gel was fixed for 2 h and stained for 12 h. After staining, the gel was washed with water until the bands were visualized. Finally, the stained gel was scanned by ImageScanner (GE Healthcare, USA) at the resolution of 300dpi.

**Proteolysis**

According to the measured protein concentration, take the same quality protein from each sample, and dilute different groups of samples to the same concentration and volume. Add 25mm DTT of corresponding volume into the above protein solution to make the DTT final concentration about 5mM, and incubate at 55°C for 30-60min. Then add the corresponding volume of iodoacetamide so that the final concentration is about 10mm, and place in the dark for 15-30min at room temperature. Then 6 times of the volume of precooled acetone in the above system to precipitate the protein, and place it at - 20 °Cfor more than four hours or overnight. After precipitation, take out the sample and centrifuge at 8000g for 10 min at 4 °C for collecting the precipitate. According to the amount of protein, add the corresponding volume of enzymolysis diluent (protein:enzyme =50:1(m/m),100ug of protein add 2ug of enzyme) to redissolve the protein precipitate, then the solutions were incubated for digestion at 37°C for 12h.

**Desalination**

The solutions were added 2-4μL H3PO4 to each sample to adjust the pH value to 1-3.

The digested peptides were desalted by C18-Reverse-Phase SPE Column. Firstly, rinse once with methanol 1mL, once with 90% acetonitrile-water 1mL , once with water ，All solutions contain 0.1% TFA. And the samples were loaded on the column 3 times. Then the column was washed by 0.1%TFA/H2O 3 times. Finally, the peptides was eluted with 90% ACN/H2O (containing 0.1%TFA) 3 times and were lyophilized.

**LC-MS/MS high resolution mass spectrometry**

**RPLC analysis**

RP separation was performed on an 1100 HPLC System (Agilent) using an Agilent Zorbax Extend RP column (5 μm, 150 mm × 2.1 mm). Mobile phases A (2% acetonitrile in HPLC water) and B (90% acetonitrile in HPLC water) were used for RP gradient. The solvent gradient was set as follows: 0~10 min, 98% A; 10~10.01 min, 98%~95% A; 10.01~37 min, 95%~80% A; 37~48 min, 80~60% A; 48~48.01 min, 60~10% A; 48.01~58 min, 10% A; 58~58.01 min, 10~98% A; 58.01~63 min, 98% A. Tryptic peptides were separated at an fluent flow rate of 250μL/min and monitored at 210 nm. Samples were collected for 10-50 minutes, and eluent was collected in centrifugal tube 1-10 every minute in turn. Samples were recycled in this order until the end of gradient. The separated peptides were lyophilized for mass spectrometry.

**Mass spectrometry analysis**

All analyses were performed by a Q-Exactive HF mass spectrometer (Thermo, USA) equipped with a Nanospray Flex source (Thermo, USA). Samples were loaded and separated by a C18 column (50 cm × 75 μm) on an EASY-nLCTM 1200 system (Thermo, USA). The flow rate was 300 nL/min and linear gradient was 90 min (0~82min, 5-44%B; 82~84 min, 44-90% B; 84~90 min, 90% B; mobile phase A = 0.1% FA in water and B = 0.1% FA in 80%ACN).

**DDA**

Full MS scans were acquired in the mass range of 350 – 1650 m/z with a mass resolution of 120000 and the AGC target value was set at 3e6. The 20 most intense peaks in MS were fragmented with higher-energy collisional dissociation (HCD) with collision energy of 27. MS/MS spectra were obtained with a resolution of 30000 with an AGC target of 2e5 and a max injection time of 80 ms. The Q Exactive HF dynamic exclusion was set for 40.0s and run under positive mode.

**DIA**

Full MS scans were acquired in the mass range of 350 - 1250 m/z with a mass resolution of 120000 and the AGC target value was set at 3e6. The 32 acquisition windows in MS were fragmented with higher-energy collisional dissociation (HCD) with collision energy of 28 and each acquisition window has 26 m/z. MS/MS spectra were obtained with a resolution of 30000 with an AGC target of 1e6 and a max injection time is set to auto and run under positive mode.

**Database search**

Spectronaut was used to search all of the raw data thoroughly against the sample protein database. Database search was performed with Trypsin digestion specificity. Alkylation on cysteine was considered as fixed modifications in the database searching. Protein, peptide and PSM’s false discovery rate (FDR) all set to 0.01. For DIA data, the quantification FDR also set to 0.05. Quantity MS-level was set at MS2.

**Parallel reaction monitoring (PRM)**

**Reversed-phase high performance liquid chromatography (RP-HPLC)**

RP separation was performed on an 1100 HPLC System (Agilent) using an Agilent Zorbax Extend RP column (5 μm, 150 mm×2.1 mm). Mobile phases A (2% acetonitrile in HPLC water) and B (90% acetonitrile in HPLC water) were used for RP gradient. The solvent gradient was set as follows: 0-10 min, 98% A; 10-10.01 min, 98%-95% A; 10.01-37 min, 95%-80% A; 37-48 min, 80-60% A; 48-48.01 min, 60-10% A; 48.01-58 min, 10% A; 58-58.01 min, 10-98% A; 58.01-63 min, 98% A. Tryptic peptides were separated at an fluent flow rate of 250μL/min and monitored at 210 nm. Samples were collected for 10-50 minutes, and eluent was collected in centrifugal tube 1-10 every minute in turn. Samples were recycled in this order until the end of gradient. The separated peptides were lyophilized for mass spectrometry.

**PRM Pre-scan**

The sample used for prescanning is equal mix of all samples to correct the retention time of the peptide segment, as the retention time of the entire liquid mass system may shift after a period of time, so prescanning correction must be carried out.

(1) Pre-sweep method setting

A target protein list is imported into SpectroDive and matched to the corresponding Library (i.e. the above DDA Library), the software will automatically calculate the mass/charge ratio of the theoretical peptide sequence of the target protein, and retain unique peptides that satisfy the following conditions:

I: ProteinGroup specific or Proteoctypic specific (which one is selected depends on the degree of redundancy of the FastA library used. If the degree of redundancy is very low, the latter is selected.)

II: Miss cut to zero

III: Immutable modification

IV: Charge number: 2-3

1. Export the above list and set up a mix pre-scan on the mass spectrometry method.
2. Import the pre-scanned RAW data into SpectrDive and correct the retention time according to the iRT standard peptide sections.

(4) Set the scheduled method and export.

**PRM mode detection**

The chromatographic conditions

Mobile phase A: ACN-H2O-FA (0:10, 0:0.1, V/V/V);

Mobile phase B: ACN-H2O-FA (80:20:0.1, V/V/V);

Flow rate: 300 nL/min;

Gradient elution conditions: 0~ 78min, 2-40% B; 78~80 min, 40-90% B; 80~90 min, 90% B.

**Mass spectrometry conditions:**

Import the pre-scan method into the Inclusion list of the Xcalibur-PRM method editing module. Main scanning parameters include: MS2 resolution 17500, Isolation Window 1.m /z, AGC setting as 1e5, maximum accumulated time of 100 ms, HCD energy of 28.

**Bioinformatics analysis process**

Trusted proteins and differential proteins were screened, and then functional analysis was performed on the differential proteins, including GO analysis, Pathway analysis, and protein interaction. Meanwhile, correlation analysis was performed on the differential comparison data, expression pattern clustering heat map, Venn analysis, etc. In addition, key proteins and their functions or pathways should be selected according to the data and the relevant or interested parts to conduct subsequent key research and verification.



Fig.3 Bioinformatics analysis process