

分类号:

密级:

UDC:

学号: T10411201201037

南昌大学专业学位博士研究生

# 学位论文

## 基于液相色谱质谱联用的结肠癌 FFPE 组织 蛋白组学研究

Study of proteomics on colon cancer FFPE tissue based on liquid  
chromatography-mass spectrometry

周仪华

培养单位(院、系): 南昌大学第二附属医院

指导教师姓名、职称: 程晓曙 教授 主任医师

专业学位种类: 临床医学博士

专业领域名称: 内科学(心血管病)

论文答辩日期: 2019年5月

答辩委员会主席: 辛洪波

评阅人: 盲审

2019年5月

## 一、学位论文独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得南昌大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名（手写）：周华 签字日期：2019年6月3日

## 二、学位论文版权使用授权书

本学位论文作者完全了解南昌大学有关保留、使用学位论文的规定，同意学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权南昌大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编本学位论文。同时授权北京万方数据股份有限公司和中国学术期刊（光盘版）电子杂志社将本学位论文收录到《中国学位论文全文数据库》和《中国优秀博硕士学位论文全文数据库》中全文发表，并通过网络向社会公众提供信息服务，同意按“章程”规定享受相关权益。

学位论文作者签名（手写）：周华 导师签名（手写）：李德岭

签字日期：2019年6月3日 签字日期：2019年6月3日

论文题目				
姓名	学号	论文级别	博士 <input checked="" type="checkbox"/> 硕士 <input type="checkbox"/>	
院/系/所	专业			
E_mail				
备注：				

公开 保密（向校学位办申请获批准为“保密”，\_\_\_\_年\_\_月后公开）

## 摘要

### 背景

结直肠癌（CRC）是全世界范围内癌症相关死亡率的主要来源之一。大约50%的治疗病例在首次诊断后5年内复发并导致死亡<sup>[1]</sup>。早期发现CRC为预防CRC死亡提供了最好的机会:对于转移的CRC患者，手术切除局部肿瘤后的5年生存率为90%，显著降低至10%<sup>[2, 3]</sup>。尽管目前广泛使用的大肠癌筛查技术、乙状结肠镜检查 and 结肠镜检查等可以减少CRC患者的死亡率，但是接受检查的患者数量仍然有限，部分是因为侵入性检查以及对检查的带来的不适感的恐惧<sup>[4-8]</sup>。因此，寻找新的生物标志物作为非侵入性分子检测的靶点，有助于将CRC筛查引入临床分析的常规程序。蛋白质生物标志物非常适合基于血液和粪便分析的新型体外分子检测的发展。目前，癌胚抗原(CEA)<sup>[9]</sup>和粪血红蛋白(f-Hb)是唯一被批准用于CRC临床的可溶性蛋白生物标志物。然而，CEA也可能在健康、重度吸烟者对炎症条件的反应、I型和II型糖尿病、溃疡性结肠炎、胰腺炎和肝硬化中过度表达<sup>[10]</sup>。因此，CEA试验可用于监测CRC的进展，因此可作为预后标志物<sup>[11, 12-14]</sup>，但这并不是早期检测CRC的可靠方法<sup>[15-17]</sup>。CRC最常用的筛查方式是愈创木脂化学法粪便潜血检测(gFOBT)。不幸的是，这项检测并不能检测出大多数息肉和癌症，需要多种粪便样本才能很好地解释结果。另外，gFOBT容易出现假阳性结果，这是正常良性环境的结果，如摄入某些食物和消炎药。目前正在进行广泛的研究，以寻找有用的生物标志物，改善目前CRC筛查的诊断手段，并可用于预测治疗结果，发现非侵入性早期检测方法的新工具是一个优先事项。

美国国家癌症研究所(National Cancer Institute, USA)将生物标志物或特征分子定义为血液、其他体液或组织中可客观测量的任何生物分子，可作为正常/异常生物过程或病理状态的指标。生物标志物可以揭示疾病，预测预后或预测药物或治疗干预的反应<sup>[18]</sup>。通常，一个给定的生物标志物要在临床实践中使用必须具有延长预期寿命或提高生活质量的作用。根据使用它们的目的，我们可以区分三种主要类型的生物标志物:诊断、预后和预测。任何可检测到的癌细胞在DNA、RNA、蛋白质或代谢物水平上的分子变异都可以被视为“癌症生物标志物”<sup>[19]</sup>。癌症是由于基因突变的积累导致细胞过程的改变，如血管生成、增殖、

凋亡和衰老<sup>[20,21]</sup>。因此，最初使用基因组学和转录组学方法寻找标记物<sup>[22,23]</sup>，这些方法详细阐述了癌症的遗传基础。通过选择性剪接 mRNA 并结合大量翻译后修饰，一个基因可以编码多种蛋白<sup>[24]</sup>。平均而言，一个给定的基因可以编码 4 个具有不同序列和活性的选择性剪接变异体<sup>[24]</sup>。蛋白质组比基因组更有动态性；因此，它更准确地反映了细胞的机制。通过蛋白质组学，可以分析成千上万的癌细胞蛋白，从而为 CRC 产生新的治疗靶点和标志物。在这种背景下，蛋白质组学是一种理想的、高度可翻译的研究工具，可用于发现新的癌症生物标志物。蛋白质组学方法已被广泛应用于寻找新的 CRC 生物标志物，并阐明 CRC 的分子机制；这导致了許多蛋白质的鉴定，这些蛋白质有可能被用作解决临床需要的生物标志物，采用蛋白质组学方法对不同来源的各种样品进行了分析。

如何在不违背伦理学的前提下，有效获取并保存人体肿瘤资源并用于科学研究，是尚未妥善解决的问题。低温冻存被公认为是组织与器官长期保存的最有效方法之一，其原理是通过降低细胞代谢率保护组织细胞以及细胞组分。但是，由于新鲜标本保存成本较高且复杂，难以在大多数地市级医院病理科开展，及时保存肿瘤组织是病理医师和临床医师开展临床科研工作的基础。

福尔马林固定石蜡包埋( Formalin Fixed Paraffin Embedded, FFPE) 是病理科普遍采用的组织处理和保存方法；组织标本经取材、固定、脱水、透明、浸蜡以及包埋等程序处理之后可长时间保存；与新鲜标本不同的是，FFPE 组织操作流程简单，易于掌握，可以实现高度自动化处理，而且代价低廉，适合大多数医院病理科开展。因此，FFPE 组织已广泛应用于病理科的常规病理临床工作中。与新鲜冰冻组织以及 OCT 组织相比，尽管 FFPE 组织具有这么多优势，但是其作为科研的样本劣势也非常明显，其中共价交联、低蛋白溶解率以及低肽段回收率就是目前研究人员难以突破的瓶颈。如何尽早突破该瓶颈显得特别重要。

## **第一部分 结肠癌 FFPE 组织多种蛋白提取缓冲液蛋白提取效率的比较以及蛋白定量方法**

**目的：**为了优化临床 FFPE 组织蛋白的提取效率，进一步为蛋白组学研究质谱分析提供较好的样本，在目前尚缺乏可靠、可重复的 FFPE 蛋白组学标准样品制备方法的条件下，我们拟寻求一种高效稳健的 FFPE 组织蛋白提取方法。

**方法：**基于成本、易用性以及是否适用于质谱分析，选取多种含有一定浓度的非离子、离子去污剂（十二烷基硫酸钠，SDS）和还原剂（二硫苏糖醇，

DTT) 的蛋白提取缓冲液。这些缓冲液成份经过一定的改良, 使组织的溶解度升高, 同时, 蛋白裂解过程中通过高温孵育以及超声波碎裂等方法以期最大程度地打开共价键交联。蛋白提取后, 采用 BCA 蛋白定量分析方法以及 SDS-PAGE 方法检测所提取蛋白的相对浓度, 以备下一步使用。

**结果:** 我们比较了 3 种含有不同浓度的非离子、离子去污剂 SDS 和还原剂 DTT 的蛋白提取缓冲液对结肠癌 FFPE 组织的蛋白提取效率, BCA 检测结果以及 SDS-PAGE 结果表明含有 2% SDS 的缓冲液较 SDS 浓度相对较低的缓冲液具有较高的提取效率。

**结论:** 我们通过优化和改良了 FFPE 组织蛋白提取缓冲液的组成成份和蛋白提取流程, 成功从 FFPE 组织中获得较高浓度的蛋白溶液, 并且发现含有相对更高的去污剂 SDS 的缓冲液具有更高的蛋白提取效率。

## **第二部分 结肠癌 FFPE 组织溶液内消化, 凝胶内消化以及悬浮 S-Trap 肽段回收效率的比较**

**目的:** 自下而上的蛋白质组学策略依赖于有效地将蛋白质消化成肽段进行质谱分析, 本研究将探索一种新的、强劲的 FFPE 组织蛋白消化方法以便从临床 FFPE 组织中获取可供质谱分析的组织全蛋白组。

**方法:** 溶液内消化以及凝胶内消化已经常规用于新鲜冰冻组织以及 OCT 组织的蛋白消化中, 对于 FFPE 组织, 因为化学交联和常规蛋白裂解缓冲液的低溶解度以及所提取的蛋白含有高浓度的 SDS 等原因, 目前尚无可靠稳健的蛋白消化方法。尽管近年来流行的去污剂去除和消化方法是过滤器辅助样品制备法 (FASP), 可以一定程度去除 SDS 等去污剂, 但是去实验方案的繁杂性以及实验结果的不稳定性常常阻碍其在高通量蛋白质组学研究中的应用。基于这些原因, 我们开发了新的技术来辅助基于 SDS 的蛋白质组样本的准备。

我们通过第一章描述的 FFPE 蛋白提取方法中提取结肠癌 FFPE 组织蛋白后, 分别以溶液内消化、凝胶内消化以及新近报道的 S-Trap 消化方法对蛋白样品进行消化以寻求最有效的液相色谱质谱分析样品准备方法。每种消化方法, 我们使用相同的 8 个原发性结肠腺癌的临床 FFPE 样本, 每一个样本蛋白含量为 100 $\mu$ g, 所有样品用胰蛋白酶和赖氨酸-C 混合酶双重消化。因为含有样本中含有高浓度的去污剂 SDS, SDS 与质谱分析仪不兼容, 故在溶液内消化 C18 脱盐步骤前, 采用 Pierce 去污剂旋转柱试剂盒去除残留的 SDS, 凝胶内消化以及 S-Trap 消化

方法不需要去除去污剂以及 C18 脱盐。蛋白消化成肽段后用 Pierce 定量比色法以及 Q-Exactive 质谱仪测定每个样本肽段的含量，比较三种方法的平均肽段回收率。

**结果：**Pierce 定量比色法测定肽段的含量结果提示在溶液内消化、凝胶内消化以及 S-Trap 消化方法平均肽段恢复率分别是 9.1%, 9.5% 和 93.5%。Q-Exactive 质谱分析结果显示溶液内消化方法所得的蛋白、肽段以及谱系的数量分别是：443、1126 和 1423；凝胶内消化方法所得的蛋白、肽段以及谱系的数量分别是：418、1211 和 1229；S-Trap 消化方法所得的蛋白、肽段以及谱系的数量分别是：1855、9433 和 11311。

**结论：**与传统的溶液内和凝胶内消化方法相比，新的 S-Trap 消化方法在 FFPE 样品中肽段回收率明显高于其它两种方法。S-Trap 操作方法简单，允许溶液内含有较高浓度的去污剂 SDS，肽段回收率高，为深入蛋白组学研究提供较好的质谱样本。

### **第三部分 基于液相色谱质谱联用的蛋白组学结肠癌 FFPE 组织差异蛋白的检测及肝转移过程的认识**

**目的：**本研究将利用超高分辨率傅里叶变换质谱技术对 FFPE 库存样品进行深入蛋白质组分析，为 FFPE 开发一种稳健的深入蛋白质组分析方法，并将其应用于原发性和匹配转移性结肠癌的蛋白质组差异表达蛋白的检测并了解肿瘤转移过程。

**方法：**我们选择了 18 个结肠腺癌病人共 58 个匹配有正常组织、原发性肿瘤组织和肝转移组织的样本，其中正常组织样本 18 个、原发性肿瘤组织 21 个和肝转移瘤组织 19 个。所有样本均取自美国纪念斯隆凯特琳癌症中心(Memorial Sloan Kettering Cancer Center, MSKCC) 生物样本库。蛋白提取方法见本文第 2 章结肠癌 FFPE 组织多种蛋白提取缓冲液蛋白提取效率的比较以及蛋白定量方法部分，蛋白提取缓冲液为含有 2%SDS 的缓冲液 1。质控样本从所有的 58 个样本的蛋白提取物中获取。蛋白质消化方法选用 S-Trap 法，消化酶选用胰蛋白酶/赖氨酸混合酶，详细方法见第 3 章 S-Trap 消化样本准备方法。非标记定量(LFQ)方法行液相色谱质谱联用的质谱分析 (LC-MS/MS)，质谱分析仪是最新一代的傅立叶变换 Orbitrap Fusion Lumos 高分辨质谱仪；MaxQuant 进行数据库搜索匹配蛋白质，统计学分析使用 Perseus 软件，GO 功能注释和 KEGG 通路分析使用

DAVID 等相关数据库工具。

**结果：**58 个结肠癌 FFPE 样本，共检测出 6052 种蛋白质，这些蛋白质中有 3200 种共同存在于正常组织、原发性肿瘤组织和肝转移组织中，而 373 种、923 种和 764 种蛋白在正常组织、原发性肿瘤组织和肝转移组织所特有。

GO 功能注释能发现原发性肿瘤与转移肿瘤等共有的蛋白质，更能发现转移组织特有的蛋白质。同样，KEGG 信号通路分析能发现肿瘤组织中所含有各种特异性的信号通路。

转移样本的无监督聚类分析检测到了 590 个差异表达蛋白，这些蛋白被 70% 或更多的结肠癌样本所共享。在转移灶中发现三个表达明显的聚类，聚类 1 集中有 188 个蛋白，聚类 2 集中有 118 个蛋白，聚类 3 集中有 284 个蛋白。药物与基因相互作用数据库 (DGIdb) 进一步分析表明，聚类 1 匹配有 FDA 批准的 137 种转移特异性的药物治疗靶点的 66 种上调蛋白，聚类 2 匹配有 FDA 批准的 230 种转移特异性的药物治疗靶点的 54 种上调蛋白，聚类 3 匹配有 FDA 批准的 148 种转移特异性的药物治疗靶点的 42 种上调蛋白。

**结论：**将深入蛋白质组分析应用于原发性和肝转移性结肠癌的同步组，揭示了转移中的新的特异性信号通路和簇聚类。这些结果加深了我们对癌症转移生物学性质的理解，并可能导致新的药物靶点和预后标志物。

**关键词：**结肠癌；无标记定量；质谱分析；生物标志物；治疗靶点

## **ABSTRACT**

### **Background**

Colorectal cancer (CRC) is one of the leading sources of cancer-related mortality worldwide. About 50% of treated cases relapse within 5 years of initial diagnosis and result in death. Early detection of CRC provides the best opportunity to prevent CRC death: For metastatic CRC patients, the 5-year survival rate after surgical resection of the local tumor is 90%, significantly reduced to 10%. Although colorectal cancer screening techniques, sigmoidoscopy, and colonoscopy are now widely used to reduce the mortality of CRC patients, the number of patients undergoing screening is still limited, in part because of invasive testing and fear of discomfort associated with testing. Therefore, the search for new biomarkers as targets for non-invasive molecular detection helps to introduce CRC screening into the routine procedures of clinical analysis. Protein biomarkers are well suited for the development of novel in vitro molecular assays based on blood and stool analysis. Currently, CEA and Fecal hemoglobin (f-Hb) are the only soluble protein biomarkers approved for clinical use in CRC. However, CEA may also be overexpressed in healthy, heavy smokers in response to inflammatory conditions, type I and type II diabetes, ulcerative colitis, pancreatitis, and cirrhosis. Therefore, CEA test can be used to monitor the progress of CRC and therefore can be used as a prognostic marker. But this is not a reliable method for early detection of CRC. The most commonly used screening method for CRC is the guaiacin chemical fecal occult blood test (gFOBT). Unfortunately, this test does not detect most polyps and cancers, and requires multiple stool samples to interpret the results well. In addition, gFOBT is prone to false positive results, which are the result of a normal benign environment, such as intake of certain foods and anti-inflammatory drugs. Extensive research is currently underway to find useful biomarkers, improve current diagnostic approaches to CRC screening, and can be used to predict treatment outcomes. The discovery of new tools for non-invasive early detection methods is a priority.

The National Cancer Institute (USA) defines biomarkers or characteristic

molecules as any biomolecules that can be objectively measured in blood, other body fluids or tissues and can serve as indicators of normal/abnormal biological processes or pathological states. Biomarkers can reveal disease, predict prognosis or predict response to drug or therapeutic interventions. In general, for a given biomarker to be used in clinical practice, it must have the function of extending life expectancy or improving quality of life. Based on the purposes for which they are used, three main types of biomarkers can be distinguished: diagnosis, prognosis, and prediction. Any detectable molecular variation in the DNA, RNA, protein, or metabolite levels of a cancer cell can be considered a "cancer biomarker." Cancer is caused by the accumulation of genetic mutations that lead to changes in cellular processes, such as angiogenesis, proliferation, apoptosis, and aging. Thus, the initial use of genomics and transcriptome methods to search for markers detailed the genetic basis of cancer. By selectively splicing mRNA and combining with extensive post-translational modification, one gene can encode multiple proteins. On average, a given gene can encode four selective splicing variants with different sequences and activities. The proteome is more dynamic than the genome; Therefore, it more accurately reflects the mechanism of the cell. Through proteomics, thousands of cancer cell proteins can be analyzed to generate new therapeutic targets and markers for CRC. In this context, proteomics is an ideal, highly translatable research tool for the discovery of new cancer biomarkers. Proteomics has been widely used to find new CRC biomarkers and elucidate the molecular mechanism of CRC. This has led to the identification of many proteins that may be used as biomarkers to address clinical needs, and the analysis of various samples from different sources using proteomics methods.

How to effectively obtain and preserve human tumor resources for scientific research without violating ethics is a problem that has not been properly solved. Cryopreservation is recognized as one of the most effective methods for long-term preservation of tissues and organs. The principle of cryopreservation is to protect tissue cells and cell components by reducing the metabolic rate of cells. However, due to the high cost and complexity of preservation of fresh specimens, it is difficult to develop in the department of pathology of most prefecture-level hospitals. Timely preservation of tumor tissue is the basis for pathologists and clinicians to carry out

clinical research work.

Formalin Fixed Paraffin Embedded (FFPE) is a widely used method for tissue treatment and preservation in pathology. Tissue samples can be preserved for a long time after being processed by sampling, fixation, dehydration, transparency, wax immersion and embedding. Different from fresh specimens, FFPE tissue has simple operation process, easy to master, can achieve a high degree of automation, and low cost, suitable for most hospitals to carry out the department of pathology. Therefore, FFPE tissue has been widely used in the routine pathological clinical work of pathology department. Compared with fresh frozen tissue and OCT tissue, FFPE tissue has many advantages, but it also has obvious disadvantages as a research sample. Covalent cross-linking, low protein solubility and low peptide recovery rate are the bottlenecks that researchers are unable to break through at present. How to break through the bottleneck as soon as possible seems particularly important.

### **Part I Comparison of protein extraction efficiency and protein quantification method of FFPE protein extraction buffer in colon cancer**

**Objective:** In order to optimize the extraction efficiency of FFPE tissue protein and provide a better sample preparation for the mass spectrometry analysis of proteomics, In the absence of a reliable and reproducible FFPE proteomics standard sample preparation method, we intend to seek an efficient and robust FFPE tissue protein extraction method.

**Methods:** Based on the cost, ease of use and whether it is suitable for mass spectrometry analysis, a variety of protein extraction buffers containing a certain concentration of non-ionic, ion stain remover (sodium dodecyl sulfate, SDS) and reductant (dithiothreitol, DTT) were selected. The components of these buffers were modified to increase the solubility of the tissue. Meanwhile, during protein lysis, methods such as high-temperature incubation and ultrasonic fragmentation were used to maximize the opening of covalent bond cross-linking. After protein extraction, BCA protein quantitative analysis method and SDS-PAGE method were used to

detect the relative concentration of the extracted protein for the next step.

**Results:** We compared the protein extraction efficiency of three protein extraction buffers containing different concentrations of non-ionic, ion stain remover SDS and reducer DTT on FFPE tissue of colon cancer. BCA test results and SDS-PAGE results showed that the buffer containing 2% SDS had higher extraction efficiency than the buffer with lower SDS concentration.

**Conclusions:** By optimizing and improving the components of FFPE tissue protein extraction buffer and protein extraction process, we successfully obtained a relatively high concentration of protein solution from FFPE tissue, and found that the buffer containing relatively higher detergent SDS has a higher protein extraction efficiency.

## **Part II Comparison of the efficiency of FFPE digestion in solution, gel digestion and suspension S-trap peptide recovery**

**Objective:** The bottom-up proteomics strategy relies on the efficient digestion of proteins into peptides for mass spectrometry, In this study we will develop a new and robust FFPE tissue digestion method for obtaining a whole proteome from clinical FFPE tissues for mass spectrometry analysis.

**Methods:** In-solution digestion and in-gel digestion have been routinely used for protein digestion in fresh frozen tissue and OCT tissue. For FFPE tissues, due to chemical cross-linking and the low solubility of conventional protein lysis buffer and the high concentration of SDS in the extracted protein, there is no reliable and robust protein digestion method at present. Although in recent years, the popular method of detergent removal and digestion is the Filter Assistant Sample Preparation method (FASP), which can remove SDS and other detergents to a certain extent, the complexity of the experimental protocol and the instability of the experimental results often hinder its application in high-throughput proteomics research. For these reasons, we have developed a new technique to assist the preparation of SDS-based proteome samples. We extracted the FFPE tissue protein of colon cancer from the FFPE protein extraction method described in the first chapter, and then digested the protein samples

in solution, in gel and the newly reported S-trap digestion method to seek the most effective sample preparation method by liquid chromatography mass spectrometry. For each digestion method, we used the same 8 clinical FFPE samples of primary colorectal adenocarcinoma, each sample had a protein amount of 100  $\mu$ g, and all samples were digested with trypsin and lysin-c mixed enzyme. Because the sample contains a high concentration of detergent SDS, which is incompatible with the mass spectrometer, the residual SDS is removed by Pierce detergent spin column before the C18 desalting step is digested in the solution. In gel digestion and S-trap digestion methods do not need detergent removal and C18 desalination step. After the protein was digested into peptides, the amount of each sample was determined by Pierce quantitative colorimetry peptide assay and Q-exactive mass spectrometer, the average peptide recovery rates of the three methods were compared.

**Results:** The results of Pierce quantitative colorimetric determination of peptides suggest that the average recovery rates of peptides in solution, in gel and S-trap digestion are 9.1%, 9.5% and 93.5%, respectively. The results of Q-exactive mass spectrometry showed that the number of protein, peptide and spectra obtained by in-solution digestion were 443, 1126 and 1423, respectively. The number of protein, peptide and spectra obtained by digestion in gel were 418, 1211 and 1229, respectively. The number of protein, peptide and spectra obtained by S-Trap digestion method were 1855, 9433 and 11311, respectively.

**Conclusions:** Compared with the traditional in-solution and in-gel digestion methods, the new S-trap digestion method has a significantly higher recovery rate of peptides in FFPE samples than the other two methods. The simple operation method of S-Trap allows the solution to contain a relatively high concentration of detergent SDS, and the peptide recovery rate is high, providing a good mass spectrum sample for in-depth proteomics research.

### **Part III Detection of differentially expressed protein and understanding of liver metastasis process for colon cancer FFPE tissue based on proteomics by liquid chromatography-mass spectrometry**

**Objective:** In this study, FFPE archive samples will be analyzed by ultra-high

resolution Fourier transform mass spectrometry for in-depth proteomics analysis. A robust in-depth proteome assay was developed for FFPE and it would be applied to the detection of differential expression protein matching primary and metastatic colon cancer and to understand the process of tumor metastasis.

**Methods:** We selected a total of 58 samples from 18 patients with colorectal adenocarcinoma matched with normal, primary tumor and liver metastatic tissue, including 18 samples of normal, 21 samples of primary tumor and 19 samples of liver metastatic tissue. All samples were taken from the biobank of Memorial Sloan Kettering Cancer Center (MSKCC). For protein extraction methods, please refer to the comparison of protein extraction efficiency and protein quantitative methods of FFPE buffer for multiple protein extraction from colon cancer tissues in chapter 2. The protein extraction buffer was buffer 1 containing 2%SDS. Quality control samples were obtained from pooled samples. S-trap method was used for protein digestion, and trypsin/lysine mixed enzyme was used for digestive enzymes. For detailed methods, see chapter 3 S-trap method for sample preparation. Label free quantitative (LFQ) method was used to perform mass spectrometry (LC/MS/MS). The mass spectrometer is the latest generation of the Fourier transform Orbitrap Fusion Lumos. MaxQuant searched the database for matching proteins, Perseus software was used for statistical analysis, and DAVID and other related database tools were used for GO functional annotation and KEGG pathway analysis.

**Results:** A total of 6,052 proteins were detected in 58 colon cancer FFPE samples, and 3,200 of these proteins were present in normal tissues, primary tumor tissues and liver metastatic tissues. However, 373, 923 and 764 proteins were specific to normal tissues, primary tumor tissues and liver metastatic tissues.

GO functional annotation can find common proteins such as primary tumor and metastatic tumor, and more specifically proteins specific to metastatic tissue. Similarly, KEGG signaling pathway analysis revealed various specific signaling pathways in tumor tissues. Unsupervised cluster analysis of the metastatic samples detected 590 differentially expressed proteins, which were Shared by 70% or more of the colon cancer samples. Three clusters with obvious expression were found in the

metastases, with 188 proteins in cluster 1, 118 proteins in cluster 2 and 284 proteins in cluster 3. Further analysis of the drug-gene interaction database (DGIdb) showed that Cluster 1 matched 66 upregulated proteins with 137 FDA approved drug treatment targets, cluster 2 matched 54 upregulated proteins with 230 FDA approved drug treatment targets and cluster 3 matched 42 upregulated proteins with 148 FDA approved drug treatment targets with metastasis specificity.

**Conclusions:** Deep proteomic analysis was applied to the synchronous group of primary and liver metastatic colon cancer, revealing new specific signaling pathways and clustering in metastasis. These results enhance our understanding of the biological nature of cancer metastasis and may lead to new drug targets and prognostic markers.

**Key words:** colon cancer; Unlabeled quantification; Mass spectrometry; Biomarkers; Therapeutic targets

## 目 录

第 1 章 引 言.....	1
第 2 章 结肠癌 FFPE 组织多种蛋白提取缓冲液蛋白提取效率的比较以及蛋白定量方法.....	4
2.1 前言.....	4
2.2 材料.....	5
2.2.1 结肠癌 FFPE 组织的获取.....	5
2.2.2 购置试剂.....	5
2.2.3 主要仪器设备及耗材.....	6
2.3 方法.....	6
2.3.1 人结肠癌 FFPE 组织蛋白提取方法.....	6
2.3.2 二喹啉甲酸 (BCA) 蛋白定量分析试剂盒蛋白浓度检测方法.....	7
2.3.3 1D SDS-PAGE 检测蛋白相对含量的方法.....	8
2.4 结果.....	9
2.4.1 不同缓冲液蛋白提取效率的比较.....	9
2.4.2 二喹啉甲酸 (BCA) 蛋白定量分析结果.....	9
2.4.3 SDS-PAGE 蛋白定量分析结果.....	10
2.5 讨论.....	10
2.6 结论.....	12
第 3 章 结肠癌 FFPE 组织溶液内消化, 凝胶内消化以及悬浮 S-Trap 肽段获取效率的比较.....	13
3.1 前言.....	13
3.2 材料.....	14
3.2.1 购置试剂.....	14
3.2.2 主要仪器设备及耗材.....	15
3.3 方法.....	15
3.3.1 溶液内消化样本准备方法.....	15

3.3.2 Pierce®去污剂移除方法.....	16
3.3.3 C18 脱盐方法.....	16
3.3.4 凝胶内消化样本准备方法.....	17
3.3.5 S-Trap 消化样本准备方法.....	18
3.3.6 Pierce™定量比色肽分析方法.....	19
3.4 结果.....	21
3.4.1 Pierce™定量比色肽分析方法鉴定各个样品肽段的含量.....	21
3.4.2 LC-MS/MS 分析方法鉴定各个样品蛋白含量.....	21
3.5 讨论.....	23
3.6 结论.....	24
第 4 章 基于液相色谱质谱 (LC-MS/MS) 联用的结肠癌 FFPE 组织差异表达蛋白的检测及肝转移过程的认识.....	25
4.1 前言.....	25
4.2 材料.....	26
4.2.1 购置试剂.....	26
4.2.2 主要仪器设备及耗材.....	27
4.3 方法.....	27
4.3.1 结肠癌 FFPE 组织 LC-MS/MS 样本的准备.....	27
4.3.2 FFPE 结肠癌组织 LC-MS/MS 分析.....	27
4.3.3 蛋白质鉴定和定量.....	28
4.3.4 统计学分析.....	28
4.3.5 差异表达蛋白的生物信息学分析方法.....	28
4.4 结果.....	29
4.4.1 各组样本质谱分析获取蛋白情况分析.....	29
4.4.2 不同样本间的稳健性、精确性和再现性比较以及数据标准化... 30	30
4.4.3 各组样本差异表达蛋白以及上调蛋白的聚类情况分析.....	31
4.4.4 结肠癌肝转移过程的探索.....	33
4.5 讨论.....	38
4.6 结论.....	40

目录

---

第 5 章 全文总结与展望.....	41
致 谢.....	43
参考文献.....	44
攻读学位期间的研究成果.....	51
综 述.....	52

## 中英文缩略词表

英文缩写	英文全称	中文全名
CRC	Colorectal cancer	结直肠癌
FFPE	Formalin Fixed Paraffin Embedded	福尔马林固定石蜡包埋
CEA	Carcinoembryonic antigen	癌胚抗原
SDS	Sodium Dodecyl Sulfonate	十二烷基磺酸钠
DTT	Dithiothreitol	二硫苏糖醇
IAA	Iodoacetamide	碘乙酰胺
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis	十二烷基硫酸钠聚丙烯酰胺 凝胶电泳
BCA	Bicinchoninic acid	二喹啉甲酸
S-Trap	Suspension Trap	悬浮捕获
OCT	Optimal cutting temperature compound	最佳切削温度化合物
FASP	Filter-aided sample preparation	过滤辅助样品制备
LFQ	Label Free Quantitation	非标记蛋白定量
LC-MS/MS	liquid chromatography-tandem mass chromatography	液相色谱-串联质谱
DAVID	Database for annotation, Visualization and Integrated Discovery	注释、可视化和集成发现的 数据库
GO	Gene Oncology	基因本体论
KEGG	Kyoto Encyclopedia of Genes and Genomes	京都基因与基因组百科全书
DGIdb	Drug-Gene Interaction database	药物基因相互作用数据库
PPI	Protein-protein interaction	蛋白蛋白相互作用
iTRAQ	Isobaric tags for relative and absolute quantitation	相对和绝对定量同位素标记
AGC	Automatic gain control	自动增益控制

中英文缩略词表

---

DEPs	Differentially expressed protein	差异表达蛋白
SILAC	Stable Isotope Labeling by/with Amino acids in Cell culture	在细胞培养中通过氨基酸进行稳定同位素标记
mTRAQ	Non-isobaric amine labeling	非异压胺标记
ICAT	Isotopic coding of affinity markers	使用同位素编码亲和标记
ICPL	Dimethyl labeling and isotope coded protein labeling	甲基标记和同位素编码蛋白质标记

---

## 第1章 引言

结肠癌是人类高发的恶性肿瘤，在过去的十年中虽然致死率有所下降，但是仍然是全球范围内的癌症相关死亡的第四个主要原因。每年仍有超过一百万的新增病例，已有五十万的死亡病例。在我国，结肠癌的死亡率已经成为位列肿瘤死亡率的第五位，已经严重危害人类的健康。结肠癌发病的原因很多，与遗传、饮食、环境等因素密切相关，同时，结肠癌等发生也涉及到多种癌基因多个阶段的过程，是一个多基因多过程协同作用的结果。结肠组织发生癌变的可能机制包括：(1) 癌基因的激活或者过度表达；(2) 抑癌基因的突变或丢失；(3) 信号转导调控紊乱；(4) 微卫星不稳定；(5) 凋亡机制发生障碍；(6) DNA损伤修复相关基因功能丧失；(7) 端粒酶过度表达等。目前，尽管有关结肠癌等研究取得了进步，许多先进的医疗设备、诊断及治疗技术应用到结肠癌的诊断及治疗中，但是近年来结肠癌的发病率及致死率仍居高不下，究其原因，可能有以下三个方面：(1) 结肠癌的生物特性比较复杂，恶性程度相对较高；(2) 结肠癌的发现一般比较晚，难以在早起发现，多数病人在发现时已经处于癌症晚期；(3) 目前的一些治疗结肠癌的手段如放疗、化疗以及手术治疗都难有新的突破。所以，结肠癌的早期诊断以及新的治疗手段仍需加强。

尽管目前关于大肠癌发生方面已经做了许多研究，但是有助于判断预后的生物指标及可利用的治疗靶基因仍然很少。因此，应用一种高通量技术同时检测大肠癌中多个分子的变化，能够提高其候选分子生物标志物的筛选率。已有的结直肠癌标志物研究大多是直接针对血清样本或只对肿瘤发展的某一个阶段进行肿瘤标志物的研究，由于血清样本的限制，一些重要的低丰度蛋白难以得到有效鉴定；而只选择某一阶段检测，不能对肿瘤发展进程中蛋白质的变化进行整体评估，会降低有效候选标志物的占比，给后续验证工作带来难度，不利于标志物的发现。近年来医学研究发展的新技术包括了各种“组学”研究技术，其中蛋白组学正成为生物医学研究领域的新热点。

蛋白质组学 (proteomics) 是以蛋白质组为研究对象，在蛋白质水平对疾病机理、细胞模式、功能联系等方面进行探讨，具有定量性、动态性、整体性等特点。通过研究相关蛋白质在肿瘤中的表达和功能，整体、动态、多角度的研究肿瘤发生发展过程中的蛋白质组的变化，筛选出与其发生发展相关的特征性

蛋白，可以发现具有特异性和较高敏感性的肿瘤标记物，有助于肿瘤的早期诊断、治疗以及预后判断。蛋白质组学常用研究方法主要包括蛋白质的分离，鉴定及生物信息学技术等。蛋白质的分离技术主要有双向凝胶电泳技术、差异凝胶电泳技术、高效液相色谱及毛细管电泳等；蛋白质鉴定技术主要包括质谱技术、色谱与质谱联用技术、同位素亲和标签技术化及 iTRAQ 技术。质谱法检测蛋白质的原理是蛋白质经过酶切消化成肽段，质谱仪根据肽段所带电荷不同的质核比，确定蛋白质的一级质谱图；选取目标肽段进行二级质谱分析，通过质谱图进行对比而鉴定出蛋白质。生物信息学技术利用相关蛋白质组数据库，对生物信息进行储存、检索和分析，在蛋白质的结构和功能预测、序列比对、蛋白质分析、系统发育分析等方面发挥越来越重要的作用。

生物信息学方法可对差异表达蛋白进行分析，有助于发现相关差异蛋白及其通路。GO 分析是 gene ontology 的简称，GO 数据库分别从功能、参与的生物途径及细胞中的定位对基因产物进行了标准化描述，即对基因产物进行简单注释，通过 GO 富集分析可以粗略了解差异基因富集在哪些生物学功能、途径或者细胞定位。GO 分析实现了各种与基因相关数据的转换、统一以及挖掘，可以使各个参与合作的数据库间进行有效的查询<sup>[25]</sup>。GO 分析有 3 个相对独立的本体论，分别为：细胞成分、生物学过程和分子功能。KEGG (Kyoto Encyclopedia of Genes and Genomes) 数据库能对基因功能进行系统分析，将基因组信息与更高阶的功能信息联系起来<sup>[26]</sup>。DAVID 是一个富集分析工具，能分析网络中差异蛋白参与的 GO 功能和 KEGG 通路<sup>[27]</sup>。STRING(Search Tool for the Retrieval of Interacting Genes)数据库能预测分析差异表达基因编码的蛋白质间相互作用关系<sup>[28]</sup>。Cytoscape 软件则被用来构建 PPI (protein-protein interaction) 网络图<sup>[29]</sup>。

在病理科中主要有三种不同类型的组织标本类型：福尔马林固定石蜡包埋 (FFPE)组织、液氮中储存的新鲜冰冻组织以及 OCT 复合冷冻组织。FFPE 组织标本是临床研究的一个潜在的有价值的标本来源，因为这些标本广泛储存在全球范围内各大医院样本库中，可以在短时间内收集大量样本，且这些样本都与患者的临床病史密切相关。因此，FFPE 组织在蛋白质生物标志物发现研究中的应用越来越受到关注。然而，尽管文献中关于 FFPE 蛋白组学的研究越来越多，但由于 FFPE 组织样本含有共价蛋白交联、低蛋白溶解率以及比较低的肽段回收率，目前尚缺乏有效的 FFPE 蛋白组学研究方法，包括 FFPE 组织蛋白的提取、肽段的获取以及高效的质谱分析结果。因此，急需开发一种稳健可靠的 FFPE 组织蛋

白组学质谱分析样本的方法，通过这种方法可获得高效、可重复提取全长 FFPE 蛋白，并制定标准流程。

本研究将以结肠癌患者 FFPE 组织为研究对象，优化结肠癌 FFPE 组织蛋白的提取、消化以及基于液相色谱质谱联用的蛋白组学研究方法以期发现新的生物标志物并了解肿瘤转移的生物过程。

## 第 2 章 结肠癌 FFPE 组织多种蛋白提取缓冲液蛋白提取效率的比较以及蛋白定量方法

### 2.1 前言

生物标志物对提高诊断水平、对患者分层进行靶向治疗和监测对各种疾病的治疗反应有很高的需求。为了找到这样的生物标志物，组织经常被用于组学领域对生物分子的识别和定量。由于与疾病相关的蛋白质和多肽(以及代谢物)在病变组织内部或附近的浓度明显较高，因此当需要了解疾病相关过程时，通常首选基于组织的蛋白质组学<sup>[30]</sup>。因此，对新鲜冷冻的临床样本的转化研究需求量很大。然而，这些新鲜组织标本的有限可用性成为生物标志物发现研究的一个严重缺陷<sup>[31]</sup>。此外，长期保存这些新鲜冷冻样本的成本很高，在短时间内收集大量样本也很困难<sup>[32]</sup>。近年来，福尔马林固定石蜡包埋(FFPE)组织作为世界范围内长期保存组织的黄金标准，被认为是一种有价值的新鲜组织替代品<sup>[32,33,34]</sup>。由于 FFPE 组织是病理研究的常规组织，所以医院档案中有上百万的样本，可以在短时间内收集到大量的样本<sup>[35,36]</sup>。此外，存档的资料具有与病理、临床和结果信息相关联的优势，因此，既适合前瞻性的临床研究，也适合回顾性的临床研究<sup>[30,36,37]</sup>。

不幸的是，十多年来，人们一直认为 FFPE 组织不适合质谱分析为基础的蛋白质组学研究，因为甲醛诱导分子内和分子间的蛋白质(以及 DNA/RNA)交联，这种交联实际上是在组织内捕获蛋白质和其他分子<sup>[38-43]</sup>。这种交联过程的不可逆性阻碍了全长蛋白的提取，使得基于质谱的蛋白质鉴定具有很大的挑战性。1991 年，随着抗原回收(AR)技术的发展，该技术被用于免疫组化研究中抗原的恢复，提取程序得到了改进。Ikeda 等(1998)是第一个利用 AR 技术成功从 FFPE 组织中提取蛋白质的人<sup>[44]</sup>。最近，一些研究小组开发了他们自己的蛋白质提取方法，使他们能够获得 FFPE 蛋白质组<sup>[45-50]</sup>。

近年来，许多 FFPE 蛋白组学研究发表在文献上。本 FFPE 研究主要分为三类:研究 FFPE 蛋白提取方法的研究，常伴有新鲜冷冻组织与 FFPE 组织的比较<sup>[50-55]</sup>，确定 FFPE 组织在生物标志物研究中的适用性的研究<sup>[56-60]</sup>，以及寻找(前)分析因素影响的研究<sup>[61-65]</sup>。所有这些研究表明，不同的研究小组使用不同的方法。

这使得严格的比较变得复杂，并得到了不同的结果。因此，对具有相同蛋白质组工作流程的多个蛋白质提取缓冲液进行严格的比较是非常必要的。本研究采用结肠癌 FFPE 组织在同等条件下比较文献报道成功率较高的 3 种不同蛋白提取缓冲液的蛋白提取效率，蛋白提取后分别以 BCA 蛋白测定法以及 SDS-PAGE 分别检测提取蛋白的浓度，以制定 FFPE 组织蛋白组学的蛋白提取标准流程，为后续成功的蛋白组学研究奠定良好的基础。

## 2.2 材料

### 2.2.1 结肠癌 FFPE 组织的获取

结肠癌 FFPE 组织取自美国纪念斯隆凯特琳癌症中心（Memorial Sloan Kettering Cancer Center, MSKCC）生物样本库。

### 2.2.2 购置试剂

Tris-HCl 缓冲液	美国 Sigma 公司
十二烷基硫酸钠	美国 Sigma 公司
碘乙酰胺	美国 Sigma 公司
1, 4-二硫代苏糖醇	美国 Sigma 公司
丙三醇	美国 Fisher Chemical 公司
二甲苯	美国 Sigma 公司
1%蛋白酶抑制剂	美国 Thermo Scientific™公司
乙基苯基聚乙二醇	美国 Bufferad 公司
3-[3-(胆酰胺丙基)二甲氨基]丙磺酸内盐	加拿大 MEMD Millipore 公司
脱氧胆酸钠	美国 Sigma 公司
4xLaemmli 样本缓冲液	美国 Bio-Rad 公司
考马斯亮蓝染色剂	美国 Bio-Rad 公司
10xTris-Glycine-SDS 缓冲液	美国 Bio-Rad 公司
质谱级 W6-1 去离子水	美国 Fisher Chemical 公司
酒精	美国 Fisher Chemical 公司
BCA 试剂盒	美国 Thermo Scientific™公司
Mini-PROTEAN®TGX™胶	美国 Bio-Rad 公司

### 2.2.3 主要仪器设备及耗材

Thermomixer 恒温混匀仪	德国 Eppendorf 公司
Sorvall™ Legend™ Micro 21 微量离心机	美国 Thermo Scientific™公司
真空浓缩器	美国 Thermo Scientific™公司
5430R Eppendorf 离心机	德国 Eppendorf 公司
超声波破碎仪	美国 Thermo Scientific™公司
-80℃冰箱	美国 Thermo Scientific™公司
-20℃冰箱	美国 Thermo Scientific™公司
4℃冰箱	美国 Thermo Scientific™公司
移液器	德国 Eppendorf 公司
电子天平	美国 Mettler Toledo 公司
PCR 微型管	德国 Eppendorf 公司
定转速涡流混合器	美国 Thermo Scientific™公司
电泳仪	美国 Bio-Rad 公司
摇床	美国 Thermo Scientific™公司

## 2.3 方法

### 2.3.1 人结肠癌 FFPE 组织蛋白提取方法

(1) 每个样本通过切片获得 10mm 厚度的 FFPE 组织切片, 该切片由 MSKCC 生物样本库工作人员完成。

(2) 每个 FFPE 样品在二甲苯中脱蜡 10 分钟, 在常温下 12000g 转速离心 3 分钟。

(3) 此孵育和离心步骤重复两次。

(4) 然后用乙醇梯度从 100%到 95%到 70%对组织颗粒进行再水化, 每次 3 分钟。

(5) 接下来, 将组织颗粒风干, 悬浮在 150 ml 蛋白提取缓冲液中, 在超声波破碎仪中超声碎裂 3 分钟。

(6) 随后, 在 Thermomixer 恒温混匀仪 98℃ 下孵育 20 分钟, 20 分钟后迅速转移至 80℃ 孵育 2 小时。

(7) 然后, 在 14000g 和 4℃ 下离心 30 分钟。

(8) 收集上清液至新的 Eppendorf PCR 离心管并存储至-80°C 冰箱备用。

### 2.3.2 二喹啉甲酸 (BCA) 蛋白定量分析试剂盒蛋白浓度检测方法

#### 2.3.2.1 梯度稀释牛血清白蛋白 (BSA) 标准品的制备

取 8 个 PCR 微型试管按照表 2.1 方法制定蛋白标准品。

表 2.1 蛋白质标准品的制备 (检测范围= 20-2,000 $\mu\text{g}/\text{mL}$ )

编号	稀释液的体积 ( $\mu\text{l}$ )	BSA 的体积和来源	BSA 终浓度 ( $\mu\text{g}/\text{ml}$ )
A	0	300 $\mu\text{l}$ 原液	2000
B	125	375 $\mu\text{l}$ 原液	1500
C	325	375 $\mu\text{l}$ 原液	1000
D	175	175 $\mu\text{l}$ B 瓶稀释液	750
E	325	325 $\mu\text{l}$ C 瓶稀释液	500
F	325	325 $\mu\text{l}$ E 瓶稀释液	250
G	325	325 $\mu\text{l}$ F 瓶稀释液	125
H	400	100 $\mu\text{l}$ G 瓶稀释液	25
I	400	0	0=空白

#### 2.3.2.2 BCA 工作液 (WR) 的制备

(1) 使用下述公式来确定所需的工作液的总体积:

(标准品的个数 + 待测蛋白质样品的个数)  $\times$  (实验重复次数)  $\times$  (用于每个样品的工作液的体积) = 所需的工作液总体积。

(2) 将 50 份 BCA 试剂 A 与 1 份 BCA 试剂 B 混合 (试剂 A 与试剂 B 的比率=50:1), 制备工作液。

将 5mL 试剂 A 和 0.1mL 试剂 B 混合, 当试剂 B 加入到试剂 A 中时, 开始可观察到有浑浊产生, 经搅拌后浑浊会迅速消失, 得到绿色澄清工作液。

#### 2.3.2.3 蛋白浓度检测方法

(1) 取各个稀释浓度的蛋白质标准品和待测蛋白质样品各 25 $\mu\text{L}$ , 加入到 1.5ml Eppendorf Lobind 试管中 (检测范围=20-2000 $\mu\text{g}/\text{mL}$ )。

(2) 在每一个蛋白标准品和待测蛋白质样品中加入 200 $\mu$ L 工作液，并在振荡器上振荡 30 秒，使其充分混合。

(3) 将试管密封，在 Thermomixer 恒温混匀仪 37 $^{\circ}$ C 孵育 30 分钟。

(4) 将试管冷却到室温，使用 Nanodrop One 分光光度计测量样品在 562nm 或该波长附近的吸光值。

### 2.3.3 1D SDS-PAGE 检测蛋白相对含量的方法

#### 2.3.3.1 所需仪器及主要溶液：

(1) 4-20% Mini-PROTEAN 预制胶，最佳运行分子量范围：10-200kD。

(2) SDS-PAGE 电泳槽。

(3) 1X 电泳缓冲液：100ml 10X 电泳缓冲液（成份：25 mM Tris, 192 mM glycine, 0.1% SDS）加入 900ml 去离子水。

(4) 4X Laemmli 上样缓冲液：成份：250 mM Tris-HCl, pH 6.8, 4% LDS, 40% (w/v) glycerol, 0.02% bromophenol blue, 和 200 mM DTT (新鲜配制)。

#### 2.3.3.2 1 D SDS-PAGE 操作流程

(1) 取下梳子:拇指置于梳子压痕处(梳子中间)，取下梳子以一个平稳的动作向上拉。

(2) 拆下胶带:轻拉，将胶带从盒式磁带底部取出。

(3) 冲洗井: 使用注射器、冲洗瓶或一次性转移管，用流动的缓冲液冲洗井。

(4) 将电极组件打开放置在干净、平整的表面上

(5) 将凝胶盒放入电极中组装。

(6) 轻轻地把两种凝胶推向对方，同时轻轻挤压凝胶盒(或卡带和缓冲坝)对着绿色垫圈(保持恒压和两种凝胶都到位)，滑动绿色的手臂的夹紧框架，一次一个以上凝胶，锁住它们电极组件的翼夹抬起每个凝胶盒靠在缺口上绿色垫圈，形成密封。

(7) 将电泳模块放入电泳槽中用 1X 电泳缓冲液填充缓冲室：内缓冲室 200ml，或在外部缓冲室 800ml。

(8) 用流动的缓冲液清洗样品井。

(9) 取 5 $\mu$ l 待测蛋白溶液加入 4.25 $\mu$ l 含有 DTT 的上样缓冲液，用去离子水稀释到 18 $\mu$ l（上样缓冲液与样本最后体积比为 1:3），95 $^{\circ}$ C 加热 5 分钟。

(10) 样本冷却后用移液器轻轻地将样本加入到电泳孔，注意避免气泡的产生，运行凝胶使用运行。

(11) 运行条件：电流 55mA，电压 200mV，运行时间：30 分钟。

(12) 电泳完成后，关闭电源，断开电源线。

(13) 打开盒式磁带，轻轻取出凝胶。

### 2.3.3.3 考马斯亮蓝胶染色、脱染及凝胶成像

(1) 从盒子里取下 SDS-PAGE 凝胶，在去离子水中冲洗一次，放入有盖的容器中，尽量不要使用比凝胶大或小得多的容器。

(2) 将装有去离子水覆盖凝胶的盒子放在摇床上轻轻搅拌，固定凝胶 15 分钟，期间每 5 分钟换去离子水一次。

(3) 吸尽盒子里的去离子水，在凝胶中加入 QC 胶体考马斯亮蓝染色剂 50ml，室温下在摇床上轻轻搅拌孵育一晚，注意盖上容器盖子避免染色剂挥发。

(4) 次日从摇床上取出盒子，用去离子水将考马斯亮蓝染色剂冲洗干净，装上适量去离子水继续室温摇床上室温孵育 3 小时以脱染，期间至少换水 3 次。

(5) 取出凝胶，在 BIO-RAD ChemiDoc™ MP 成像系统摄片，Image-Lab6.01 (Bio-Rad) 软件计算蛋白相对含量。

## 2.4 结果

### 2.4.1 不同缓冲液蛋白提取效率的比较

为了建立 FFPE 组织蛋白组学标准操作规程，选取 3 种含有一定浓度的去污剂 SDS 和还原剂 DTT 的蛋白提取缓冲液，这些缓冲液的选择是基于提取效率、成本、易用性以及是否适用于质谱分析而确定的。比较了 3 种蛋白提取缓冲液的蛋白提取效率。蛋白提取后，采用 BCA 蛋白定量分析方法以及 SDS-PAGE 方法检测蛋白的相对浓度。

### 2.4.2 二喹啉甲酸 (BCA) 蛋白定量分析结果

将 3 种不同蛋白提取缓冲液获得的蛋白提取溶液，采用 BCA 蛋白浓度测定法，具体操作步骤见实验方法部分，缓冲液成份以及不同缓冲液提取的蛋白浓度 BCA 检测结果如表 2.2 所示，从该结果不难看出，缓冲液 1 BCA 测定值明显高于缓冲液 2 和 3。

表 2.2 缓冲液成份以及不同缓冲液提取的蛋白浓度 BCA 检测结果

缓冲液	成份	BCA 值 ( $\mu\text{g}/\mu\text{L}$ )
1	20mM 三羟甲基氨基甲烷, 2%十二烷基硫酸钠, 200mM 二硫苏糖醇, 20%丙三醇, 1%蛋白酶抑制剂, pH 8.8 (Addis et al. 2009 <sup>[43]</sup> 已修正)	11.2
2	RIPA 裂解缓冲液, 2%十二烷基硫酸钠, pH 8 (Hwang et al.,2007 <sup>[44]</sup> )	2.7
3	20mM 三羟甲基氨基甲烷, 2%十二烷基硫酸钠, 1.5% 3-[(3-胆固醇氨基)二甲基氨基]-1-丙磺酸, 200mM 二硫苏糖醇, 10%丙三醇,pH 8.8 (Addis et al.2009 <sup>[43]</sup> )	3.8

### 2.4.3 SDS-PAGE 蛋白定量分析结果

为了准确测定各种蛋白提取缓冲液 FFPE 蛋白提取物的浓度, 增加实验的可靠性, 我们在 BCA 检测的基础上对蛋白提取物进行 SDS-PAGE 检测。实验结果用 Image Lab 6.01(Bio-Rad)软件检测。从图 1 可以发现, 缓冲液 1 的蛋白提取效率明显高于缓冲液 2 和 3, 说明缓冲液 1 为 FFPE 组织蛋白提取最佳缓冲液。该结果与 BCA 检测结果一致。

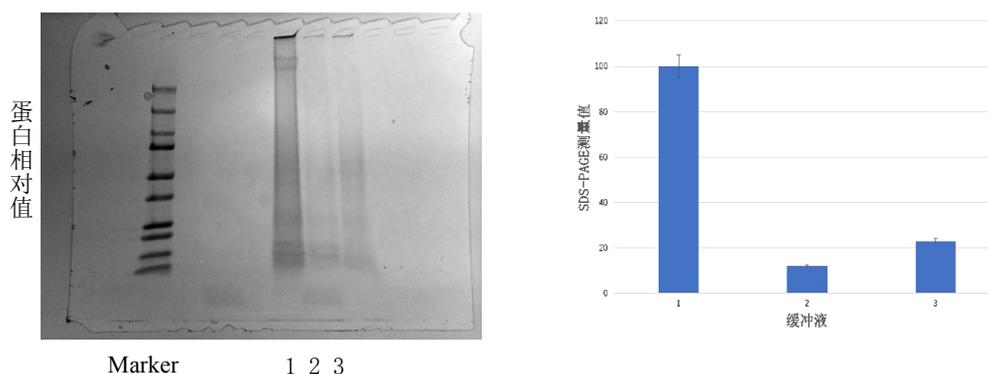


图 2.1 不同缓冲液蛋白相对浓度的 SDS-PAGE 检测结果

## 2.5 讨论

目前没有可靠稳健的 FFPE 组织蛋白组学研究方法, 故对于生物医学项目来

说 FFPE 组织作为研究对象并不是优先考虑的。FFPE 组织蛋白组学研究的一个主要挑战是几乎所有现有的研究方案均不相同,包括蛋白提取、消化以及质谱分析等等,没有严格的操作流程。在本研究中,我们从文献报道中<sup>[43-45,66,67]</sup>选取了3种含有相对较高浓度的离子型去污剂 SDS 和还原剂 DTT 且提取效率较高适用于 FFPE 组织以及质谱分析的蛋白提取缓冲液,经过一定改良后比较各缓冲液的蛋白提取效率,以期待得出标准的 FFPE 组织蛋白提取流程,为下一步深入蛋白组学研究提供基础。

本研究采用两种蛋白质浓度测定方法:BCA 蛋白定量分析法以及 SDS-PAGE 法。Valerie 等认为蛋白质浓度的误差幅度似乎与蛋白质提取缓冲液的组成有关。通过对3种不同蛋白提取缓冲液比较,我们可以指出哪种缓冲液成分组合是最优的,以便从 FFPE 组织中提取尽可能多的全长 FFPE 蛋白。在人类结肠癌 FFPE 组织中,蛋白提取缓冲液 1 一直被认为是最有效的蛋白提取缓冲液,该蛋白提取缓冲液 1 含有 2% SDS, 20% 甘油, pH 值为 8.8。Magdeldin 和 Yamamoto 证明,与不含 SDS 的提取缓冲液相比,含有 2% SDS 的蛋白质提取缓冲液从 FFPE 组织中提取的蛋白质要多 15 倍,这种提取效率可以归因于 SDS 作为洗涤剂和蛋白质变性剂的双重作用<sup>[68]</sup>。此外,其他研究表明,碱性(pH 8.0-9.5)三氯化氢缓冲液比中性或酸性缓冲液提取蛋白质的质量更高<sup>[69,70]</sup>。因此,我们认为碱性培养基有利于蛋白质交联逆转,从而从 FFPE 组织中成功提取蛋白质。然而,甘油在蛋白质提取中没有作用,因为添加甘油(或甘氨酸)到提取缓冲液中不能提高蛋白质的回收率。但是甘油常被添加到蛋白质提取缓冲液,因为它具有增加渗透压和稳定蛋白质。此外,甘油可以作为冷冻保护剂有益于 FFPE 蛋白质提取液在 -80℃ 的长期保存<sup>[69,70]</sup>。综上所述,蛋白提取缓冲液 1 的蛋白提取效率可归因于 SDS、甘油与 pH 8.8 的碱性介质的最佳组合。然而,有研究认为,缓冲成分对蛋白质提取效率的影响相对较小。他们认为,在高于 60℃ 的温度下加热样品对高蛋白回收至关重要<sup>[71,72]</sup>。尽管 Sprung(2009)<sup>[71]</sup>和 Yamashita(2007)<sup>[72]</sup>并没有发现改变缓冲液的主要差异,但当使用相同的蛋白质提取方案时,我们确实看到了改变缓冲液成分的效果。

尽管目前对 FFPE 组织蛋白组学的研究已经取得了一定的进展,但基于质谱分析的 FFPE 组织蛋白质组分析仍具有较高的技术变异水平。技术变异可细分为两部分:样品制备的重复性和质谱分析的重复性。相比之下,由于甲醛诱导的蛋白质交联的还原是一个不可控的过程,样品制备技术变化影响较大。此外,FFPE

组织蛋白质回收效率也受组织固定方案、固定时间和保存时间等多重因素的影响。考虑到人类 FFPE 组织样本从来没有以标准化的方式制备，人类 FFPE 组织样本的技术总变异水平将更高。由于性别、年龄、健康状况、饮食等因素的影响，人体组织 FFPE 样本中的生物变异水平很高。因此，在进行统计上可靠的实验时，需要考虑这些因素，一个实验的统计可靠性是由统计能力决定的，因此，为了防止得出假阴性结论，需要高统计能力。

在现有的研究中，新鲜冷冻组织和 FFPE 组织之间有许多关键的比较，这些研究的普遍共识是，在过去几年中取得了非常令人鼓舞的成果。FFPE 组织样本的蛋白质组已被证明与成对新鲜冷冻组织样本的蛋白质组具有高度可比性，尽管这主要适用于鸟枪法蛋白组学实验<sup>[73]</sup>。另外，FFPE 组织固定和后续组织处理的效果(蛋白质交联和修饰)可以对提取尽可能多的全长 FFPE 蛋白以及以明确的方式鉴定提取的蛋白产生影响。

## 2.6 结论

我们通过优化和改良了 FFPE 组织蛋白提取缓冲液的组成成份和蛋白提取流程，成功从 FFPE 组织中获得较高浓度的蛋白溶液，并且发现含有相对更高的去污剂 SDS 的缓冲液具有更高的蛋白提取效率。

## 第 3 章 结肠癌 FFPE 组织溶液内消化, 凝胶内消化以及悬浮 S-Trap 肽段获取效率的比较

### 3.1 前言

全蛋白质和多肽的鉴定和定量已成为分析生物系统的重要工具。利用超高效液相色谱(UPLC)或毛细管区带电泳(CZE)与质谱仪联用的质谱研究通常用于鉴定感兴趣的样品中的数千种蛋白质。大规模的蛋白质组学研究往往依赖于利用胰蛋白酶将蛋白质消化成肽段, 并利用相应的肽离子强度进行定量。有效地将蛋白质消化成相应的肽段是实现蛋白质自下而上定量的关键。蛋白质组的覆盖深度在很大程度上取决于选择的样品制备和分离技术。质谱法上游的样品制备方法因样品类型、裂解条件、消化和离线分馏方法的不同而有很大的差异, 每种方法都有各自的优缺点。之前的一些研究比较了裂解条件<sup>[74,75]</sup>、分馏方法<sup>[76,77]</sup>和消化条件<sup>[78,79]</sup>以得出最佳的蛋白质组制备方案。

大多数研究强调去污剂能够检测在细胞中发挥许多重要功能的疏水性膜蛋白的重要性。为了达到这个目的, 最常用的洗涤剂是 SDS, 它很容易溶解生物基质中的蛋白质。对于 FFPE 组织目前切实可靠的蛋白提取缓冲液中都含有相对较高浓度的去污剂 SDS。然而, SDS 的去除在质谱分析之前是至关重要的, 因为任何残留的 SDS 都会污染或者损坏质谱分析仪。近年来最流行的去污剂去除和消化方法是滤器辅助样品制备法(FASP)。FASP 将含 SDS 的蛋白样品置于过滤器中, 用 8 M 尿素洗涤, 以破坏 SDS 胶束, 然后用各种缓冲液清洗过滤器, 去除多余的尿素, 再加入胰酶进行消化, 最后洗脱多肽。尽管这种方法已经在许多应用中获得了广泛的成功<sup>[80-82]</sup>, 但是实验方案的繁杂性和结果的不稳定性常常阻碍其在高通量蛋白质组学研究中的应用。基于这些原因, 最近又开发了新的技术来辅助基于 SDS 的蛋白质组样本质谱分析的准备。

Zougman 等人描述了一种悬浮捕获(S-Trap)蛋白消化方法, 该方法可以在 FASP 方案所需的一小部分时间内制备包含 SDS 的蛋白裂解物<sup>[83]</sup>。在此方法中, 蛋白质在 5%SDS 中裂解, 通过添加磷酸和甲醇缓冲液, 形成精细的蛋白质颗粒悬浮液。悬浮液被 S-Trap 过滤柱捕获结合在过滤柱上, 剩余的 SDS 通过洗涤的方法被冲走。在 LC-MS/MS 分析之前, 蛋白质在过滤器中使用所选择的蛋白酶

消化。S-Trap 具有 FASP 和其他基于过滤器的方法的优点，同时减少了样品处理步骤和质谱分析前的时间。

在本研究中，我们检测了结肠癌患者 FFPE 组织在不同条件下裂解和消化的蛋白质的数量。我们使用 S-Trap 方法，该方法允许含有高浓度的离子型和非离子型的去污剂 SDS，并与传统的溶液内消化方法、凝胶内消化方法肽段获取效率进行比较。裂解之后所得肽段的使用 Pierce™ 定量比色肽分析方法以及 LC-MS/MS 分析方法检测。

## 3.2 材料

### 3.2.1 购置试剂

Tris-HCl 缓冲液	美国 Sigma 公司
十二烷基硫酸钠	美国 Sigma 公司
碘乙酰胺	美国 Sigma 公司
1, 4-二硫代苏糖醇	美国 Sigma 公司
丙三醇	美国 Fisher Chemical 公司
二甲苯	美国 Sigma 公司
1%蛋白酶抑制剂	美国 Thermo Scientific™公司
乙基苯基聚乙二醇	美国 Bufferad 公司
3-[3-(胆酰胺丙基)二甲氨基]丙磺酸内盐	加拿大 MEMD Millipore 公司
脱氧胆酸钠	美国 Sigma 公司
4xLaemmli 样本缓冲液	美国 Bio-Rad 公司
考马斯亮蓝染色剂	美国 Bio-Rad 公司
10xTris-Glycine-SDS 缓冲液	美国 Bio-Rad 公司
质谱级 W6-1 去离子水	美国 Fisher Chemical 公司
酒精	美国 Fisher Chemical 公司
BCA 试剂盒	美国 Thermo Scientific™公司
Mini-PROTEAN®TGX™胶	美国 Bio-Rad 公司
碳酸氢铵	美国 Sigma 公司
甲醇	美国 Fisher Chemical 公司
S-Trap 捕获柱	美国 Protifi 公司

### 3.2.2 主要仪器设备及耗材

Thermomixer 恒温混匀仪	德国 Eppendorf 公司
Sorvall™ Legend™ Micro 21 微量离心机	美国 Thermo Scientific™公司
真空浓缩器	美国 Thermo Scientific™公司
5430R Eppendorf 离心机	德国 Eppendorf 公司
超声波破碎仪	美国 Thermo Scientific™公司
-80℃冰箱	美国 Thermo Scientific™公司
-20℃冰箱	美国 Thermo Scientific™公司
4℃冰箱	美国 Thermo Scientific™公司
移液器	德国 Eppendorf 公司
电子天平	美国 Mettler Toledo 公司
PCR 微型管	德国 Eppendorf 公司
定转速涡流混合器	美国 Thermo Scientific™公司
电泳仪	美国 Bio-Rad 公司
摇床	美国 Thermo Scientific™公司
Q-Exactive 质谱分析仪	美国 Thermo Scientific™公司

## 3.3 方法

### 3.3.1 溶液内消化样本准备方法

- (1) 100 $\mu$ g 蛋白溶液加入 9 $\mu$ l 100mM 碳酸氢铵溶液中。
- (2) 加入 1.1 $\mu$ l 50mM DTT (DTT 终浓度为 5mM) 样本溶液中。
- (3) 在 450 转/分 Thermomixer 恒温混匀 56℃ 孵育器中孵育 30 分钟。
- (4) 冷却样本至室温。
- (5) 加入 1.2 $\mu$ l 110mM/L IAA 使 IAA 浓度为 11.0mM。
- (6) 室温下暗处孵育 30 分钟。
- (7) 30 分钟后加入 1.3 $\mu$ l 50mM DTT 以终止 IAA 反应。
- (8) 室温暗处孵育 15 分钟。
- (9) 加入 50mM 碳酸氢铵溶液 67.5 $\mu$ l 以 1:6 体积比稀释样本，总的样本量为 81.0 $\mu$ l。

(10) 以样本与胰酶的质量比 50:1 加入胰酶溶液 2 $\mu$ g，样本溶液总体积为 91.0 $\mu$ l。

(11) 37 $^{\circ}$ C 450 转/分 Thermomixer 恒温混匀器孵育过夜。

### 3.3.2 Pierce®去污剂移除方法

(1) 将 Pierce®去污剂自旋柱底部覆盖物去除并松开自旋柱盖子，不要去盖子。

(2) 将自旋柱放进一个 2ml 的收集管中。

(3) 1500g 离心 1 分钟去除储存液。在柱子的一边做上标记以确保每次离心标记物都朝向外侧。

(4) 加入 400 $\mu$ l 50mM 碳酸氢铵溶液后以 1500g 转速离心 1 分钟，弃去缓冲液。

(5) 重复步骤 4 两次。

(6) 将自旋柱移至一个新的 1.5mL 的收集管中。

(7) 缓慢的将样本溶液加入自旋柱顶部的树脂床上，室温孵育 2 分钟。

(8) 1500g 转速离心 2 分钟收集以去除去污剂的样本溶液。

### 3.3.3 C18 脱盐方法

(1) 在去除去污剂后的样本溶液中加入 20%的 TFA，使样本中的 TFA 浓度为 1%。PH 试纸检测 PH，最佳 PH 值在 1~2 之间。

(2) 室温下放置酸化样本溶液 10 分钟。

(3) 10 $^{\circ}$ C 10000 转/分离心 10 分钟以清除样本中的杂质。

(4) 将含有肽段的上清液移至一个新的 1.5ml 离心管中。

(5) 用真空蒸发器不加热的条件下浓缩液体至 50 $\mu$ l。

(6) 再次检测 PH 值以确保溶液 PH 值在 2.0 左右。

(7) 用 50 $\mu$ l 100%乙腈预湿带有两个 C18 圆形滤片的 P200 吸头，轻弹以去除气泡，376g 离心，1 分钟或 1 分钟以上直至液体全部通过滤纸。

(8) 加入 50  $\mu$ l 0.5%乙酸 50%乙腈溶液洗涤，376g 离心 1 分钟或 1 分钟以上。

(9) 加入 50 $\mu$ l 0.1%TFA 洗涤，376g 离心 1 分钟或 1 分钟以上，弃去洗涤液。

(10) 将含有肽段的样本溶液加入经多次洗涤的 C18 柱上，240g 离心，2 分钟或者 2 分钟以上。

(11) 加入 50 $\mu$ l 0.1% TFA 洗涤 C18 柱，376g 离心 1 分钟或 1 分钟以上。

(12) 用 50 $\mu$ l 0.5%乙酸 50%乙腈溶液洗脱肽段，240g 离心 2 分钟或者 2 分钟以上。再次加入 30 $\mu$ l 0.5%乙酸 50%乙腈溶液，240g 离心 7 分钟后，6000 转/分离心 2 分钟以便从 C8 柱中洗脱获得更多的肽段。

(13) 将所收集的溶液置于 -80 $^{\circ}$ C 冷冻 20 分钟后用真空蒸发器烘干，加入 20 $\mu$ l 3%乙腈 0.1%甲酸溶液以备质谱分析。

### 3.3.4 凝胶内消化样本准备方法

(1) 将 100 $\mu$ g 蛋白质样本经 SDS-PAGE 分离（分离时间为 7 分钟，其它条件及方法同第二章所述）。

(2) 使用洁净刀片、刮铲、玻璃平皿，截取待鉴定的蛋白质条带，从凝胶中非蛋白区域截取同样大小的一段凝胶片，与蛋白质条带同样处理，用作胶内蛋白消化及后续质谱分析的对照。尽量在与蛋白质接近的位置取，以便减少背景量。

(3) 将截取的凝胶切成 1mm 小块，转移至 1.5ml 离心管中。

(4) 用水/乙腈 (1:1) 混合液洗涤凝胶 15min，使用的体积应至少为凝胶体积的 2 倍，迅速离心 (1s) 以使液体从管壁中集中到管底，使用装有洁净枪头的移液器吸去液体。

(5) 添加乙腈至完全覆盖凝胶。

(6) 待凝胶收缩至完全显乳白色后，吸去乙腈。

(7) 加入 100mmol/L 的碳酸氢铵。

(8) 5 分钟后，加等体积的乙腈（达到 1:1 的比例）。

(9) 孵育 15 分钟，倾去液体。

(10) 用真空蒸发器抽干凝胶。（避免采用加热来提高干燥速度）

(11) 加入适量 10mmol/L 的 DTT 溶液和 100mmol/L 的碳酸氢铵溶液，以完全浸没凝胶为宜。如溶液凝胶吸收，可以补加更多的溶液，65 $^{\circ}$ C 孵育 45 分钟。

(12) 吸去该溶液，并迅速加入等体积的 50mmol/L 的 IAA 和 100mmol/L 的碳酸氢铵缓冲液，于室温下暗处孵育 30 分钟。

- (13) 吸去 IAA 溶液，按照步骤 4~步骤 6 所述的方法洗涤凝胶。
- (14) 用真空蒸发器抽干凝胶。
- (15) 加入消化液浸没凝胶，将凝胶置于冰上孵育 45 分钟，如果加入的消化液被吸收则进行适当添加，以浸没凝胶为宜。
- (16) 37℃ 孵育过夜。
- (17) 次日添加适量 100mmol/L 碳酸氢铵溶液以浸没凝胶，室温孵育 15 分钟。
- (18) 添加等量乙腈，室温孵育 15 分钟。
- (19) 用装有洁净吸头的移液器将上清液转移至一个新的 1.5ml 离心管内。
- (20) 重复此提取过程（步骤 17~步骤 19）两次，但用 5%的甲酸溶液取代 100 mmol/L 的碳酸氢铵溶液，以保护肽段。将所得的上清液都混合至 1.5mL 离心管中。
- (21) 将收集的上清液在干冰上冷冻。
- (22) 将提取的肽段冷冻干燥至近干，但不可过于干燥，最理想状况是离心管内残留 1~2 $\mu$ l 液体即可停止。
- (23) 加入 20 $\mu$ l 3%乙腈 0.1%甲酸溶液复溶肽段，储存于-80℃ 冰箱中备用。

### 3.3.5 S-Trap 消化样本准备方法

- (1) 100 $\mu$ g 结肠癌 FFPE 蛋白溶液加入 100mM 碳酸氢铵缓冲溶液至 25 $\mu$ l。
- (2) 加入 25 $\mu$ l 2X 裂解缓冲液至步骤 1 含 100mM 碳酸氢铵缓冲溶液样本中，溶液总量为 50 $\mu$ l。
- (3) 在 13000g 的条件下离心 8 分钟以澄清样本，离心后更换新的 1.5mL 离心管。
- (4) 将 1.02 $\mu$ l 1 M DTT 加入含有 SDS 的样品溶液中，最终浓度为 20mM，以还原二硫化物。
- (5) 在 Thermomixer 恒温混匀仪上孵育 10 分钟，孵育条件为 95℃，450 转/分。孵育后常温下冷却 10 分钟。
- (6) 加入 2.68 $\mu$ l 800mM IAA 至最终浓度 40mM 以烷基化样品溶液。
- (7) 室温黑暗条件下，450 转/分继续孵育半小时。
- (8) 13000g 的条件下离心 8 分钟以继续清除未溶解的杂质，离心后更换新的 1.5mL 离心管。此时溶液总量为 53.9 $\mu$ l。

(9) 在 53.9 $\mu$ l 样本溶液中加入 5.39 $\mu$ l 12%磷酸溶液以酸化样本，总的溶液量为 59.3 $\mu$ l。(12%磷酸与样本溶液的体积比为 1:10)

(10) 在样本溶液中加入 415 $\mu$ l S-Trap 缓冲液，S-Trap 缓冲液与样本溶液的体积比为 1: 7。

(11) 将含裂解液的样本加入含有迷你 S-Trap 旋转柱的 2ml 收集试管中，将样本加入 S-Trap 旋转柱。

(12) 在台式离心机中以 4000g 转速旋转 60 秒，或直到所有溶液都通过旋转柱。蛋白质会与 S-Trap 旋转柱蛋白质捕获剂结合。弃去滤过溶液。

(13) 在旋转柱中加入 400 $\mu$ l S-Trap 缓冲液洗涤捕获蛋白 3 次，以 4000g 离心 60 min，每台离心柱旋转 180 度。

(14) 将结合有蛋白质的 S-Trap 旋转柱移至一个新的干净的 2ml 收集试管中。

(15) 在旋转柱上加 125 $\mu$ l 含有胰蛋白酶/赖氨酸酶混合物的消化缓冲液，酶与蛋白质的质量为 1:20。

(16) 如果有含胰酶的溶液渗透入旋转柱到收集管中，可以用移液器将其重新收集至捕获有目的蛋白的旋转柱上，确保蛋白捕获柱上无气泡产生，如果必要，轻弹收集管。

(17) 37 $^{\circ}$ C 条件下在 Thermomixer 恒温混匀仪孵育过夜，避免搅拌或轻摇。

(18) 次日，将 S-Trap 旋转柱取出放置在一个新的 Lobind 1.5ml 的离心管中。

(19) 将 80 $\mu$ l 消化缓冲液加入含有蛋白酶的 S-trap 自旋柱中。1000g 离心 60 秒，将肽段洗脱。

(20) 在 S-Trap 蛋白捕获基质中加入 80 $\mu$ l 0.2%甲酸，以 1000g 的速度旋转 60 秒。

(21) 用 80 $\mu$ l 含有 50%乙腈，0.2%甲酸混合溶液在 4000g 的转速下洗脱肽段。

(22) 将洗脱液置入-80 $^{\circ}$ C 冰箱中冰冻 20 分钟使溶液成固态后取出，置入 SpeedVac 真空浓缩器中将含肽段溶液烘干，加入 3% ACN, 0.1 FA 20 $\mu$ l，超声水浴分离 15min 后置入-80 $^{\circ}$ C 冰箱中备用。

### 3.3.6 Pierce™定量比色肽分析方法

### 3.3.6.1 标准液的制备

使用表 1 的程序，制备肽消化试验标准的稀释系列，生成标准曲线。将肽消化试验标准稀释成干净的小瓶，最好使用与样品相同的稀释剂。这将提供足够的容量来运行了 8 分标准曲线(从 0-1000 $\mu\text{g}/\text{ml}$ )一式三份。

表 3.1 制备稀释肽消化测定标准品

离心管	去离子水	消化液体积 ( $\mu\text{l}$ )	消化液终浓度 ( $\mu\text{g}/\text{ml}$ )
A	0	150ml 储存液	1000
B	75	75ml A 溶液	500
C	75	75ml B 溶液	250
D	75	75ml C 溶液	125
E	75	75ml D 溶液	62.5
F	75	75ml E 溶液	31.3
G	75	75ml F 溶液	15.6
Blank		去离子水	0

### 3.3.6.2 工作液的制备

(1) 使用以下公式确定所需工作试剂的总体积(WR):

$$[(\text{标准})+(\text{未知})]\times(\text{重复})\times(\text{每样 } 0.18\text{mL}) = \text{所需总体积 WR}$$

(2) WR 混合液的制备:

50 份比色肽检测试剂 A+48 份比色肽检测试剂 B+2 部分比色肽检测试剂 C

注:WR 制备完成后，室温不能超过 30 分钟。

### 3.3.6.3 肽段浓度监测

(1) 准备样品和标准。

(2) 吸管 20 $\mu\text{l}$  每个标准或未知样本至 1.5 $\mu\text{l}$  离心管内。

(3) 在每个 PCR 试管内添加 180 $\mu\text{l}$  WR 充分混匀，盖上盖子后 37 $^{\circ}\text{C}$  孵育半小时。

(4) NanoDrop One 分光光度计检测肽段含量，检测波长 205nm.

### 3.4 结果

临床 FFPE 结肠腺癌组织，蛋白按照上述（第2章）蛋白提取以及定量方法获取蛋白之后，所有样品均用 S-Trap 胰蛋白酶和赖氨酸-C 酶双重消化。

#### 3.4.1 Pierce™定量比色肽分析方法鉴定各个样品肽段的含量

表 3.1 显示来自 8 个不同的病人 FFPE 结肠癌组织蛋白的肽段恢复量。使用溶液内消化、凝胶内消化和 S-Trap 消化所得的平均肽段恢复量分别是 9.1%，9.5% 和 93.5%。从该研究结果明显可以看出 S-Trap 消化方法的肽段提取效率明显高于溶液内消化方法和凝胶内消化方法。

表 3.1 Pierce™定量比色肽分析方法鉴定各个样品肽段的含量 (n=8)

	蛋白质 ( $\mu\text{g}$ )	酶/蛋白质 质量比	平均肽段 ( $\mu\text{g}$ )	平均肽段恢复率 (%)
溶液内消化	100	1:50	9.1	9.1
凝胶内消化	100	1:50	9.5	9.5
S-Trap 消化	100	1:20	93.5	93.5

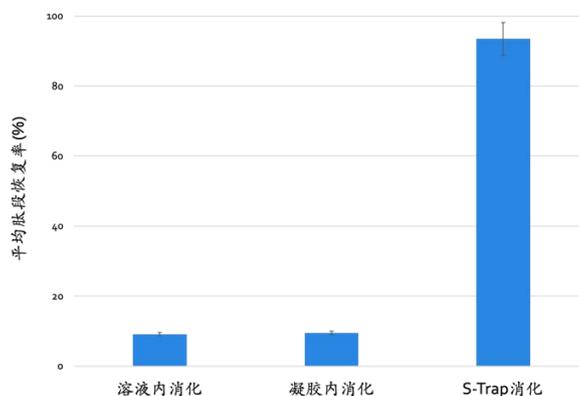


图 3.1 不同消化方法肽段的含量比较

#### 3.4.2 LC-MS/MS 分析方法鉴定各个样品蛋白含量

为了进一步验证 Pierce™定量比色肽分析方法检测结果的准确性以及 S-Trap 消化方法的肽段提取高效性、精确性以及再现性，以及与质谱分析的兼容性，为临床 FFPE 组织的蛋白组学研究提供更强劲的质谱分析样本准备方法，我们使

第 3 章 结肠癌 FFPE 组织溶液内消化，凝胶内消化以及悬浮 S-Trap 肽段获取效率的比较

用一种高性能台式四极杆-轨道阱 LC-MS/MS 系统 Q-Exactive 质谱分析仪。

将样本充分溶解，然后上样到质谱仪，质谱仪使用美国 Thermo Scientific 公司的 Q-Exactive 质谱仪。高效液相系统为黛安 NCS3500 系统。液相色谱 A 相为 99.9%水、0.1%甲酸。B 相为 99.9%乙腈、0.1%甲酸。分析时长：90 分钟，全扫描范围  $m/z$ :350-1800。一级质谱扫描范围  $m/z$  200，分辨率 7 万。自动增益控制 (AGC) 为：2 万。一级最大 IT 50ms，采用“最高速度”算法进行数据依赖扫描。全扫描中选择离子强度 TOP 20 的母离子进行二级质谱鉴定。二级质谱扫描范围  $m/z$  200，分辨率：1.75 万，AGC: 2 万，二级最大 IT 100ms，未填充率:1.0%，离子强度阈值:  $2.0e4$ ，动态排除时间: 18.0 s。母离子使用 HCD 方法碎裂，进行二级质谱序列测定，用报告离子的比例定量，生成质谱检测数据。

质谱分析后得到的蛋白质数量，肽段的数量使用 Scaffold 软件进行搜库分析（蛋白质阈值 1.0%FDR，最小肽段 1，肽段阈值 1.0%FDR），其结果结果见表 4 和图 4。从图中可以看出，S-Trap 消化方法所得蛋白、肽段和谱图数量明显高于溶液内和凝胶内消化方法。

表 3.2 三种不同方法消化效率的比较 (n=8)

消化方法	蛋白质	肽段
溶液内消化	443	1126
凝胶内消化	465	1211
S-Trap	1855	9433

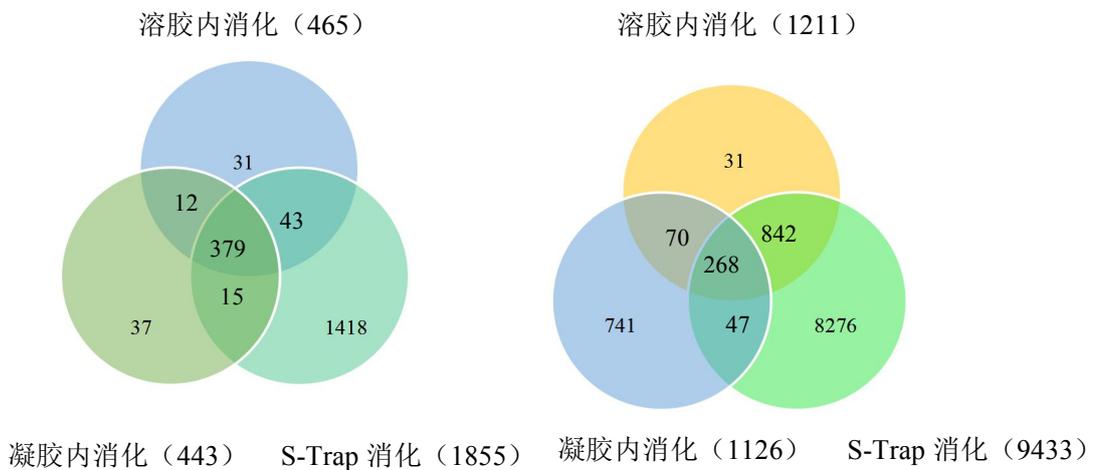


图 3.2 不同消化方法所得蛋白质和肽段比较

### 3.5 讨论

自下而上的蛋白质组研究方案依赖于从待测样品中提取蛋白质，蛋白质提取之后，完整的蛋白质在注入质谱分析仪前需要在蛋白酶消化成肽段。含有 SDS 的裂解缓冲液通常是蛋白质组学分析的首选，因为它们能够提取更多的蛋白质<sup>[74,75]</sup>。溶液内消化和凝胶内消化是新鲜冰冻组织标本最传统以及使用最广泛的两种方法；然而，去污剂的去除（例如 SDS）往往比较困难并且去除不完全。任何残留的离子 SDS 将污染和损害质谱分析仪。

溶液内消化时新鲜冰冻组织和 OCT 组织蛋白组学质谱分析样本准备的常规方法，被广泛使用，但是对于含有高浓度去污剂 SDS 的 FFPE 组织来说，溶液内消化并不适合，因为离子型去污剂与质谱分析仪不兼容（污染或损坏质谱分析仪）。为了解决这一问题，我们尝试了在常规溶液内消化的基础上通过去污剂清除过滤柱以清除残存的 SDS，但是实验结果不尽人意，通过该方法我们仅能得到 50% 肽段恢复率，这与新鲜冰冻组织或 OCT 组织肽段恢复率相比，肽段产量过低，不利于基于质谱分析的蛋白组学研究方案。我们认为 SDS 清除过程丢失了大部分肽段。

凝胶内消化是另一种新鲜冰冻组织和 OCT 组织蛋白组学质谱分析样本准备的常规方法。理论上讲，对于 FFPE 组织，使用完整的 FFPE 组织切片提取蛋白进行凝胶电泳是有利的，因为在将样品加载到液相色谱柱上之前，凝胶可以去除残留石蜡和质谱分析不相容的缓冲成分如离子型去污剂 SDS 等。我们的研究结果发现，来自 FFPE 组织的蛋白提取液经过常规的一维 SDS-PAGE，凝胶内消化之后检测到的肽段同样很低，其原因我们考虑经甲醛处理后的 FFPE 组织因化学交联的形成后，部分蛋白并不能顺利通过凝胶而停留在凝胶上样井口或溢出至凝胶外，导致凝胶内蛋白含量减低；另外，残存石蜡可能也是影响结果的一个原因，但是对于化学交联和残存石蜡的检测目前尚无比较成熟可靠的方法。对于化学交联，蛋白提取过程中的高温孵育以及强有力的超声碎裂是大部分研究者比较认同的减少化学交联提高肽段恢复率的方法。反复多次二甲苯和不同浓度酒精清洗是尽量减少样本石蜡残余的可靠方法。

有研究报道，一个新的基于滤器的 S-Trap 消化方法允许样本中含有高浓度的离子或非离子成份的 SDS，在蛋白消化之后注入质谱分析仪之前可以通过数次洗涤去除，并且该洗涤方法并不影响肽段的产生量。我们比较了溶液内消化方法、凝胶内消化方法和 S-Trap 消化方法的蛋白消化效率和肽段恢复产生量。

在给予质谱分析鉴定之前，我们使用 Pierce™定量比色肽分析方法鉴定各个样品肽段的产生量。从表 3 我们可以清楚的看到，使用溶液内消化、凝胶内消化和 S-Trap 消化所得的平均肽段恢复量分别是 9.1%，9.5%和 93.5%。S-Trap 消化方法的肽段提取效率明显优于溶液内消化方法和凝胶内消化方法，且与质谱分析方法兼容。该研究结论与 Katelyn<sup>[84]</sup>等的研究结果基本相同。

显然，仅仅根据定量比色肽分析方法对肽段进行定量是不完全可靠的，为了证实我们的研究结果，我们使用同样的样本进行质谱分析评估不同消化方法蛋白质和肽段恢复率。基于成本、质谱分析等待时间等综合考虑，我们选用 Q-Exactive 质谱分析仪对样本进行检测。

Q-Exactive 是 2011 年问世的一种高性能台式四极杆-轨道阱 LC-MS/MS 系统，具有新型 C-Trap 离子光学系统和碰撞诱导解离(HCD)碰撞池，提供了快速 HCD MS/MS 扫描并改善了低质量数离子的传递，从而提高灵敏度和定量性能，尤其适用于使用同位素标签的实验;多重检测提高了整体系统工作周期的效率，能够更好的与 UHPLC 兼容，并在 Orbitrap 进行同时检测之前收集并保存多种母离子功能。检测结果进一步证实 S-Trap 消化方法的肽段提取效率以及蛋白恢复率明显优于溶液内消化方法和凝胶内消化方法（约 4 倍）。

### 3.6 结论

与传统的溶液内和凝胶内消化方法相比，新的 S-Trap 消化方法在 FFPE 样品中肽段回收率明显高于其它两种方法。S-Trap 操作方法简单，允许溶液内含有较高浓度的去污剂 SDS，肽段回收率高，为深入蛋白组学研究提供较好的质谱样本。

## 第 4 章 基于液相色谱质谱 (LC-MS/MS) 联用的结肠癌 FFPE 组织差异表达蛋白的检测及肝转移过程的认识

### 4.1 前言

通过蛋白质组学, 可以分析成千上万的癌细胞蛋白, 从而为 CRC 产生新的治疗靶点和标志物。发现癌症生物标志物最常用和最方便的蛋白质组学策略包括识别“差异”表达的蛋白质和多肽。这里的“差异”一词是指: 1) 肿瘤组织与相应的正常组织相比, 或在不同的肿瘤发生阶段之间, 其水平是不同的; 2) 周围生物环境和液体(如血浆)对肿瘤发生、存在和生长的反应水平不同; 3) 肿瘤或周围液体在特定治疗条件下的表达变化。前两种情况下明显指的是具有潜在诊断或预后效用的因素, 而最后一种情况则与预测标志物的发现有关, 这些标志物可能有助于临床医生在不同的治疗方案之间做出选择。当然, 候选标记物的潜在临床应用在很大程度上取决于发现策略的实验设计, 特别是通过: 1) 生物模型(如细胞培养、动物、患者); 2) 样本类型(例如细胞提取物、组织提取物、生物液体); 3) 蛋白组学技术及样本量(即样本数); 4) 验证的类型和样本量。

肝脏是结直肠癌最常见的转移部位, 结直肠癌患者中, 超过 50% 的病例最终进展为肝脏转移<sup>[85]</sup>。肝转移是结直肠癌患者的主要死亡原因<sup>[86]</sup>, 外科手术是使结直肠癌肝转移病人长期生存的重要治疗策略, 接受手术后, 结直肠癌肝转移患者 5 年生存率可达 16%-71%<sup>[87]</sup>, 尽管不同研究中对其生存率的分析存在差异, 但都肯定了手术在结直肠癌肝转移治疗中的重要地位。然而, 多数结直肠癌肝转移无法获得根治性切除, 通过新辅助化疗等转化治疗, 明显增加了手术切除率<sup>[88]</sup>, 但即使接受手术治疗, 1 年内复发率达 30%<sup>[87]</sup>。此外, 结直肠癌合并肝脏及肝外转移也屡见报道, 如果不进行治疗, 转移性结直肠癌患者的平均预期寿命约为 8 个月<sup>[89]</sup>。复发和转移往往导致治疗失败。近些年来, 随着转化治疗模式的推广应用, 多学科团队诊断与治疗模式的确立, 诊断技术的进步, 新型化疗药物及靶向药物的应用使得结直肠癌肝转移的预后有了明显改善, 但结直肠癌肝转移的复发及包括肝脏转移在内的远处转移仍面临严峻挑战。故进一步了解结直肠癌肝转移的生物学过程, 对发现肝转移的特定标志物以及对结肠癌晚期肝转移的病人提供新的治疗靶点有着重要的临床意义。

来自患者的生物样本是最有价值的蛋白质组学标记发现的生物系统。本研究利用临床病人的匹配有正常组织、原发性肿瘤组织以及肝转移组织的结肠癌 FFPE 组织作为研究对象, 通过液相色谱质谱 (LC-MS/MS) 联用的方法以期发现新的肿瘤生物标志物, 并了解肿瘤的转移过程。

## 4.2 材料

### 4.2.1 购置试剂

Tris-HCl 缓冲液	美国 Sigma 公司
十二烷基硫酸钠	美国 Sigma 公司
碘乙酰胺	美国 Sigma 公司
1, 4-二硫代苏糖醇	美国 Sigma 公司
丙三醇	美国 Fisher Chemical 公司
二甲苯	美国 Sigma 公司
1%蛋白酶抑制剂	美国 Thermo Scientific™公司
乙基苯基聚乙二醇	美国 Bufferad 公司
3-[3-(胆酰胺丙基)二甲氨基]丙磺酸内盐	加拿大 MEMD Millipore 公司
脱氧胆酸钠	美国 Sigma 公司
4xLaemmli 样本缓冲液	美国 Bio-Rad 公司
考马斯亮蓝染色剂	美国 Bio-Rad 公司
10xTris-Glycine-SDS 缓冲液	美国 Bio-Rad 公司
质谱级 W6-1 去离子水	美国 Fisher Chemical 公司
酒精	美国 Fisher Chemical 公司
BCA 试剂盒	美国 Thermo Scientific™公司
Mini-PROTEAN®TGX™胶	美国 Bio-Rad 公司
碳酸氢铵	美国 Sigma 公司
甲醇	美国 Fisher Chemical 公司
S-Trap 捕获柱	美国 Protifi 公司
胰酶/赖氨酸酶混合酶	美国 Promega 公司

## 4.2.2 主要仪器设备及耗材

Thermomixer 恒温混匀仪	德国 Eppendorf 公司
Sorvall™ Legend™ Micro 21 微量离心机	美国 Thermo Scientific™公司
真空浓缩器	美国 Thermo Scientific™公司
5430R Eppendorf 离心机	德国 Eppendorf 公司
超声波破碎仪	美国 Thermo Scientific™公司
-80℃冰箱	美国 Thermo Scientific™公司
-20℃冰箱	美国 Thermo Scientific™公司
4℃冰箱	美国 Thermo Scientific™公司
移液器	德国 Eppendorf 公司
电子天平	美国 Mettler Toledo 公司
PCR 微型管	德国 Eppendorf 公司
定转速涡流混合器	美国 Thermo Scientific™公司
电泳仪	美国 Bio-Rad 公司
摇床	美国 Thermo Scientific™公司
Orbitrap Fusion Lumos 质谱仪	美国 Thermo Scientific™公司

## 4.3 方法

### 4.3.1 结肠癌 FFPE 组织 LC-MS/MS 样本的准备

我们选择了 18 个结肠腺癌病人共 58 个匹配有正常组织、原发性肿瘤组织和肝转移组织的样本。所有样本均取自美国纪念斯隆凯特琳癌症中心 (Memorial Sloan Kettering Cancer Center, MSKCC) 生物样本库。蛋白提取方法见第二章, 提取缓冲液为含有 2%SDS 的缓冲液 1。每个标本中含有从 FFPE 组织中获得的 10 个薄层切片, 每片厚度为 10 $\mu$ m。蛋白提取之后每个样本吸取 3 $\mu$ l 含蛋白提取液混匀作为质量控制标本 (QC); 蛋白质消化方法选用 S-Trap 法, 消化酶选用胰酶/赖氨酸酶混合酶, 详细操作方法见第三章 S-Trap 消化样本准备方法部分。

### 4.3.2 FFPE 结肠癌组织 LC-MS/MS 分析

将样本充分溶解, 然后上样到质谱仪, 该实验部分质使用美国 Thermo Scientific 公司的 Orbitrap Fusion Lumos 高分辨质谱仪。液相色谱为 Dionex

UltiMate 3000 系统; 色谱分析柱为 Acclaim Pepmap RSLC(C18,75 $\mu$ m $\times$ 15cm,2 $\mu$ m) 预柱为 Acclaim Pep-map100( C18,75 $\mu$ m $\times$ 2cm,3 $\mu$ m) ;流动相 A 为 0.1%甲酸溶液, 流动相 B 为 80%乙腈溶液(含 0.1%甲酸)。梯度洗脱程序:0~5min, 5%B; 5~104min,5%~45% B; 104 ~105 min, 45%~99% B; 105~110 min , 99%B;110~110.1min, 99%~5%B; 110.1~120 min,5% B。流速为 300 nl/min; 进样量 2 $\mu$ l。正离子扫描方式, 一级质谱扫描范围 m/z 300~2000, 分辨率 12 万。自动增益控制(AGC) 为 10 万, 采用高能诱导裂解(HCD) 方式裂解, 归一化碰撞能量 30%, 采用 FT 采集二级质谱数据,分辨率 6 万, 采用“最高速度”算法进行数据依赖扫描, 离子强度高于 2.5e4 一级离子进行二级碎裂, 动态排除时间为 60 秒。

### 4.3.3 蛋白质鉴定和定量

质谱分析所得原始数据使用 MaxQuant 软件(版本 1.6.0.16)处理, 搜索引擎 Andromeda, 使用默认参数进行蛋白质鉴定和定量, 数据来自 Uniprot 数据库。可变修饰:蛋氨酸氧化和 N 末端乙酰化; 固定的修饰:半胱氨酸脲基甲基化; 无标记定量方法 (LFQ); 误检率 (FDR) :0.01; 并设置运行之间的匹配进行搜索和量化。UniProt 数据库下载人类参考蛋白组的 FASTA 文件。

### 4.3.4 统计学分析

差异表达蛋白(DEPs)使用 Perseus 软件进行鉴定(版本 1.5.4.1)。将 MaxQuant 搜索产生的 LFQ 强度导入 Perseus: 删除仅位点鉴定, 反向和污染物; 将表达式值转换为以 2 为底的对数; 将样本分成三组: 正常组织组、原发肿瘤组和肝转移瘤组; 样本根据各自的组进行注释; 对有效值进行筛选, 至少有 1 组中出现 70%以上的有效值; R2 相关系数表示样本相关性。采用 2 样本 t 检验比较两组间蛋白表达情况, p 值设置为<0.05。2 样本 t 检验后生成火山图; 计算 Z 分数后, 对转移样本 DEPs 进行无监管分层聚类。

### 4.3.5 差异表达蛋白的生物信息学分析方法

#### 4.3.5.1 差异表达蛋白的筛选

为了筛选出显著的差异表达蛋白, 我们又对样本进行了生物信息学分析。将样本信息分成三组: 正常组、原发肿瘤组和肝转移瘤组。我们使用 t 检验的

方法分别计算正常组 VS 原发肿瘤组、正常组 VS 肝转移瘤组和原发肿瘤组 VS 肝转移瘤组。统计学分析工具为 Perseus 软件。

#### 4.3.4.2 差异表达蛋白 GO 功能注释与 KEGG 通路分析

使用常用的富集分析工具 DAVID(The database for annotation, visualization and integrated discovery)分别分析网络中差异蛋白参与 GO 功能和 KEGG 通路。参数富集基因个数  $\text{count} \geq 2$ ，超几何检验显著性阈值  $P < 0.05$  (视为显著富集结果)。

#### 4.3.4.3 蛋白互作网络 (protein-protein interaction network, PPI) 网络分析

使用 STRING 数据库预测分析差异表达基因编码的蛋白间互作关系，该数据库预测方法来源于基因邻接、基因融合、共现分析、共表达实验、数据库分析、系统发生谱和基因芯片数据的基因共表达。输入为差异表达蛋白，物种为人类。参数 PPI combine score 设为 0.4 (表示最高的置信度)，要求发生互作的蛋白质节点均为差异表达蛋白。然后使用 Cytoscape 软件构建网络图。蛋白质网络中的拓扑性质可被用来研究网络中的关键基因，我们使用网络中也性方法来研巧关键基因，即节点的连接度来描述蛋白质互作网络中节点的重要性。

## 4.4 结果

### 4.4.1 各组样本质谱分析获取蛋白情况分析

来自 18 个病人的匹配有正常组织、原发肿瘤组织以及肝转移组织的 58 个结肠癌 FFPE 样本，其中正常组织样本 18 个，原发肿瘤组织 21 个，肝转移组织样本各 19 个，共检测出 6052 种蛋白质，如下图 4.1 中所示。这些蛋白有 3200 种共同存在于正常组织、原发肿瘤组织和肝转移组织中，而 373 种、923 种和 764 种蛋白分别在正常组织、原发性肿瘤组织和肝转移组织所特有。该研究结果蛋白质组覆盖率非常高，说明我们的 FFPE 质谱样本处理方法切实可靠，不光有效的去除了去污剂 SDS，并且可能打开了或者部分打开了 FFPE 样本的化学交联。

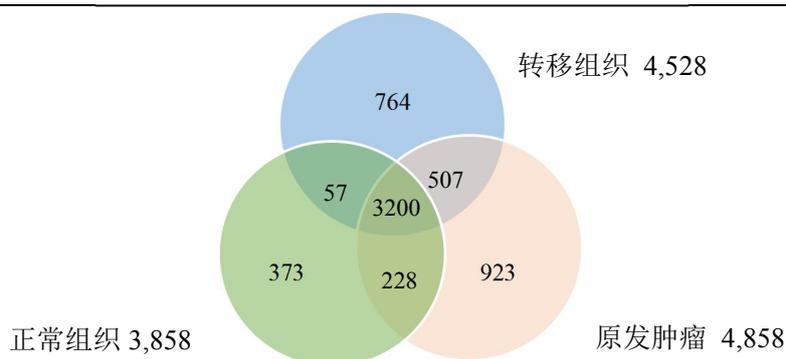


图 4.1 各组样本质谱分析获取蛋白情况

#### 4.4.2 不同样本间的稳健性、精确性和再现性比较以及数据标准化

为了验证我们新技术的稳健性、精确性和再现性，将其广泛的应用于基础及临床研究中，帮助临床诊断及药物开发等现实世界中。58 个样本加上 7 个质量控制 (QC) 样本一共分 10 次进行质谱分析。QC 样本来自所有已提取蛋白质溶液。我们对 QC 样本进行质控，研究结果发现，QC 样本之间的总蛋白质组含量相关性非常高，R2 相关系数始终大于 0.8，如下图 4.2 相关矩阵所示。

我们得到的 LFQ 强度样本数据可能是多个维度的，即一个样本是用多个特征来表征的，这些特征的量纲和数值得量级可能都是不一样的，故在统计学分析前将数据标准化处理，使其符合正态化分布，这样，不同特征对参数的影响程度就一样了，保证后续分析结果的准确性和可靠性。见图 4.3 所示。



图 4.2 各 QC 样本之间的 R2 相关性分析

## 第 4 章 基于液相色谱质谱 (LC-MS/MS) 联用的结肠癌 FFPE 组织差异表达蛋白的检测及肝转移过程的认识

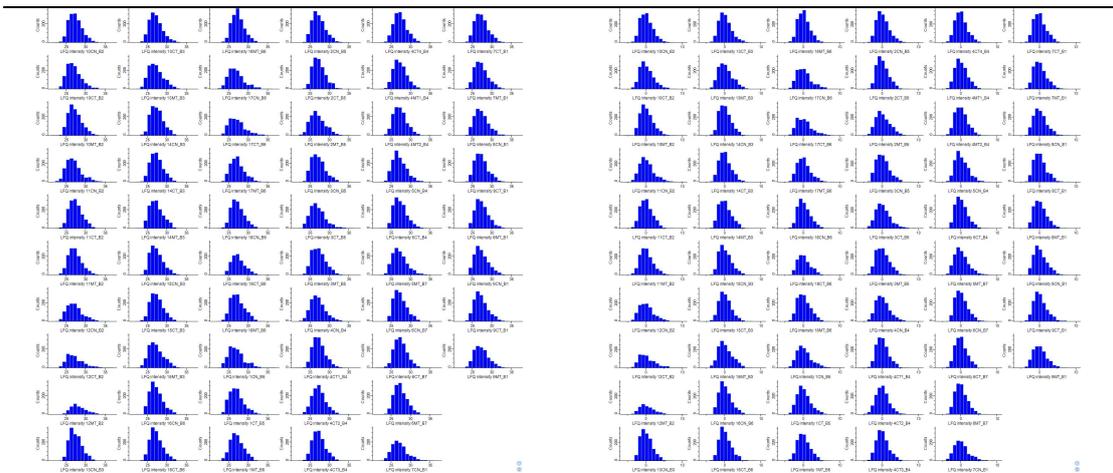
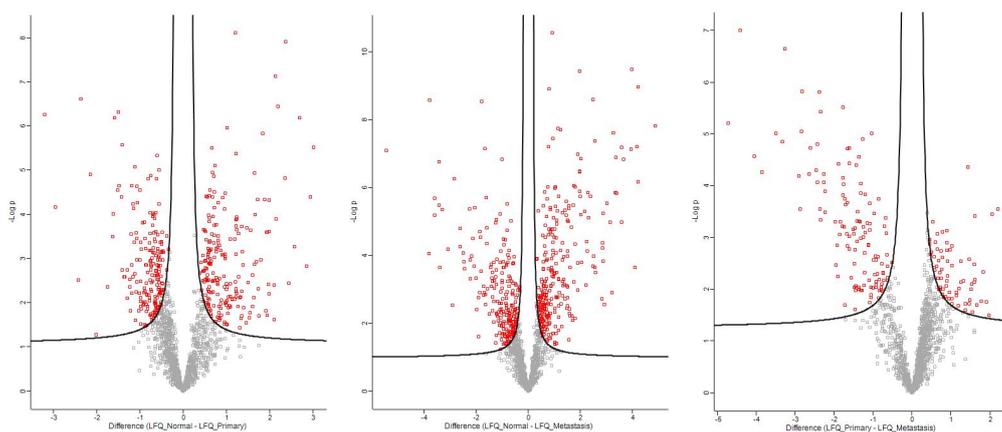


图 4.3 所有样本 LFQ 强度数据的标准化处理 (左: 标准化处理前, 右标准化处理后)

### 4.4.3 各组样本差异表达蛋白以及上调蛋白的聚类情况分析

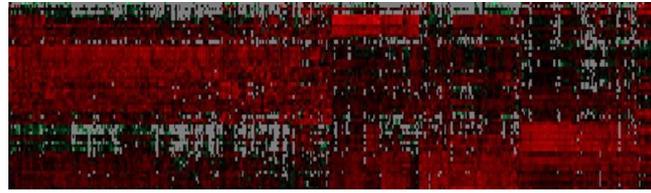
#### 4.4.3.1 差异表达蛋白的筛选结果

为了了解各组样本差异蛋白表达情况, 我们将 Maxquant 数据库搜索匹配后的 LFQ 数据导入 Perseus 软件进行分析, 分析结果提示正常组织组 VS 原发性肿瘤组、正常组织组 VS 肝转移组以及原发性肿瘤组 VS 肝转移组得到的差异表达蛋白分别是 411、502 和 193; 其中各组差异蛋白上调蛋白分别为 236、213 和 158。差异表达蛋白情况如图 4.4 所示, 上调蛋白表达聚类情况如图 4.5 红色部分所示。

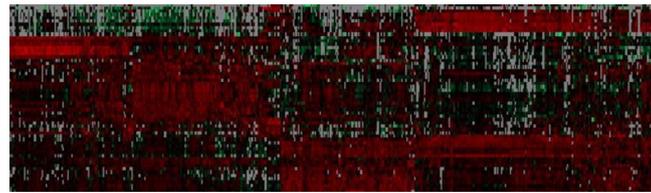


A. 正常组 VS 原发性肿瘤组 B. 正常组 VS 肝转移组 C. 原发性肿瘤组 VS 肝转移组

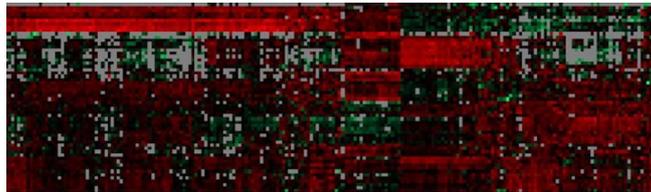
图 4.4 各组样本差异蛋白表达情况



A. 正常组 VS 原发肿瘤组



B. 正常组 VS 肝转移组



C. 原发肿瘤组 VS 肝转移组

图 4.5 各组样本差异上调蛋白表达聚类情况 (红色)

#### 4.4.3.2 差异表达蛋白的 GO 功能注释以及 KEGG 信号通路分析

我们利用 DAVID 数据库将各组差异表达的蛋白进行 GO 功能富集分析, 分析结果表明, 正常组 VS 原发肿瘤组差异表达蛋白功能主要集中在蛋白质折叠, 泛素蛋白质连接酶负性调节、RNA 的加工、tRNA 氨基酰化蛋白翻译、mRNA 稳定性调节和基因表达等功能上, 如图 4.6A 左所示; 而 KEGG 信号通路主要集中在 cGMP-PKG 信号通路、粘附连接、剪接体、血管平滑肌收缩以及紧密连接等信号通路上, 如图 4.6A 右所示。正常组 VS 肝转移组差异表达蛋白功能主要集中在 ATP 生物合成、线粒体呼吸复合体合成、黑质发育、肌肉收缩、有丝分裂细胞周期 G2/M 过渡等功能上, 如图 4.6B 左所示, 而 KEGG 信号通路主要集中在紧密连接、剪接体形成、粘附作用、细胞外基质受体相互作用、代谢途径、血管平滑肌收缩等信号通路上, 如图 4.6B 右所示。原发肿瘤组 VS 肝转移组差异表达蛋白功能主要集中在补体活化调节、酮体生物合成过程、糖原新生、蛋白质四聚体化、异型生物质代谢过程等功能上, 如图 4.6C 左所示, 而 KEGG 信号通路主要集中在氨基酸合成、丙酮酸代谢、精氨酸生物合成、糖酵解/葡萄糖

## 第 4 章 基于液相色谱质谱 (LC-MS/MS) 联用的结肠癌 FFPE 组织差异表达蛋白的检测及肝转移过程的认识

酵素、精氨酸与脯氨酸代谢途径上, 见图 4.6C 右所示。

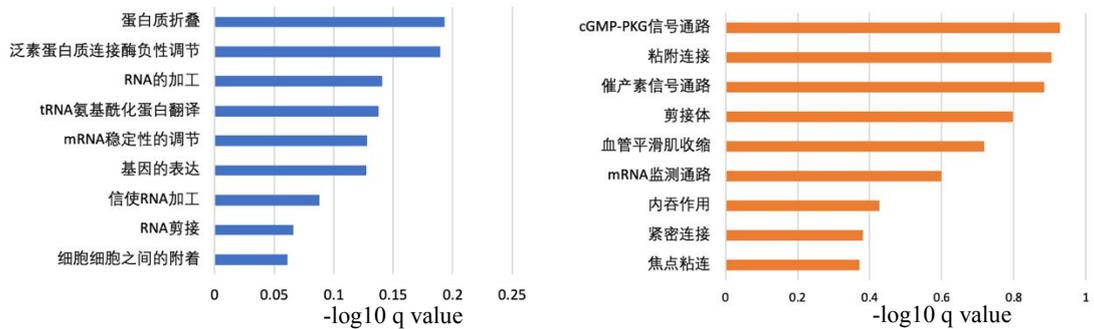


图 4.6A 正常组 VS 原发性肿瘤组差异表达蛋白 GO 功能注释(左)与 KEGG 信号通路 (右)

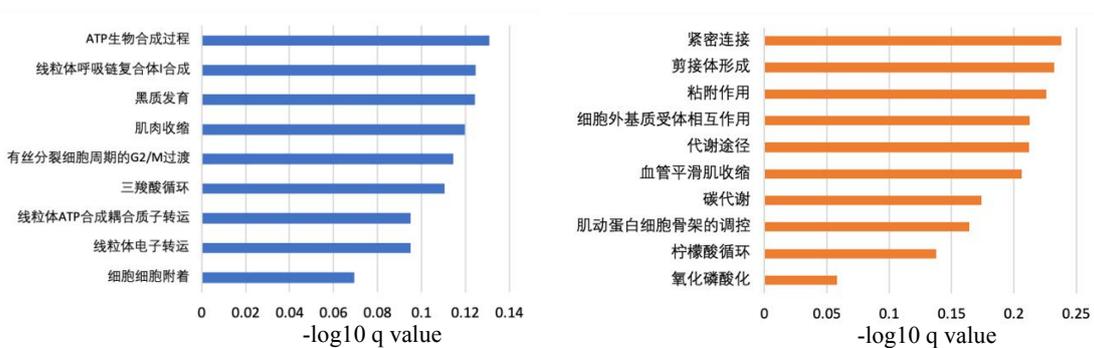


图 4.6B 正常组 VS 肝转移组差异表达蛋白 GO 功能注释(左)与 KEGG 信号通路 (右)

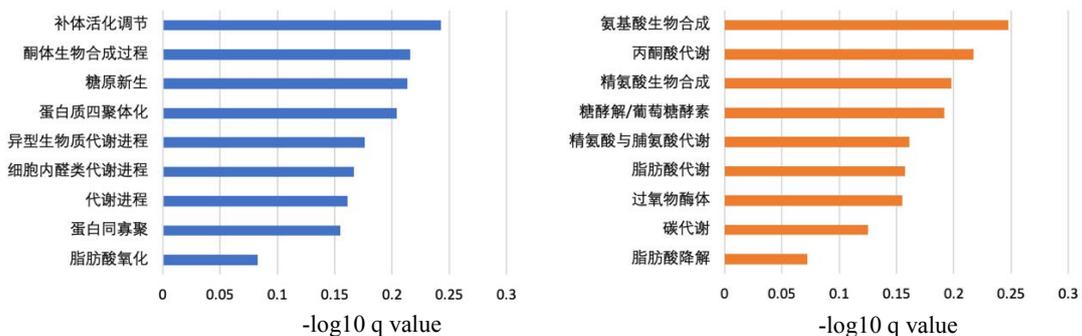


图 4.6C 原发性肿瘤组 VS 肝转移组差异表达蛋白 GO 功能注释(左)与 KEGG 信号通路 (右)

### 4.4.4 结肠癌肝转移过程的探索

#### 4.4.4.1 肝转移病变的无监督聚类分析

转移样本的无监督聚类分析检测到了 590 个表达蛋白, 这些蛋白被 70%或更多的结肠癌样本所共享。在转移灶中发现三个不同的大簇, 分别以聚类转移 1、

转移 2 和转移 3 表示。转移 1 有 188 个蛋白，转移 2 有 118 个蛋白，转移 3 有 284 个蛋白。见图 4.7 所示。

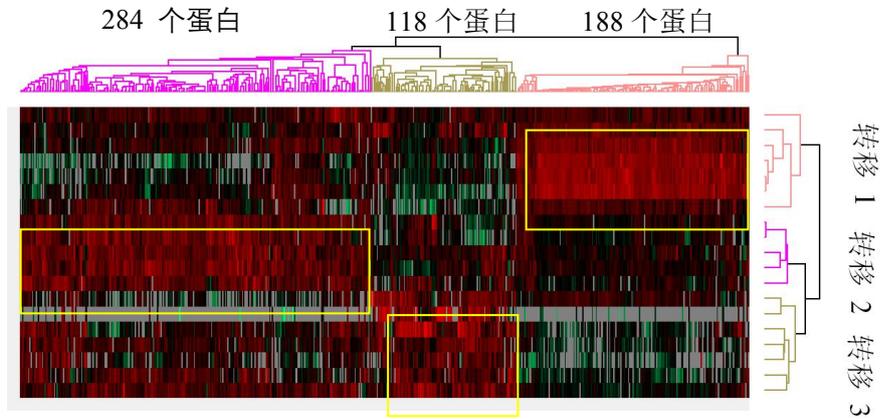


图 4.7 肝转移病变的无监督聚类热图

#### 4.4.4.2 肝转移病变差异表达基因富集 (GO) 功能注释

为了进一步了解结肠癌肝脏转移的过程，我们将所有共享蛋白和差异表达蛋白进行 GO 功能富集分析，分析结果表明，所有的共享蛋白功能主要集中在蛋白质翻译，端粒酶 RNA 的调控，氧化磷酸化和 mRNA 的剪接功能上，而转移特异性的差异表达蛋白除集中在蛋白质翻译、mRNA 的剪接功能上，还体现在转录后修饰、内吞作用以及线粒体呼吸功能途径上。见图 4.8 所示。

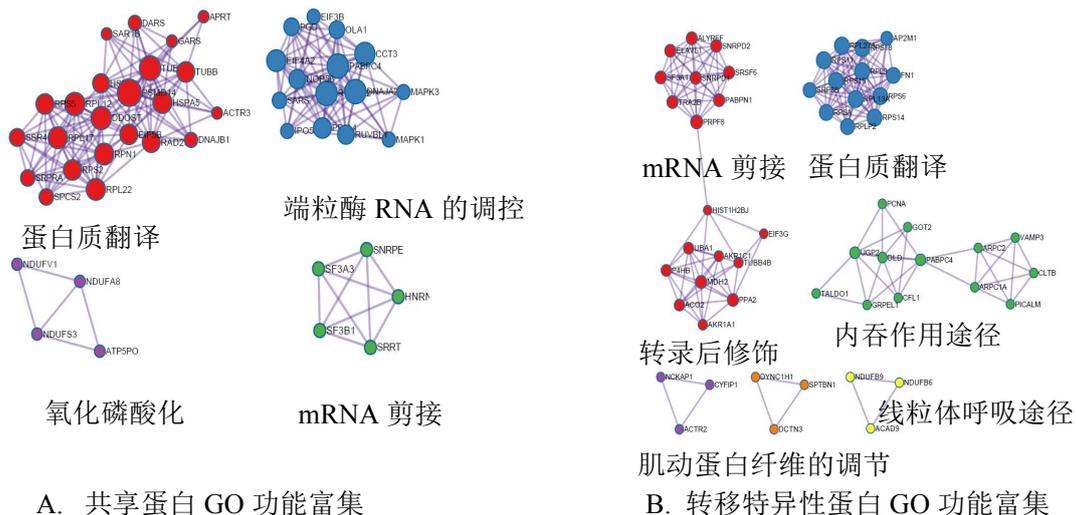


图 4.8 共享及差异表达蛋白 GO 功能富集分析

#### 4.4.4.3 肝转移病变差异表达聚类通路 (KEGG) 分析

通过肝转移病变差异蛋白表达 KEGG 通路分析,我们发现转移聚类 1 KEGG 通路主要富集在药物的代谢进程、异型物质代谢过程、分解代谢过程和氨基酸代谢的过程中;转移聚类 2 KEGG 通路主要富集在 mRNA 剪接、无义调节的 RNA 降解、细胞凋亡和 ATP 的合成等生物过程中;转移聚类 3 KEGG 通路主要富集在补体途径、凝血、肌肉结构以及类维生素 A 的代谢过程中。见图 4.9 所示。

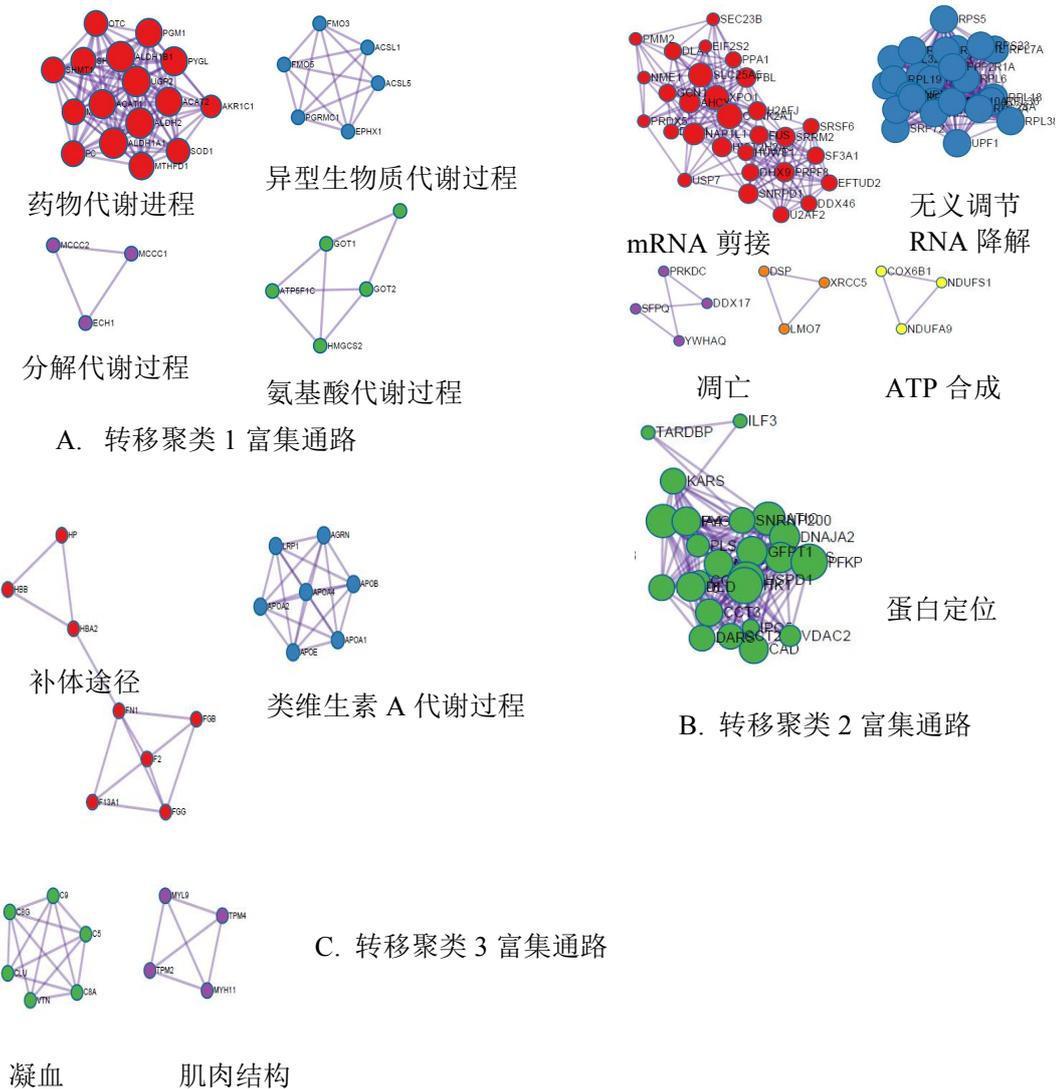


图 4.9 肝转移病变差异蛋白表达聚类的 KEGG 通路分析

#### 4.4.4.4 肝转移病变 PPI 网络分析

使用 STRING 数据库对肝脏转移病变的差异表达基因编码的 193 个蛋白间互作关系进行分析 (见图 4.10)。图中不同颜色的节点代表不同的蛋白质。剪接异构体或翻译后修饰被折叠, 即每个节点代表一个单一的蛋白质编码基因位点产生的所有蛋白质。直线代表蛋白相互作用, 其中深绿色和粉红色代表已知的蛋白相互作用 (深绿色代表蛋白之间相互作用来自已知数据库, 粉红色代表蛋白之间相互作用决定于不同的实验方法); 绿色、红色和蓝色代表预测的相互作用 (绿色代表基因邻接, 红色代表基因融合, 蓝色代表基因共现); 浅绿色、黑色和浅蓝色代表其它相互作用如蛋白同源性以及基因共表达等。这种相互作用是有特定意义的, 即蛋白质参与一个共同的功能, 并不意味着它们在物理上相互结合。

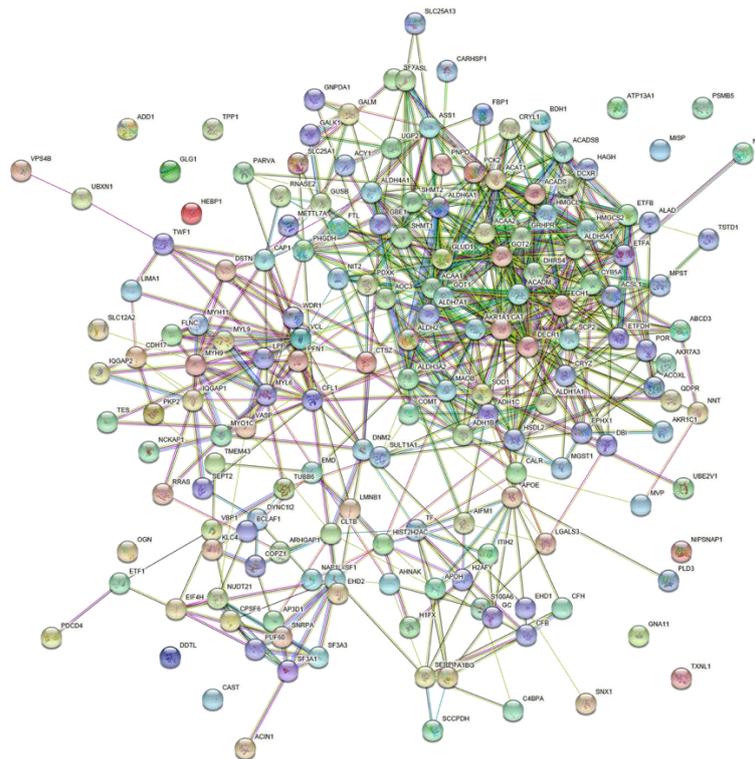
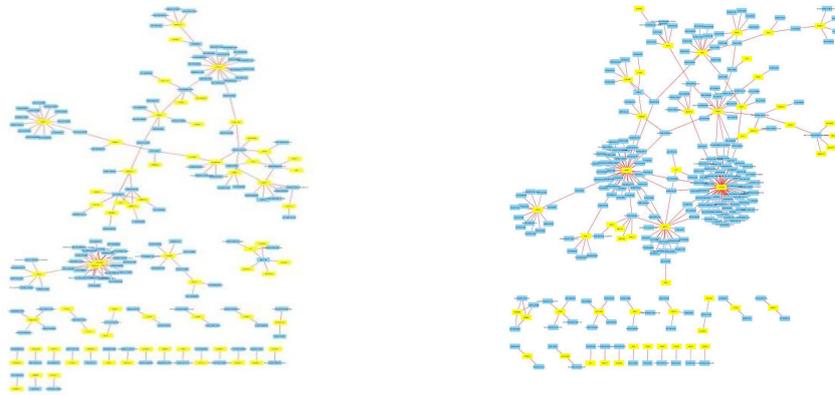


图 4.10 蛋白质相互作用网路图

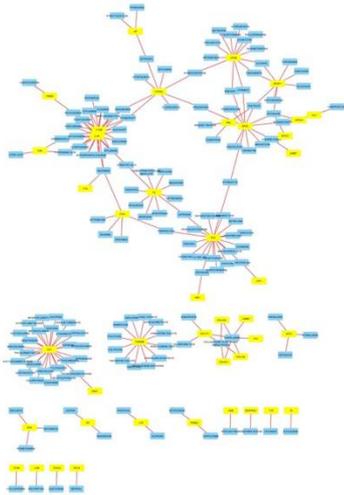
#### 4.4.4.5 结肠癌肝转移瘤特异性的上调蛋白药物治疗靶点

通过结肠癌肝转移特异性差异表达蛋白的聚类以及 GO 功能富集、KEGG 信号通路分析后，我们通过药物基因相互作用数据库 (DGIdb) 对特异性肝转移上调蛋白的药物靶点进行进一步分析。研究发现，转移聚类 1 的 66 种上调蛋白匹配有 FDA 批准的 137 种可能治疗有效的药物，转移聚类 2 的 54 种上调蛋白匹配有 FDA 批准的 230 种可能治疗有效的药物，而转移聚类 3 的 42 种上调蛋白匹配有 FDA 批准的 148 种可能治疗有效的药物。图 8 中黄色代表靶点蛋白，蓝色代表 FDA 批准的治疗药物。



A. 转移聚类 1 的 66 种上调蛋白匹配有 FDA 批准的 137 种可能治疗有效的药物

B. 转移聚类 2 的 54 种上调蛋白匹配有 FDA 批准的 230 种可能治疗有效的药物



C. 转移聚类 3 的 42 种上调蛋白匹配有 FDA 批准的 148 种可能治疗有效的药物

图 4.11 结肠癌肝转移瘤特异性的上调蛋白药物治疗靶点

## 4.5 讨论

近年来, 无凝胶法被广泛应用于定量蛋白质组学。通常由蛋白质酶解产生复杂的多肽混合物, 并且使用不同的分馏策略(如反相液相色谱)。同位素标记是常用的定量蛋白质或多肽的方法。典型的同位素标记方法包括使用同位素编码亲和标记(ICAT)<sup>[90,91]</sup>, 使用等压标记进行相对和绝对量化(iTRAQ)<sup>[92]</sup>, 或使用细胞培养中的氨基酸进行稳定同位素标记(SILAC)<sup>[93]</sup>。化学标记的其他变体包括非异压胺标记(mTRAQ)、二甲基标记和同位素编码蛋白标记(ICPL)<sup>[94-96]</sup>。这些标记方法在量化方面是可靠的, 但是它们需要使用昂贵的分子、特定的软件和一定的熟悉程度来分析结果。而且, 可以同时分析的样本数量受可用标签数量的限制, 并不是所有的策略都适用于所有类型的样本。

新的仪器和先进的生物信息学帮助开启了无标签蛋白质组学的时代<sup>[97]</sup>。无标签量化, 是一个直接的、廉价的而且不涉及同位素标记的策略。有两种测量方法可以获得定量结果: 峰面积(或离子强度)和光谱计数。峰面积法是根据提取的离子色谱的总峰面积来测量分析物数量的一种方法。其原理是检测到的离子信号与分析物浓度成正比, 这种无标签方法的实际局限性与 LC-MS 方法有关, LC-MS 方法必须具有高度的可重复性, 任何在滞留时间和  $m/z$  上的漂移都将极大地复杂化峰对齐过程; 其次, 使用的 MS 仪器必须是高分辨率的, 因为低分辨率仪器在区分重叠峰时是可变的。另一种基于光谱计数的定量测量方法, 是一种简单直接的测量策略。光谱计数定义为分配给一个蛋白质的 MS/MS 的数量。因此, 在 LC-MS/MS 分析过程中, 它通过观察一个肽段的 MS/MS 谱的数量来测量属于一个蛋白的每个肽段执行 MS/MS 事件的次数, 这与蛋白质的丰度成正比。这种方法的一个缺点是, 由于离子竞争、动态范围限制和 MS 敏感性, 并不是所有的肽都能以相同的方式检测到<sup>[98]</sup>。然而, 无标签质谱已广泛应用于 CRC 生物标志物的发现。本研究利用无标签质谱法, 选用最新一代的高分辨傅立叶变换 Orbitrap Fusion Lumos 质谱仪, 单次注入鸟枪法蛋白组学分析方法, 尝试通过基于 LC-MS/MS 联用的结肠癌 FFPE 组织来发现新的结肠癌的生物标志物并了解其肝转移过程。

患者的生物样本是最有价值的蛋白质组学标记发现的生物系统。我们选取来自 18 个病人的匹配有正常组织、原发肿瘤组织以及肝转移组织的 58 个结肠癌 FFPE 样本, 通过优化 FFPE 组织蛋白提取缓冲液, 采用 S-Trap 蛋白消化方法,

共检测出 6052 种蛋白质，该研究结果蛋白质组覆盖率非常高，几乎与目前使用较为成熟的新鲜冰冻组织或 OCT 组织相当或者接近，说明我们的 FFPE 质谱样本处理方法切实可靠，不光有效的去除了去污剂 SDS，并且可能打开了或者部分打开了 FFPE 样本的化学交联，所有样本与高灵敏度高分辨率的质谱仪兼容性良好，样本之间的相关性很高，我们对比了所有质控样本的相关性，R2 均在 0.8 以上。

我们比较了各组样本之间的差异蛋白表达情况，并对上调蛋白进行分析，见图 4.4 和图 4.5，这些蛋白可能是 CRC 新的候选生物标志物。我们相信在未来几年还会发现其他 CRC 相关蛋白和潜在的生物标志物。重要的是，CRC 的异质性要求有多个标记，每个标记代表一种不同的 CRC 亚型。一个单一的生物标记很难既非常敏感又非常具体。因此，蛋白质组学的目标应该是识别特定的蛋白质特征，以便对患者进行最佳的分层和分类，从而简化临床管理。

事实上，在使用不同实验策略进行的不同研究中，候选生物标志物的列表之间通常不存在或几乎不存在重叠。不同的蛋白质组学方法应该被认为是互补的，而不是替代的，多个蛋白质组学技术的并行应用导致候选蛋白质的重叠列表将有更多的机会提出有用的标记。值得注意的是，蛋白质组学研究已经提出了大量新的假定的生物标志物，但在临床实践中这些标志物的引入并没有取得显著的改善。在这方面，一个关键的步骤是标记验证策略，应该使用最标准化的技术，并使用来自统计上显著数量的患者的样本进行验证。最佳做法应是使用妥善回收和储存的临床组织样本的标准化档案，并提供适当的临床文件和有关临床结果的详细说明。此外，基于新型生物标志物的新检测方法的性能应严格与大量(成百上千)患者的现有检测方法(如 FOBT)进行比较。最后，规范的实验室间的检测分析和验证方案，将有助于克服发现策略和结果临床应用之间的瓶颈<sup>[99]</sup>。

为了进一步了解结肠癌肝脏转移的过程，我们对肝转移病变的进行无监督聚类分析，如图 8 所示，在转移病变中发现三个不同的聚类相对应着差异表达的蛋白。GO 功能富集发现所有的共享蛋白功能主要集中在蛋白质翻译，端粒酶 RNA 的调控，氧化磷酸化和 mRNA 的剪接功能上，而转移特异性的差异表达蛋白除集中在蛋白质翻译、mRNA 的剪接功能上，还体现在转录后修饰、内吞作用以及线粒体呼吸功能途径上，如图 9 所示。KEGG 通路分析发现转移病变中不同的表达聚类参与着不同的代谢和合成过程如图 10 所示。

我们再通过 DGIdb 数据库对特异性肝转移上调蛋白的药物治疗靶点进行进一步分析发现, 不同的上调蛋白的表达, 可能有多种 FDA 批准的转移聚类 1 的 66 种上调蛋白匹配有 FDA 批准的 137 种可能治疗有效的药物, 转移聚类 2 的 54 种上调蛋白匹配有 FDA 批准的 230 种可能治疗有效的药物, 而转移聚类 3 的 42 种上调蛋白匹配有 FDA 批准的 148 种可能治疗有效的药物。这些结果加深了我们对癌症转移生物学性质的理解, 并可能导致新的药物靶点和预后标志物。

## 4.6 结论

将深入蛋白质组分析应用于原发性和肝转移性结肠癌的同步组, 揭示了转移中的新的特异性信号通路和功能簇聚类。这些结果加深了我们对癌症转移生物学性质的理解, 并可能导致新的药物靶点和预后标志物。

## 第 5 章 全文总结与展望

本研究主要以 FFPE 结肠癌组织为研究对象，利用基于液相色谱质谱联用的蛋白质组学研究方法以期发现结肠癌病变的新的生物标志物，并通过生物信息学分析方法对结肠癌肝转移生物进程进行分析，以期发现新的治疗靶点和药物。

FFPE 组织是临床研究的一个潜在的有价值的标本来源，广泛储存在全球范围内各大医院样本库中；然而，因其含有共价蛋白交联、低蛋白溶解率以及肽段回收率等，目前尚缺乏有效的 FFPE 组织蛋白质组学研究方法或标准流程。

自下而上蛋白质组学研究方法中，FFPE 样本注入质谱分析仪之前，合格标本的准备非常重要。从石蜡包埋的蜡块组织中提取足量的蛋白质是实验成功的第一步；蛋白提取之后，通过消化酶（通常是胰酶）将其消化为肽段以供质谱分析是实验成功的第二步。然而，到目前为止，对于 FFPE 组织的研究来说这两个重要的过程恰恰是目前难以逾越的瓶颈。

幸运的是，我们的研究突破了这两大瓶颈。首先，我们通过阅读大量的文献，选取目前文献报道蛋白提取效率较高的含有浓度相对较高的去污剂 SDS 和还原剂 DTT 的蛋白提取缓冲液并对其组成成份和实验流程进行优化，成功从 FFPE 结肠癌组织中提取了蛋白。

接下来，成功提取蛋白之后的肽段获取即蛋白质的消化是整个实验成功的关键。对于细胞株或者其它组织来说，溶液内消化和凝胶内消化实验技术已经非常成熟。但是对于 FFPE 组织来说，因其含有共价的蛋白交联、蛋白溶解率较低以及其蛋白提取缓冲液内含有与质谱分析仪不兼容的离子型和非离子型的去污剂 SDS 等原因，常规的溶液内和凝胶内消化方法肽段回收率非常低。因此，我们选用新近报道的 S-Trap 消化方法，并与溶液内消化和凝胶内消化方法的肽段获取效率进行了比较，研究结果表明，S-Trap 消化方法的肽段获取效率远远高于另外两种方法（见第 3 章），重要的是，S-Trap 允许样本中含有高浓度的 SDS。

我们获取来自 18 个病人匹配有正常组织、原发性肿瘤组织以及肝转移组织的总共 58 个 FFPE 结肠癌样本，通过我们优化的蛋白提取方法提取蛋白，以 S-Trap 消化方法获取肽段，以最新一代的傅立叶变换高灵敏度质谱分析仪对所准备的样本进行分析，共检测出 6052 种蛋白质。该研究结果显示蛋白质组覆盖率非常高，远远超过目前文献报道的数据。说明我们的 FFPE 质谱样本处理方法切实可

靠，不光有效的去除了去污剂 SDS，并且可能打开了或者部分打开了 FFPE 样本的化学交联。

最后，我们以生物信息学方法对质谱分析所得的原始数据进行分析得到可能为结肠癌组织的生物标志物的差异表达的蛋白，并将这些上调蛋白和下调蛋白进行 GO 功能分析和 KEGG 信号通路分析以及蛋白蛋白相互作用分析等，得到了这些差异表达蛋白等功能以及信号通路的聚类，并进一步的对结肠癌组织肝转移过程有了更深层次的认识。通过 DGIdb 数据库初步筛选出了一些与肿瘤治疗有关的 FDA 批准的药物。

总之，我们的研究成功的优化了 FFPE 结肠癌组织的的蛋白提取效率，开发了允许含有高浓度离子型去污剂 SDS 的并与质谱分析兼容的 S-Trap 蛋白消化方法，突破了目前 FFPE 组织作为研究对象难以逾越的瓶颈。同时，发现了可能作为结肠癌生物标记物的差异表达蛋白，对其进行功能注释以及信号通路分析；进一步通过蛋白组学研究的方法对结肠癌肝转移生物进程深入的理解，为结肠癌的诊断和治疗提供新的靶点。

我们相信我们的研究已经打开了一个全球范围内的巨大样本资源库（FFPE 组织资源库），我们将会发现更多的 CRC 特异性组织生物标志物，这些标志物可以促进新药的开发，未来可能会有更多的蛋白质组学研究关注这些细胞的生物标志物，从而为特定 CRC 的靶向治疗铺平道路。

## 致 谢

本课题在课题设计、资料分析和论文撰写等方面得到导师程晓曙教授的精心指导,在此首先衷心感谢导师程晓曙教授三年来对我的临床实践和教学科研上的严格要求和悉心培养!

在三年的博士研究生学习中,导师程晓曙教授严谨求实的科学作风、广博丰富的理论和实践知识,给我留给我极其深刻的印象,使我终生受益。特别是在我在美国学习期间,在学习生活上碰到什么困难之处,他都毫不犹豫的第一时间给予我帮助。

衷心感谢南昌大学二附院心内科和综合 ICU 的全体老师。

感谢同门师兄、师姐、师弟、师妹们对我的帮助和支持。

最后,我还要向多年来一直鼓励和支持我的家人献上最真诚的谢意,是她们在精神上和生活上的无私奉献,给了我知难而进的勇气和动力,才使自己的学业攀上了一个新的平台!

2019 年 5 月

## 参考文献

- [1] Newton K, Newman W, Hill J. Review of biomarkers in colorectal cancer. *Colorectal Dis.* 2012,14(1):3-17.
- [2] Hagggar FA, Boushey RP. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin colon rectal surg.* 2009, 22(4):191-7.
- [3] Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin.* 2012, 62(1):10-29.
- [4] Coghlin C, Murray GI. Progress in the identification of plasma biomarkers of colorectal cancer. *Proteomics.* 2013, 13(15):2227-8.
- [5] Moreno CC, Mittal PK, Sullivan PS, et al. Colorectal Cancer Initial Diagnosis: Screening Colonoscopy, Diagnostic Colonoscopy, or Emergent Surgery, and Tumor Stage and Size at Initial Presentation. *Clin Colorectal Cancer.* 2016,15(1):67-73.
- [6] Adler A, Geiger S, Keil A, et al. Improving compliance to colorectal cancer screening using blood and stool based tests in patients refusing screening colonoscopy in Germany. *BMC Gastroenterology.* 2014, 14:183.
- [7] Taylor DP, Cannon-Albright LA, Sweeney C, et al. Comparison of compliance for colorectal cancer screening and surveillance by colonoscopy based on risk. *Genet Med.* 2011, 13(8):737-43.
- [8] Menees SB, Kim HM, Elliott EE, et al. The impact of fair colonoscopy preparation on colonoscopy use and adenoma miss rates in patients undergoing outpatient colonoscopy. *Gastrointestinal Endoscopy.* 2013, 78(3):510-6.
- [9] Gemei M, Corbo C, Salvatore F, et al. Carcinoembryonic Antigen Family Cell Adhesion Molecules (CEACAM) as Colorectal Cancer Biomarkers. *Biomarkers in Cancer.* 2015, 685-705.
- [10] Tanaka T, Tanaka M, Tanaka T, et al. Biomarkers for colorectal cancer. *Int J Mol sci.* 2010, 11(9):3209-25.
- [11] Tan E, Gouvas N, Nicholls RJ, et al. Diagnostic precision of carcinoembryonic antigen in the detection of recurrence of colorectal cancer. *Surg Oncol.* 2009, 18(1):15-24.
- [12] Araujo RL, Gonen M, Allen P, et al. Positive Postoperative CEA is a Strong Predictor of Recurrence for Patients After Resection for Colorectal Liver Metastases. *Annals of Surgical Oncology.* 2015, 22(9):3087-93.
- [13] Fakih MG, Padmanabhan A. CEA monitoring in colorectal cancer. What you should know. *Oncology.* 2006, 20(6):579-87.
- [14] Park IJ, Choi GS, Lim KH, et al. Serum Carcinoembryonic Antigen Monitoring After Curative Resection for Colorectal Cancer: Clinical Significance of the Preoperative Level. *Ann Surg Oncol.* 2009, 16(11):3087-93.

- [15] Duffy MJ. Carcinoembryonic antigen as a marker for colorectal cancer: Is it clinically useful? *Clin Chem.* 2001, 47(4):624-30.
- [16] Lyall MS, Dundas SR, Curran S, et al. Profiling markers of prognosis in colorectal cancer. *Clin Cancer Res.* 2006, 12(4):1184-91.
- [17] Kuipers EJ, Rosch T, Bretthauer M. Colorectal cancer screening--optimizing current strategies and new directions. *Nat Rev Clin Oncol.* 2013, 10(3):130-42.
- [18] Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat Clin Pract Oncology.* 2008, 5(10):588-99.
- [19] Hayes DF, Bast RC, Desch CE, et al. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst.* 1996, 88(20), 1456-66.
- [20] Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell.* 2000, 100(1):57-70.
- [21] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011, 144(5):646-74.
- [22] Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet.* 2010, 11(10), 685-96.
- [23] Hanash SM, Baik CS, Kallioniemi O. Emerging molecular biomarkers--blood-based strategies to detect and monitor cancer. *Nat Rev Clin Oncol.* 2011, 8(3), 142-50.
- [24] Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem.* 2003, 72:291-336.
- [25] Torto-Alalibo T, Purwantini E, Lomax J, et al. Genetic resources for advanced biofuel production described with the Gene Ontology. *Front Microbiol.* 2014,5:528.
- [26] Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000,28(1):27-30.
- [27] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009,4(1):44-57.
- [28] Szklarczyk D, Franceschini A, Wyder S et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015, 43:D447-52.
- [29] Shannon P, Markiel A, Ozier O et al. Cytoscape: a software environment for integrated models. of biomolecular interaction networks. *Genome Res.* 2003,13(11):2498-504.
- [30] Tanca A, Pagnozzi D, Addis F. Setting proteins free: progresses and achievements in proteomics of formalin-fixed, paraffin-embedded tissues. *Proteomics Clin Appl.* 2012, 6(1-2):7-21.
- [31] Nirmalan NJ, Harnden P, Selby PJ, et al. Mining the archival formalin-fixed paraffin-embedded tissue proteome: opportunities and challenges. *Mol BioSyst.* 2008, 4(7):712-20.
- [32] Klopffleisch R, Weiss AT, Gruber AD. Excavation of a buried treasure--DNA, mRNA, miRNA and protein analysis in formalin fixed, paraffin embedded tissues. *Histol Histopathol.* 2011, 26(6):797-810.
- [33] Reimel BA, Pan S, May DH, et al. Proteomics on Fixed Tissue Specimens-A Review. *Curr*

- Proteomics. 2009, 6(1):63–69.
- [34] Magdeldin S, Yamamoto T. Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues. *Proteomics*. 2012, 12(7):1045–58.
- [35] Thavarajah R, Mudimbaimannar VK, Elizabeth J, et al. Chemical and physical basics of routine formaldehyde fixation. *J Oral Maxillofac Pathol*. 2012, 16(3):400–5.
- [36] Maes E, Broeckx V, Mertens I, et al. Analysis of the formalin-fixed paraffin-embedded tissue proteome: pitfalls, challenges, and future perspectives. *Amino Acids*. 2013, 45(2):205–18.
- [37] Ralton LD and Murray GI. The use of formalin fixed wax embedded tissue for proteomic analysis. *J Clin Pathol*. 2011, 64(4): 297–302.
- [38] Metz B, Kersten GF, Baart GJ, et al. Identification of formaldehyde-induced modifications in proteins: reactions with insulin. *Bioconjugate Chem*. 2006,17(3):815-22.
- [39] B Metz, Kersten GF, Hoogerhout P, et al. Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. *J Biol Chem*.2003, 279(8):6235-43.
- [40] Toews J, Rogalski JC, Kast J. Accessibility governs the relative reactivity of basic residues in formaldehyde-induced protein modifications. *Anal Chim Acta*. 2010, 676(1-2): 60–7.
- [41] Fox CH, Johnson FB, Whiting J, et al. Formaldehyde fixation. *J Histochem Cytochem*.1985, 33(8):845-53.
- [42] Klockenbusch C, O’Hara JE, Kast J. Advancing formaldehyde cross-linking towards quantitative proteomic applications. *Anal Bioanal Chem*. 2012, 404(4):1057-67.
- [43] Dapson RW. Macromolecular changes caused by formalin fixation and antigen retrieval. *Biotech Histochem*. 2007, 82(3):133-40.
- [44] Ikeda K, Monden K, Kanoh T, et al. Extraction and analysis of diagnostically useful proteins from formalin-fixed paraffin-embedded tissue sections. *J Histochem Cytochem*.1998, 46(3): 397-403.
- [45] Addis MF, Tanca A, Pagnozzi D, et al. 2-D PAGE and MS analysis of proteins from formalin-fixed, paraffin-embedded tissues. *Proteomics*. 2009, 9(18):4329-39.
- [46] Fowler CB, Chesnick IE, Moore CD, et al. Elevated pressure improves the extraction and identification of proteins recovered from formalin-fixed, paraffin-embedded tissue surrogates. *PLoS One*. 2010, 5(12):e14253.
- [47] Donadio E, Giusti L, Cetani F, et al. Evaluation of formalin-fixed paraffin-embedded tissues. in the proteomic analysis of parathyroid glands. *Proteome Sci*. 2011, 9(1):29.
- [48] Fowler CB, Waybright TJ, Veenstra TD, et al. Pressure-assisted protein extraction: a novel method for recovering proteins from archival tissue for proteomic analysis. *J. Proteome Res*. 2012, 11(4): 2602-08.
- [49] Maes E, Valkenborg D, Mertens I, et al. Proteomic analysis of formalin-fixed paraffin-embedded colorectal cancer tissue using tandem mass tag protein labeling. *Mol BioSyst*. 2013, 9(11):2686-95.
- [50] Fowler CB, O’Leary TJ, Mason TJ. Improving the Proteomic Analysis of Archival Tissue by

- Using Pressure-Assisted Protein Extraction: A Mechanistic Approach. *J Proteomics Bioinf.* 2014, 7(6):151–7.
- [51] Addis MF, Tanca A, Pagnozzi D, et al. Generation of high-quality protein extracts from formalin-fixed paraffin-embedded tissues. *Proteomics.* 2009, 9(15):3815-23.
- [52] Davalieva K, Kiprijanovska S, Polenakovic M et al. Assessment of the 2-d gel-based proteomics application of clinically archived formalin-fixed paraffin embedded tissues. *Protein J.* 2014, 33(2):135-42.
- [53] Hood BL, Conrads TP, Veenstra TD. Unravelling the proteome of formalin-fixed paraffin-embedded tissue. *Brief Funct Genomic Proteomic.* 2006, 5(2):169-75.
- [54] Jiang X, Jiang X, Feng S, et al. Development of efficient protein extraction methods for shotgun proteome analysis of formalin-fixed tissues. *J Proteome Res.* 2007, 6(3):1038-47.
- [55] Sprung RW, Brock JW, et al. Equivalence of protein inventories obtained from formalin-fixed paraffin-embedded and frozen tissue in multidimensional liquid chromatography-tandem mass spectrometry shotgun proteomic analysis. *Mol Cell Proteomics.* 2009, 8(8):1988-98.
- [56] Bateman NW, Sun M, Bhargava R, et al. Differential proteomic analysis of late-stage and recurrent breast cancer from formalin-fixed paraffin-embedded tissues. *J Proteome Res.* 2011, 10(3):1323-32.
- [57] Hood BL, Darfler MM, Guiel TG, et al. Proteomic analysis of formalin-fixed prostate cancer tissue. *Mol Cell Proteomics.* 2005, 4(11):1741-53.
- [58] Matsukawa S, Morita K, Negishi A, et al. Galectin-7 as a potential predictive marker of chemo- and/or radio-therapy resistance in oral squamous cell carcinoma. *Cancer Med.* 2014, 3(2):349-61.
- [59] Takadate T, Onogawa T, Fukuda T, et al. Novel prognostic protein markers of resectable pancreatic cancer identified by coupled shotgun and targeted proteomics using formalin-fixed paraffin-embedded tissues. *Int J Cancer,* 2013, 132(6):1368-82.
- [60] Tanca A, Addis MF, Pagnozzi D, et al. Proteomic analysis of formalin-fixed paraffin-embedded lung neuroendocrine tumor samples from hospital archives. *J Proteomics.* 2011, 74(3):359-70.
- [61] Jackson D, Rowlinson RA, Eaton CK, et al. Prostatic tissue protein alterations due to delayed time to freezing. *Proteomics.* 2006, 6(13):3901-08.
- [62] Kokkat TJ, Patel MS, McGarvey D, et al. Archived formalin-fixed paraffin-embedded (FFPE) blocks: A valuable underexploited resource for extraction of DNA, RNA, and protein. *Biopreserv Biobank.* 2013, 11(2):101-06.
- [63] Tanca A, Pagnozzi D, Falchi G, et al. Impact of fixation time on GeLC-MS/MS proteomic profiling of formalin-fixed, paraffin-embedded tissues. *J Proteomics.* 2011, 74(7):1015-21.
- [64] Xie R, Chung JY, Ylaya K, et al. Factors influencing the degradation of archival formalin-fixed paraffin-embedded tissue sections. *J Histochem Cytochem.* 2011, 59(4):356-65.

- [65] Zhong J, Ye Z, Clark CR, et al. Enhanced and controlled chromatin extraction from FFPE tissues and the application to ChIP-seq. *BMC Genomics*. 2019, 20(1):249.
- [66] Pessôa-Pereira D, Medeiros MFDS, Lima VMS et al. Association between BRAF (V600E) mutation and clinicopathological features of papillary thyroid carcinoma: a Brazilian single-centre case series. *Arch Endocrinol Metab*. 2019,63(2):97-106.
- [67] Tanca A, Pagnozzi D, Burrai GP, et al. Comparability of differential proteomics data generated from paired archival fresh-frozen and formalin-fixed samples by GeLC-MS/MS and spectral counting. *J Proteomics*. 2012, 77:561-76.
- [68] Magdeldin S, Yamamoto T. Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues. *Proteomics*. 2012, 12(7):1045-58.
- [69] Chung JY, Lee SJ, Kris Y et al. A well-based reverse-phase protein array applicable to extracts from formalin-fixed paraffin-embedded tissue. *Proteomics Clin Appl*. 2008, 2(10-11): 1539–47.
- [70] Azimzadeh O, Barjaktarovic Z, Aubele M. Formalin-fixed paraffin-embedded (FFPE) proteome analysis using gel-free and gel-based proteomics. *J Proteome Res*. 2010, 9(9):4710–20.
- [71] Sprung RW, Brock JW, Tanksley JP et al. Equivalence of protein inventories obtained from formalin-fixed paraffin-embedded and frozen tissue in multidimensional liquid chromatography-tandem mass spectrometry shotgun proteomic analysis. *Mol Cell Proteomics*. 2009, 8(8):1988–98.
- [72] Yamashita S. Heat-induced antigen retrieval: mechanisms and application to histochemistry. *Prog Histochem Cytochem*. 2007, 41(3):141-200.
- [73] Tanca A, Pagnozzi D, Addis MF. Setting proteins free: progresses and achievements in proteomics of formalin-fixed, paraffin-embedded tissues. *Proteomics Clin Appl*. 2012, 6(1-2):7–21.
- [74] Weston LA, Bauer KM, Hummon AB. Comparison of bottom-up proteomic approaches for LC-MS analysis of complex proteomes. *Anal Methods*. 2013, 5 (18): 4615.
- [75] Glatter T, Ahrne E, Schmidt A, et al. Comparison of different sample preparation protocols. reveals lysis buffer-specific extraction biases in gram-negative bacteria and human cells. *J Proteome Res*. 2015, 14 (11):4472-85.
- [76] Mostovenko E, Hassan C, Rattke J. Comparison of peptide and protein fractionation methods in proteomics. *EuPa Open Proteomics*. 2013, 1, 30-37.
- [77] Cao Z, Tang HY, Wang H, et al. Systematic comparison of fractionation methods for in-depth analysis of plasma proteomes. *J Proteome Res*. 2012, 11 (6):3090–100.
- [78] Tanca A, Abbondio M, Pisanu S, et al. Critical comparison of sample preparation strategies for shotgun proteomic analysis of formalin-fixed, paraffin-embedded samples: Insights from liver tissue. *Clin Proteomics*. 2014, 11 (1):28.
- [79] Burkhart JM, Schumbrutzki C, Wortelkamp S, et al. Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS- based

- proteomics. *J Proteomics*. 2012, 75 (4):1454-62.
- [80] Wisniewski JR, Zougman A, Mann M. Combination of FASP and Stage Tip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J Proteome Res*. 2009, 8 (12):5674-78.
- [81] Wisniewski JR; Ostasiewicz P, Dusík K, Zielinska DF. Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol Syst Biol*. 2012,8:611.
- [82] Wisniewski JR, Nagaraj N, Zougman A, et al. Brain phosphoproteome obtained by a fasp-based method reveals plasma membrane protein topology. *J Proteome Res*. 2010, 9 (6):3280-89.
- [83] Zougman, A, Selby PJ, Banks RE. Suspension trapping (STrap) sample preparation method for bottom-up proteomics analysis. *Proteomics*. 2014, 14 (9):1006-10.
- [84] Ludwig KR, Schroll MM, Hummon AB. Comparison of In-Solution, FASP, and S-Trap Based Digestion Methods for Bottom-Up Proteomic Studies. *J Proteome Res*. 2018,17 (7):2480-90.
- [85] Renehan AG, Egger M, Saunders MP, et al. Impact on survival of intensive follow up after curative resection for colorectal cancer: systematic review and meta-analysis of randomised trials. *BMJ*. 2002,324(7341):813.
- [86] 许剑民, 汤文涛, 朱德祥. 结直肠癌肝转移的治疗进展. *中华消化外科杂志*. 2016,15(2): 107-12.
- [87] Abdalla EK, Vauthey JN, Ellis LM, et al. Recurrence and outcomes following hepatic resection, radiofrequency ablation, and combined resection/ablation for colorectal liver metastases. *Ann Surg*. 2004,239(6):818-25.
- [88] Misiakos EP, Karidis NP, Kouraklis G. Current treatment for colorectal liver metastases. *World J Gastroenterol*. 2011,17(36):4067-75.
- [89] Prescrire Editorial Staff. Chemotherapy of metastatic colorectal cancer. *Prescrire Int*. 2010,19(109):219-24.
- [90] Gygi SP, Rist B, Gerber SA, et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology*. 1999, 17(10):994-9.
- [91] Han DK, Eng J, Zhou H, et al. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nature biotechnology*. 2001, 19(10):946-51.
- [92] Ross PL, Huang YN, Marchese JN, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics* 2004, 3(12):1154-69.
- [93] Ong SE, Blagoev B, Kratchmarova I, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*. 2002, 1(5):376-86.
- [94] DeSouza LV, Taylor AM, Li W, et al. Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in

- cancerous and normal endometrial tissues. *J proteome Res.* 2008, 7(8):3525-34.
- [95] Lottspeich F, Kellermann J. *Gel-Free Proteomics.* Springer. 2011, 55-64.
- [96] Hsu JL, Huang SY, Chow NH, et al. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal Chem.* 2003, 75(24): 6843-52.
- [97] Neilson KA, Ali NA, Muralidharan S, et al. Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics.* 2011, 11(4):535-53.
- [98] Zhang B, VerBerkmoes NC, Langston MA, et al. Detecting differential and correlated protein expression in label-free shotgun proteomics. *Journal of proteome research.* 2006, 5(11):2909-18.
- [99] Coghlin C, Murray GI. Biomarkers of colorectal cancer: recent advances and future challenges. *Proteomics Clin Appl.* 2015, 9(1-2): 64-71.

## 攻读学位期间的研究成果

### 已发表论文:

- [1] Hu L, Zhou Y, Huang X, et al. Association between subjective sleep duration on workdays or non-workdays and uncontrolled blood pressure in Southern China. *J Am Soc Hypertens*. 2018, 12(10): 742~50.
- [2] Zhou Y, Hu L, Huang C, et al. Infection caused by amniotic fluid embolism complicated with disseminated intravascular coagulation. *Clin Exp Obstet Gynecol*. 2017, 11(2): 966~69.
- [3] Zhou Y, Yuan Y, Huang C, et al. Pathogenesis, diagnosis and treatment of cerebral fat embolism. *Chin J Traumatol*. 2015,18(2): 120~23.
- [4] 周仪华, 苑会晶, 袁影等. 艾司洛尔对脓毒症休克患者左室功能及预后的影响. *中华急诊医学杂志*. 2019,28(1):100-03.

## 综述

### 基于蛋白组学的结直肠癌早期检测和个体化用药生物标志物的发现

周仪华 综述 程晓曙 审校

#### 摘要

全世界范围内大约每年有 100 万人患结直肠癌 (CRC)，其中一半患者死于 CRC。早期诊断是 CRC 预后的一个重要的预测因子。有界 CRC 的 5 年生存率接近 90%，转移 CRC 的 5 年生存率接近 10%。因此，在防治 CRC 的战役中其中一个重要的挑战就是提高早期诊断率。大量的蛋白组学方法被用于癌症的新的生物标志物的发现。这种标志物的发现能够提高诊断的准确率并优化个体化治疗。本文综述了近年来发表的有关 CRC 的研究，这些研究通过蛋白质组学方法识别出适合临床应用的蛋白质生物标志物。根据生物标志物的应用，我们讨论了这些发现，包括蛋白质磷酸化和癌症干细胞在生物标志物发现中的作用。我们的综述提供了科学方法的一个截面，可以为未来的实验策略提供建议，供对蛋白质组学感兴趣的科学家、临床医生和研究人员为生物标志物的发现提供参考。

#### Abstract

About one million people/year develop colorectal cancer (CRC) and approximately half of them die. The extent of the disease (i.e. local invasion at the time of diagnosis) is a key prognostic factor. The 5-year survival rate is almost 90% in the case of delimited CRC and 10% in the case of metastasized CRC. Hence, one of the great challenges in the battle against CRC is to improve early diagnosis strategies. Large-scale proteomic approaches are widely used in cancer research to search for novel biomarkers. Such biomarkers can help in improving the accuracy of the diagnosis and in the optimization of personalized therapy. Herein, we provide an overview of studies published in the last 5 years on CRC that led to the identification of protein biomarkers suitable for clinical application by using proteomic approaches. We discussed these findings according to biomarker application, including also the

role of protein phosphorylation and cancer stem cells in biomarker discovery. Our review provides a cross section of scientific approaches and can furnish suggestions for future experimental strategies to be used as reference by scientists, clinicians and researchers interested in proteomics for biomarker discovery.

## CRC 的蛋白组学生物标记物研究概述

结直肠癌（CRC）是全世界范围内癌症相关死亡率的主要来源之一。大约 50% 的治疗病例在首次诊断后 5 年内复发并导致死亡<sup>[1]</sup>。早期发现 CRC 为预防 CRC 死亡提供了最好的机会。尽管目前广泛使用的大肠癌筛查技术、乙状结肠镜检查检查和结肠镜检查等可以减少 CRC 患者的死亡率，但是接受检查的患者数量仍然有限，部分是因为侵入性检查、检查的高额费用以及对检查不适感的恐惧。因此，寻找新的生物标志物作为非侵入性分子检测的靶点，有助于将 CRC 筛查引入临床分析的常规程序。蛋白质生物标志物非常适合基于血液和粪便分析的新型体外分子检测的发展。目前，癌胚抗原(CEA)<sup>[2]</sup>和粪血红蛋白(f-Hb)是唯一被批准用于 CRC 临床的可溶性蛋白生物标志物。但 CEA 也可能在健康、重度吸烟者对炎症的反应、I 型和 II 型糖尿病、溃疡性结肠炎、胰腺炎和肝硬化中过度表达<sup>[3]</sup>。CRC 最常用的筛查方式是愈创木脂化学法粪便潜血检测(gFOBT)。然而，这项检测并不能检测出大多数息肉和癌症，需要多种粪便样本才能很好地解释结果。而且，gFOBT 容易出现假阳性结果，如摄入某些食物和消炎药。

癌症是由于基因突变的积累导致细胞过程的改变，如血管生成、增殖、凋亡和衰老<sup>[4,5]</sup>。目前正在进行广泛的研究，以寻找有用的生物标志物，改善 CRC 筛查的诊断手段，并可用于预测治疗结果。发现非侵入性早期检测方法的新工具是一个优先事项。最初使用基因组学和转录组学方法寻找标记物<sup>[6,7]</sup>，这些方法详细阐述了癌症的遗传基础。通过选择性剪接 mRNA 并结合大量翻译后修饰，一个基因可以编码多种蛋白<sup>[8]</sup>。蛋白质组比基因组更有动态性；因此，它更准确地反映了细胞的机制。通过蛋白质组学，可以分析成千上万的癌细胞蛋白，从而为 CRC 产生新的治疗靶点和标志物。此外，蛋白质标记物可以很容易地在日常可用的体液(如血清)上测量，从而消除了新鲜或冷冻组织活检的需要。在这种背景下，蛋白质组学是一种理想的、高度可翻译的研究工具，可用于发现新的癌症生物标志物。

## 生物标志物发现的蛋白质组学方法

二维聚丙烯酰胺凝胶电泳(2D-PAGE)是根据蛋白质的等电点和分子量分离蛋白质的<sup>[9]</sup>。虽然这是一种成熟的蛋白质分析技术,但这种技术与变异有关,会扭曲蛋白质表达的真正差异。这些蛋白表达上的差异可能是以下几种不一致的结果: 1)由于样品处理、染色程序或图像采集引起的分析变异; 2)样品生产、加工、保存环境造成的生物变异。为了尽量减少这些变异,可以使用多种生物和重复分析;然而,这大大增加了分析的难度。在1997年,二维凝胶电泳(2D-DIGE)的发展克服了这些限制,允许在一个二维凝胶中同时分离三个样品。这保证了准确的凝胶内匹配,解决了许多重现性和定量的问题<sup>[10]</sup>。尽管在检测疏水性、碱性或小蛋白<sup>[11]</sup>方面存在局限性,但基于凝胶的方法已被广泛用于寻找潜在的CRC生物标志物<sup>[12,13-23]</sup>。

近年来,无凝胶法被广泛应用于定量蛋白质组学。复杂的肽混合物,通常由蛋白质酶解产生,使用不同的分馏策略(如反相液相色谱)来解决。为了定量感兴趣的蛋白质/多肽,同位素标记染料经常被使用。典型的同位素标记方法包括通过化学反应,使用同位素编码亲和标记(ICAT)<sup>[24,25]</sup>向蛋白质添加富含同位素的标记,使用等压标记进行相对和绝对量化(iTRAQ)<sup>[26]</sup>,或使用细胞培养中的氨基酸进行稳定同位素标记(SILAC)<sup>[27]</sup>。化学标记的其他变体包括非异压胺标记(mTRAQ)、二甲基标记和同位素编码蛋白标记(ICPL)<sup>[28-30]</sup>。ICAT是Aebersell集团最早引入定量蛋白质组学的试剂之一<sup>[24]</sup>。这些试剂有两种同位素形式,它们的<sup>12</sup>C/<sup>13</sup>C原子含量不同,因此可以在MS模式下对两个样品进行相对定量估计。ICAT标记通常用于比较成对的蛋白质样本。ICAT试剂含有巯基反应性碘乙酸盐基团,其结果是能够标记半胱氨酸(Cys)残基。因此,只有含Cys的蛋白才能通过这种方法进行分析。iTRAQ试剂包含一个由平衡和报告基团组成的等压标签。iTRAQ标签的总质量通过在平衡和报告基团中分别使用<sup>13</sup>C、<sup>15</sup>N和<sup>18</sup>O原子的不同同位素富集来保持恒定,从而使标记的肽具有等压性<sup>[31]</sup>。iTRAQ支持一次最多8个样本的多路复用。这些试剂在一级氨基上具有活性,因此可以用这种方法分析所有的蛋白质。此外,iTRAQ标记允许在MS级的碎片化过程中进行量化。由Mann集团<sup>[27]</sup>开发的SILAC标记不是一种化学物质,而是一种代谢标记过程,它是基于细胞生长在同位素丰富的氨基酸培养基中,使“重”和“轻”氨基酸被纳入细胞的蛋白质。在一个典型的实验中,细胞培养基中加入重精氨酸或赖氨酸(<sup>2</sup>H, <sup>13</sup>C或<sup>15</sup>N),而对照细胞培养基由轻氨基酸组成。这允许同位素标签的早期加入,从而减少了操作的需要,从而减少了过程中样品的变

化。另一方面, SILAC 不适合人体组织。这三种方法在量化方面是可靠的, 但是它们需要使用昂贵的分子、特定的软件和一定的熟悉程度来分析结果。而且, 可以同时分析的样本数量受可用标签数量的限制, 并不是所有的策略都适用于所有类型的样本。

新的仪器和先进的生物信息学帮助开启了无标签蛋白质组学的时代<sup>[32]</sup>。无标签量化, 是一个直接的、廉价的而且不涉及同位素标记的策略。有两种测量方法可以获得定量结果: 峰面积(或离子强度)和光谱计数。峰面积法是根据提取的离子色谱的总峰面积来测量分析物数量的一种方法。其原理是检测到的离子信号与分析物浓度成正比。这种无标签方法的实际局限性与 LC-MS 方法有关, LC-MS 方法必须具有高度的可重复性。任何在保持时间和  $m/z$  上的漂移都将极大地复杂化峰对齐过程。其次, 使用的 MS 仪器必须是高分辨率的, 因为低分辨率仪器在区分重叠峰时是可变的。另一种基于光谱计数的定量测量方法, 是一种简单直接的测量策略。光谱计数定义为分配给一个蛋白质的 MS/MS 谱的数量。因此, 在 LC-MS/MS 分析过程中, 它通过观察一个肽段的 MS/MS 谱的数量来测量属于一个蛋白的每个肽段执行 MS/MS 事件的次数。这与蛋白质的丰度成正比。这种方法的一个缺点是, 由于离子竞争、动态范围限制和 MS 敏感性, 并不是所有的肽都能以相同的方式检测到<sup>[33]</sup>。然而, 无标签质谱已广泛应用于 CRC 生物标志物的发现。

### 发现作为潜在生物标志物的差异表达因子

发现癌症生物标志物最常用和最方便的蛋白质组学策略包括识别“差异”表达的蛋白质和多肽。这里的“差异”一词是指: 1) 肿瘤组织与相应的正常组织相比, 或在不同的肿瘤发生阶段之间, 其水平是不同的; 2) 周围生物环境和液体(如血浆)对肿瘤发生、存在和生长的反应水平不同; 3) 肿瘤或周围液体在特定治疗条件下的表达变化。前两种情况下明显指的是具有潜在诊断或预后效用的因素, 而最后一种情况则与预测标志物的发现有关, 这些标志物可能有助于临床医生在不同的治疗方案之间做出选择。当然, 候选标记物的潜在临床应用在很大程度上取决于发现策略的实验设计, 特别是通过: 1) 生物模型(如细胞培养、动物、患者); 2) 样本类型(例如细胞提取物、组织提取物、生物液体); 3) 蛋白组学技术及样本量(即样本数); 4) 验证的类型和样本量。

#### 1) 诊断标记物

以 CRC 细胞系为生物模型,开展了以发现新的诊断标志物为重点的蛋白质组学研究。利用凝胶电泳和纳米 LC/MS/MS 对 5 种不同 CRC 细胞系的细胞表面蛋白组进行蛋白质组学分析,发现葡萄糖转运体 1 型膜蛋白以及 42 种其他可能的腺瘤-癌进展的生物标志物<sup>[34]</sup>。肿瘤生物标志物发现的一个关键因素是了解上皮细胞与癌细胞之间的相互作用。在这种情况下,分泌的蛋白质在细胞间的交流中起着连接的作用。Zeng 等利用 SILAC 技术评估肿瘤相关分泌蛋白<sup>[35]</sup>。其中 IGFBP6 是 IGF 介导信号的关键调控因子,是 CRC 诊断和预后的潜在生物标志物<sup>[35]</sup>。考虑到组织分泌蛋白的高度相关性,“组织分泌组学”是候选生物标志物的丰富来源。利用无标签光谱计数技术,对 4 例 CRC 患者的 5 种不同 CRC 细胞系分泌物与组织分泌物进行分析,揭示了一组潜在的临床相关的 CRC 早期检测生物标志物<sup>[36]</sup>。2 个 CRC 细胞系(即 HCT-GEO 和 Caco-2)的分泌体谱也存在差异<sup>[37]</sup>。

如果细胞培养在发现分泌标记物方面无疑具有优势,那么动物模型可以代表另一种可能的方法。在这个实验策略中,突变小鼠代表了一种宝贵的资源。腺瘤性息肉病大肠杆菌(Adenomatous polyposis coli, Apc)基因突变常发生在 CRC 中<sup>[38]</sup>。小鼠肠道上皮中 Apc 的缺失是结肠直肠肿瘤发生的良好模型<sup>[38]</sup>。Hammoudi 利用 iTRAQ 对 Apc 缺失的小鼠肠上皮细胞进行了蛋白组学分析,其中热休克 60kDa 蛋白 1、核苷、抗增值蛋白、细胞角蛋白 18、核糖体蛋白 L6、DEAD 盒多肽 5 在人 CRC 中检测,4~6 个被证实为人 CRC 的潜在血清生物标志物<sup>[39]</sup>。

然而,来自患者的生物样本是最有价值的蛋白质组学标记发现的生物系统。根据这一思路,近年来,通过 iTRAQ 定量蛋白组学比较了手术切除的原发性肿瘤及其非肿瘤性结肠黏膜(NNCM)对应物<sup>[40]</sup>。本研究的结果表明,凝乳酶是结直肠癌的生物标志物。事实上,在所有分析的肿瘤中均发现凝乳酶表达降低,并在中国人群中作为 CRC 的生物标志物<sup>[40]</sup>。此外,我们还结合两种不同的定量蛋白组学技术(即 mTRAQ 和 cICAT)分析了 CRC 和非肿瘤组织的蛋白组学特征<sup>[41]</sup>。CRC 组织中的差异表达蛋白是参与相关 CRC 信号通路,即 Wnt/ $\beta$ -连环蛋白和连环蛋白介导的内吞作用<sup>[41]</sup>。Fan 等<sup>[42]</sup>采用无标签定量分析的方法,将三种 CRC 组织与正常黏膜组织进行对比分析。在 67 个差异表达蛋白中,由于原肌凝蛋白-3、内质网驻留蛋白 29、阳离子抗菌蛋白和热休克蛋白 70 kDa 蛋白 8 在 CRC 组织中较正常表达蛋白有较强的上调作用,他们将其确定为 CRC 的潜在诊断生物标志物<sup>[42]</sup>。Sugihara 等使用 2D-DIGE 对 59 例 CRC 患者正常结肠直肠上皮组

织与肿瘤组织进行比较<sup>[43]</sup>。他们鉴定了 110 种蛋白质，这些蛋白质在折叠变化上有统计学上的显著差异，其中他们重点研究了腺瘤息肉结肠炎蛋白 EB1。IHC 验证了该蛋白在肿瘤细胞中的过表达，结果也表明其与不良预后的相关性，从而提出 EB1 作为 CRC 的候选生物标志物和治疗靶点<sup>[43]</sup>。

CRC 组织及其正常对照物(NNCM)是比较蛋白组学分析最明显的样本。然而，如上所述，肿瘤微环境的分子变化可能为诊断提供新的策略。在这一过程中，基质细胞被认为是发现 CRC 早期诊断甚至预防的可能靶点的有价值的起始样本。为了发现与癌变过程有关的基质蛋白，Mu 等人从结肠腺癌黏膜组织中分离基质细胞，并将其蛋白含量与健康结肠黏膜组织进行比较。他们的研究基于 iTRAQ 定量蛋白组学，确定间质差异表达蛋白(即 DCN、FN1、PKM2、HSP90B1、S100A9、MYH9、TUBB、YWHAZ)为结肠癌预防、诊断和治疗的潜在靶点。

在生物标志物中，体液标志物无疑是最有趣的，因为它们的潜在易用性。上面我们提到了在小鼠模型中发现血清标记物的研究;然而，患者的血浆是最有价值的生物样本来阐明体液标志物。在这方面，最近 Choi 等人为了了解肿瘤发生发展的分子机制，对腺瘤性息肉患者和 CRC 患者血浆蛋白组进行了蛋白组学比较分析<sup>[12]</sup>。通过联合使用 2DE 和 MS，他们发现了腺瘤向癌进展的潜在生物标志物。这些潜在生物标志物包括富含 $\alpha$ -2-糖蛋白,血红蛋白亚基 $\beta$ ,Ig  $\alpha$ -2 链 C, 补体因子 B, 生育酚结合蛋白, 锌- $\alpha$ -2-糖蛋白玻璃粘连蛋白, $\alpha$ 1-抗凝乳蛋白酶以及三个细胞因子(即白介素-8,射线诱发干扰素蛋白和肿瘤坏死因子 $\alpha$ )。这些都被认为是 CRC 和/或其从腺瘤进展的新的生物标志物<sup>[12]</sup>。

如前所述，其他非组织临床材料也可用于寻找生物标志物。例如，对腺癌或癌切除术后从结肠直肠道获得的洗液进行了基于凝胶的 2D 研究，并将其与由健康切片获得的洗液组成的对照组进行比较<sup>[44]</sup>。在这些差异表达的蛋白中，有 19 种已经被认为是潜在的 CRC 生物标志物。作者证明，洗液中含有一系列来源于肿瘤上皮细胞分泌的蛋白质，可能与组织的病理进展有关。因此，洗液可以作为生物标志物和药物靶点发现的样品<sup>[44]</sup>。

诊断改进的最佳策略可能应该是识别在临床实践中已经使用且以非侵入性方式收集的样本类型中可检测到的新的特异性生物标志物。在最近的一项工作中，Bosch 试图提高目前粪便血红蛋白检测的诊断准确性，通过识别和验证粪便中的蛋白质生物标志物作为补充检测，他们分析了 10 名健康人和 12 名 CRC 患者粪便中的蛋白质，并利用光谱计数量化了蛋白质表达的差异，他们的初步结果

还有待进一步验证，在 CRC 患者粪便样本中发现近 80 种蛋白表达明显增高，说明该策略的新颖和有效性<sup>[45]</sup>。

## 预后标志物

细胞系常被用作识别预后标志物的生物学模型，特别是对转移过程的特异性标志物。事实上，主要的方法是利用商业 CRC 细胞系与它们的同基因转移物配对。Luque-Garcia 等<sup>[46]</sup>使用 SILAC 比较了两种 CRC 细胞系 KM12C 和 KM12SM，这两种细胞系分别代表低转移细胞和高转移细胞。他们研究了两种细胞系在细胞表面的差异表达蛋白，发现整合素和其他细胞粘附分子(如连接斑珠蛋白)与高转移细胞系相关<sup>[46]</sup>。Barderas 等<sup>[47]</sup>也利用 SILAC 标记分析了相同的两种细胞系，试图鉴定与结肠癌转移相关的蛋白。他们在全蛋白组学上的研究结果证实了转移细胞系中存在下调的蛋白质，以及一组参与细胞粘附的蛋白质(如 NEO1、SERPINI1 和 PODXL)。此外，沉默其表达可降低转移细胞的粘附性<sup>[47]</sup>。Lin 和他的同事<sup>[48]</sup>报道了另一种基于 iTRAQ 标记结合 2D-LC 的方法，来阐明可能作为转移生物标志物的蛋白。在这种情况下，将结肠腺癌 HCT-116 细胞与转移癌 E1 细胞进行了比较，与 HCT-116 相比，Drebrin (DBN1)在 E1 中过表达。该结果经 western blotting 和 CRC 切片免疫组化验证。Ghosh 等<sup>[49]</sup>使用 iTRAQ，比较了两种同基因 CRC 细胞系之间的蛋白质组谱，例如 SW480 和转移性对应的 SW620。他们的结果表明钙周期蛋白结合蛋白(CacyBP)参与了 CRC 转移过程。几年后，同样的研究小组使用另一种方法寻找与转移相关的生物标志物，目的是为了重现肿瘤的异质性<sup>[50]</sup>。他们再次进行了基于 iTRAQ 的定量比较，这次比较的是 CRC 细胞系(HCT-116)的蛋白组和非同源转移细胞系 SW620。他们的结果证实了 CacyBP 水平与 CRC 转移之间存在直接相关性，因此 CacyBP 可能作为 CRC 转移的候选生物标志物<sup>[50]</sup>。SW480 和 SW620 等基因细胞系是 CRC 和淋巴结转移的体外模型，已用于多项研究。在这些研究中，q 亚组分结合蛋白<sup>[51]</sup>，分泌的蛋白质三叶因子 3<sup>[52]</sup>，生长/分化因子 15， $\beta$ -连环蛋白以及与 RhoA 相关的 GTP-结合蛋白<sup>[53]</sup> 与 SW620 细胞系的转移潜能相关。因此，这些蛋白被认为是预测 CRC 转移的生物标志物。相反，Lee 等<sup>[54]</sup>通过比较相同的两种同基因 CRC 细胞，最近提出了脂肪酸合成酶和组蛋白 H4 作为癌症转移的分子靶点。这些基因的沉默表明迁移减少，从而证实了它们在癌症转移中的作用<sup>[54]</sup>。CRC 和转移细胞体外模型的另一个例子是 HCT-116 和 E1 细胞系。Lin 等<sup>[55]</sup>采用多凝集素亲和层析

法富集分泌体(糖分泌体)中的糖基化蛋白。然后,他们进行了一项无标签的分析,以比较两种糖醇分泌体,发现在 E1 细胞中层流蛋白-1 分泌过量。ELISA 检测 CRC 患者血清中层流蛋白-1 水平。他们的结果显示,在区分 CRC 患者和健康对照组时,层流蛋白-1 相对于 CEA 血清水平有更好的表现<sup>[55]</sup>。

小鼠肿瘤异种移植是鉴别转移特异性蛋白因子的一种有价值的生物学模型。按照这种策略,2D-DIGE(差异凝胶电泳)被用来识别原发性结肠癌肝转移相关的分子靶点<sup>[56]</sup>。Leiphakpam 及其同事开发了一种原位小鼠模型,将高度转移的 CRC 异种移植瘤植入裸鼠结肠,诱导肝转移形成<sup>[56]</sup>。比较原发性结肠肿瘤和肝转移裂解液的全蛋白表达谱。在差异表达的蛋白中,埃兹蛋白(ezrin)在肝脏转移中被发现表达上调,更准确地说,ezrin 的磷酸化亚型(T567)被认为是一种靶向 CRC 转移的潜在治疗手段<sup>[56]</sup>。在同一研究中,与原发性 CRC 肿瘤相比,人类肝脏转移的免疫组化(IHC)也证实了 p-ezrin T567 表达的增加。

从概念上讲,发现预后标志物最直接的方法是使用患者的样本。在此背景下,我们使用 cICAT 和 2D-DIGE 的结合来定义一个蛋白组学图谱,该图谱可以帮助风险分层并预测 IV 期 CRC 的不良结果<sup>[57]</sup>。此外,O'Dwyer 等<sup>[58]</sup>采用了一种基于 2DE 的方法来分析 28 个 CRC 切片与相应的 NNCM 配对。他们在包含 515 例原发性 CRC、224 例淋巴结转移和 50 NNCM 样本的组织芯片上发现 54 种候选蛋白和 15 种蛋白经 IHC 验证。生物信息学和统计学分析允许将 14-3-3b 和醛脱氢酶 1 作为两种具有高度预后意义的蛋白标记<sup>[59]</sup>。

### 3) 预测标记物

细胞系是最常用的模型,用于阐明假定的生物标志物,这些标志物可用于预测特定治疗的反应。采用这种方法,利用凝胶蛋白组学技术比较了西妥昔单抗敏感和耐药 GEO 细胞的差异表达谱<sup>[20]</sup>。类似的方法和相同的技术也揭示了潜在的生物标志物,可以预测化疗对广泛使用的药物如 5-FU、伊立替康和奥沙利铂/顺铂的反应性<sup>[60-62]</sup>。

蛋白质组学策略也倾向于研究非规范治疗的反应。在这些情况下,基于蛋白质组学的方法的优点是,它们不仅可以识别新的潜在的生物标志物,而且还可以帮助理解与正在研究的情况相关的分子机制。Tan 等<sup>[63]</sup>以 iTRAQ 和可裂解 ICAT 为例,研究了丁酸盐处理 HCT-116 结肠癌细胞。这使得作者能够识别出由丁酸调节的通路组成的信号簇,作为 CRC 治疗的新型化学预防和治疗药物的潜在靶点<sup>[63,64]</sup>。在非传统疗法的背景下,有强有力的证据表明饮食对癌症的发病率

有影响。特别是全麦食品的假设似乎与 CRC 发病风险的降低有关<sup>[65]</sup>，可以适当地使用小鼠模型来研究这种预防策略的有效性。事实上，为了寻找新的 CRC 的化学预防生物标志物，Norris 等<sup>[65]</sup>在 CRC 的 Apcmin 小鼠模型中使用米糠饮食。这些老鼠的日常饮食中补充了不同数量的大米。采集血浆进行 MS-蛋白组学分析，结果发现脂联素是一种潜在的抑癌因子，其表达受饮食中大米的影响较大<sup>[65]</sup>。

### 2) 磷酸化作为结直肠癌的生物标志物

蛋白磷酸化和去磷酸化是翻译后修饰中最重要和最丰富的修饰之一，是一个具有生物学意义的生化过程。事实上，它作为一个分子开关，“开启”或“关闭”蛋白质活性或细胞通路，从而调节酶活性和底物特异性、蛋白质的折叠和功能、蛋白质的定位以及复合物的形成/降解。近三分之一的真核蛋白被磷酸化<sup>[66,67]</sup>。磷酸化的检测，从位点和丰度的定义来看，是亚化学计量丰度蛋白质组学的一个具有挑战性的任务。几种富集策略已被开发出来以分离和鉴定磷酸化蛋白；最常用的富集方法是固定金属亲和层析(IMAC)、金属氧化物亲和层析(MOAC)和强阳离子交换层析(SCX)<sup>[68]</sup>。鉴于激酶和磷酸酶作为信号级联的调节因子的关键作用，对结肠癌磷酸化蛋白组的分析可能揭示磷酸化与癌症机制之间的联系。Smad4 阳性 SW480 细胞 TGF- $\beta$ 驱动磷酸化蛋白质组的研究显示，细胞周期蛋白依赖性激酶家族成员在 CRC 细胞中被下游 TGF- $\beta$ 激活<sup>[69]</sup>。对人类结肠癌细胞株 HT-29 进行的一项全磷酸化蛋白组分析显示，在与细胞凋亡相关的蛋白以及其他参与细胞信号转导调控的蛋白(包括激酶及其底物)上存在磷酸化位点<sup>[70]</sup>。这些位点可以作为替代生物标志物，验证来自不同细胞条件和外源性刺激的激酶和信号通路的活性。两项不同的研究评估了酪氨酸磷酸化在结肠直肠癌中的作用<sup>[71,72]</sup>。此外，Sakai 等<sup>[60]</sup>在试图鉴定 5-氟尿嘧啶(5-FU)耐药相关蛋白时，比较人 CRC 细胞系 DLD-1 和 DLD-1/5-FU 的总蛋白组和磷酸化蛋白组，即 5-FU 抗性亚克隆，提示磷酸化 HSPB1 在 5-FU 抗性中发挥作用。Piersma 等<sup>[73]</sup>对 8 个 CRC 细胞系进行了无标签磷酸肽分析，这些不同的细胞系代表了 3 种 CRC 预后亚型。通过无标记磷酸化蛋白质组学，作者获得了生物细胞系的定量图谱，并识别出与不良预后相关的信号蛋白和过程，强调了该技术在肿瘤活检分析临床应用中的可行性<sup>[73]</sup>。

### 3) 癌症干细胞在结直肠癌生物标志物发现中的作用

根据癌症干细胞(cancer stem cells, CSC)假说<sup>[74-76]</sup>，肿瘤是由分层组织的异质性细胞组成的。其中，一小部分亚群具有触发肿瘤发生的能力(即 CSCs 或肿瘤

起始细胞(TICs))<sup>[77-79]</sup>。这些细胞诱导细胞增殖、侵袭和转移。它们也具有自我更新和多向性分化潜能<sup>[80]</sup>。虽然 CSCs 的起源尚待确定,但其作为肿瘤维持/生长“来源”的作用已被广泛接受。因此,消灭 CSCs 似乎是癌症治愈的重要前提。此外,了解它们的特征和分子机制可以为恶性肿瘤新靶向治疗的开发提供新思路。对于 CRC 来说,这显然是正确的<sup>[81-83]</sup>。因此,CSCs 是开发更有效的筛查、早期发现、预防、治疗和护理方法的合理目标<sup>[84]</sup>。在此基础上,对 CRC 来源的 CSCs 进行识别、严格筛选和表征,对于改进现有的筛查、早期诊断、风险分层预后和治疗结果个性化预测方法至关重要<sup>[85]</sup>。一些膜和细胞质标志物被用来分离假定的 CRC 干细胞(即 CD133, CD24, CD29, CD44, CD166 (ALCAM), EpCAM, Lgr5, ALDH1A1 和 ALDH1B1)。我们使用流式细胞术对结肠 CSCc 进行表型表征,发现 CD66c 表达与 CSCs 特异性相关,表明 CD66c 沉默几乎完全与 CaCo-2 细胞的致瘤性相反<sup>[86]</sup>。我们还根据 CSC 标志物 CD133 的表达,对从两个 CRC 细胞系分离的 CSCs 进行 2D-DIGE 蛋白组分析<sup>[58]</sup>。我们表明,剪接因子 SRp20,Wnt / $\beta$ -连环蛋白通路的目标基因,影响细胞增殖,当基因沉默时,导致 MCC-介导的减速 Wnt 通路<sup>[87]</sup>。我们的结果表明,CD66c 和 SRp20 是 CSCs 靶向治疗的生物标志物。Lee 等<sup>[61]</sup>利用基于 CD133 的 CSC 分选和 2-DE 的研究发现,激动蛋白相关蛋白 2/3 复合物亚基 5 和抑制蛋白 2 可能是特定的 CSC 蛋白,有助于这些细胞的干细胞性和致瘤性<sup>[61]</sup>。在一项基于质谱的蛋白质组学研究中,Fang 等<sup>[88]</sup>证实了结肠 CSCs 相关的几种新的细胞表面蛋白(即 CEACAM5、biglycan 和 cadherin 17)的过表达。他们还证明结肠 CSCs 可以在无血清干细胞培养条件下分离和繁殖。Zou 等<sup>[89]</sup>从 SW1116 中分离出人结肠 CSCs,用 2-DE 对其特异性蛋白组进行了表征。这揭示了 10 种与 CSCs 相关的蛋白,包括泛素融合降解 1L(UFD1L)、人分层蛋白(14-3-3 家族成员)和甘油醛-3-磷酸脱氢酶 (GAPDH)<sup>[89]</sup>。最后, Van Houdt 等<sup>[62]</sup>采用无标签量化的方法,识别出细胞凋亡蛋白抑制剂 BIRC6 作为肿瘤干细胞对顺铂和奥沙利铂耐药的重要介质,提示 BIRC6 靶向性是一种潜在的根除结肠 CSCs 的方法<sup>[62]</sup>。

#### 4) 结论

蛋白质组学的最新进展为癌症相关生物标志物的发现提供了新的见解。新的蛋白质组学技术以及改进的数据处理和解释工具为寻找能够预测癌症治疗和管理结果的生物标志物提供了新的手段。蛋白质组学研究中发现的大量 CRC 相关蛋白表明,这些蛋白可能是 CRC 新的候选生物标志物。对这些和其他假定的

蛋白质生物标志物的进一步研究将产生一组在临床环境中对 CRC 具有足够敏感性和特异性的标志物。我们相信在未来几年还会发现其他 CRC 相关蛋白和潜在的生物标志物。重要的是，CRC 的异质性要求有多个标记，每个标记代表一种不同的 CRC 亚型。一个单一的生物标记很难既非常敏感又非常具体。因此，蛋白质组学的目标应该是识别特定的蛋白质特征，以便对患者进行最佳的分层和分类，从而简化临床管理。事实上，在使用不同实验策略进行的不同研究中，候选生物标志物的列表之间通常不存在或几乎不存在重叠。不同的蛋白质组学方法应该被认为是互补的，而不是替代的，多个蛋白质组学技术的并行应用导致候选蛋白质的重叠列表将有更多的机会提出有用的标记。值得注意的是，蛋白质组学研究已经提出了大量新的假定的生物标志物，但在临床实践中这些标志物的引入并没有取得显著的改善。在这方面，一个关键的步骤是标记验证策略，应该使用最标准化的技术，并使用来自统计上显著数量的患者的样本进行验证。最佳做法应是使用妥善回收和储存的临床组织样本的标准化档案，并提供适当的临床文件和有关临床结果的详细说明。此外，基于新型生物标志物的新检测方法的性能应严格与大量(成百上千)患者的现有检测方法(如 FOBT)进行比较。最后，规范的实验室间的检测分析和验证协议，将有助于克服发现策略和结果临床应用之间的瓶颈<sup>[90]</sup>。

对于样本类型，在体液中发现新的 CRC 候选生物标志物是下一个挑战，可以使用新的靶向高复用 MS 方法发现。CSCs 在这种情况下的作用影响很大，必须加以考虑。我们相信更多的 CRC 特异性血清和组织生物标志物可以促进新药的开发。未来可能会有更多的蛋白质组学研究关注这些细胞的生物标志物，从而为特定 CRC CSCs 的靶向治疗铺平道路。

#### 参考文献

- [1] Newton KF, Newman W, Hill J, et al. Review of biomarkers in colorectal cancer. *Colorectal Dis.* 2012, 14(1):3-17.
- [2] Gemei M, Corbo C, Salvatore F, et al. Carcinoembryonic Antigen Family Cell. Adhesion. Molecules (CEACAM) as Colorectal Cancer Biomarkers. *Biomarkers in Cancer.* 2015, 685-705.
- [3] Tanaka T, Tanaka M, Tanaka T et al. Biomarkers for colorectal cancer. *Int J Mol Sci.* 2010,11(9): 3209-25.
- [4] Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell.* 2000, 100(1):57-70.
- [5] Hanahan D., Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011, 144(5):646-74.

- [6] Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet.* 2010, 11(10):685-96.
- [7] Hanash SM, Baik CS, Kallioniemi O. Emerging molecular biomarkers--blood-based strategies to detect and monitor cancer. *Nat rev Clin oncol.* 2011, 8(3):142-50.
- [8] Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu rev biochem.* 2003, 72:291-336.
- [9] O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem.* 1975, 250(10):4007-21.
- [10] Ünlü M, Morgan ME, Minden JS. Difference gel electrophoresis: A single gel method. For detecting changes in protein extracts. *Electrophoresis.* 1997, 18(11):2071-77.
- [11] Issaq HJ, Veenstra TD. The role of electrophoresis in disease biomarker discovery. *Electrophoresis.* 2007, 28(12):1980-88.
- [12] Choi JW, Liu H, Shin DH, et al. Proteomic and cytokine plasma biomarkers for predicting progression from colorectal adenoma to carcinoma in human patients. *Proteomics.* 2013, 13(15):2361-74.
- [13] Friedman DB, Hill S, Keller JW, et al. Proteome analysis of human colon cancer by two - dimensional difference gel electrophoresis and mass spectrometry. *Proteomics.* 2004, 4:793-811.
- [14] Alfonso P, Núñez A, Madoz Gurrpide J, et al. Proteomic expression analysis of colorectal cancer by two - dimensional differential gel electrophoresis. *Proteomics.* 2005, 5(10): 2602-11.
- [15] Kim HJ, Kang HJ, Lee H, et al. Identification of S100A8 and S100A9 as serological markers for colorectal cancer. *J Proteome Res.* 2009, 8(3): 1368-79.
- [16] Álvarez-Chaver P, Rodríguez-Piñeiro AM, Rodríguez-Berrocal FJ et al. Identification of hydrophobic proteins as biomarker candidates for colorectal cancer. *Int J Biochem Cell Biol.* 2007, 39(3): 529-40.
- [17] Xing X, Lai M, Gartner W, et al. Identification of differentially expressed proteins. In colorectal cancer by proteomics: Down regulation of secretagogin. *Proteomics.* 2006, 6(9):2916-23.
- [18] Ma YL, Peng JY, Zhang P, et al. Heterogeneous nuclear ribonucleoprotein A1 is identified as a potential biomarker for colorectal cancer based on differential proteomics technology. *J Proteome Res.* 2009, 8(10):4525-35.
- [19] Rahman MM, Seo YR. Discovery of potential targets of selenomethionine mediated chemoprevention in colorectal carcinoma mouse model using proteomics analysis. *Carcinogenesis.* 2013, 34(7):1575-84.
- [20] Corbo C, Orrù S, Gemei M et al. Protein cross - talk in CD133+ colon cancer cells indicates activation of the Wnt pathway and upregulation of SRp20 that is potentially involved in tumorigenicity. *Proteomics.* 2012, 12(12): 2045-59.
- [21] Monteleone F, Rosa R, Vitale M, et al. Increased anaerobic metabolism. is a distinctive

- signature in a colorectal cancer cellular model of resistance to antiepidermal growth factor receptor antibody. *Proteomics*. 2013, 13(5):866-77.
- [22] Kim KH, Yeo SG, Kim WK et al. Up-regulated expression of l-caldesmon associated with malignancy of colorectal cancer. *BMC cancer*. 2012, 12:601.
- [23] Lee HN, Park SH, Lee EK, et al. Proteomic profiling of tumor- initiating cells in HT-29 human. colorectal cancer cells. *Biochem Biophys Res Commun*. 2012, 427(1):171-77.
- [24] Gygi SP, Rist B, Gerber SA, et al. Quantitative analysis of complex protein. mixtures using. isotope-coded affinity tags. *Nat Biotechnol*. 1999, 17(10):994-9.
- [25] Han DK., Eng J, Zhou H, et al. Quantitative profiling of differentiation-induced. Microsomal. proteins using isotope-coded affinity tags and mass spectrometry. *Nat biotechnol*. 2001,19(10):946-51.
- [26] Ross PL, Huang YN. Marchese JN, et al. Multiplexed protein. quantitation in *Saccharomyces. cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*. 2004, 3(12):1154-69.
- [27] Ong SE, Blagoev B, Kratchmarova I et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics*. 2002, 1(5): 376-86.
- [28] DeSouza LV, Taylor AM, Li W, et al. Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in cancerous and normal endometrial tissues. *J Proteome Res*. 2008, 7(8):3525-34.
- [29] Lottspeich F, Kellermann J. *Gel-Free Proteomics*, Springer 2011, pp, 55-64.
- [30] Hsu JL, Huang SY, Chow NH, et al. Stable-isotope dimethyl labeling for quantitative proteomics. *Analytical chemistry*. 2003, 75(24):6843-52.
- [31] Iliuk A, Galan J, Tao WA. Playing tag with quantitative proteomics. *Anal Bioanal Chem* 2009, 393(2):503-13.
- [32] Neilson KA, Ali NA, Muralidharan S, et al. Less label, more free: approaches in label - free. quantitative mass spectrometry. *Proteomics*. 2011, 11(4):535-53.
- [33] Zhang B, VerBerkmoes NC, Langston MA et al. Detecting differential. and correlated protein expression in label-free shotgun proteomics. *J Proteome Research*. 2006, 5(11):2909-18.
- [34] de Wit M, Jimenez CR, Carvalho B, et al. Cell surface proteomics identifies. Glucose. transporter type 1 and prion protein as candidate biomarkers for colorectal adenoma-to-carcinoma progression. *Gut*. 2011, 61(6):855-64.
- [35] Zeng X, Yang P, Chen B, et al. Quantitative secretome analysis reveals the. Interactions. between epithelia and tumor cells by in vitro modulating colon cancer microenvironment. *Journal of proteomics*. 2013, 89:51-70.
- [36] de Wit M, Kant H, Piersma SR, et al. Colorectal cancer candidate biomarkers. identified by. tissue secretome proteome profiling. *J Proteomics*. 2014, 99: 26-39.
- [37] Imperlini E, Colavita I, Caterino M, et al. The secretome signature of colon cancer cell

- lines. *Journal of cellular biochemistry*. 2013,114(11):2577-87.
- [38] Smith G, Carey FA, Beattie J, et al. Mutations in APC, Kirsten-ras and p53-alternative. genetic pathways to colorectal cancer. *Proc Natl Acad Sci*. 2002, 99(14):9433-38.
- [39] Hammoudi A, Song F, Reed KR, et al. Proteomic profiling of a mouse model of acute intestinal Apc deletion leads to identification of potential novel biomarkers of human colorectal cancer(CRC). *Biochem Biophys Res Commun*. 2013, 440(3):364-70.
- [40] Fan NJ, Gao CF, Wang CS, et al. Discovery and verification of gelsolin as a potential biomarker of colorectal adenocarcinoma in a Chinese population: Examining differential protein expression using an iTRAQ labelling-based proteomics approach. *Can J Gastroenterol*. 2012, 26(1):41-7.
- [41] Kang UB, Yeom J, Kim HJ, et al. Expression profiling of more than 3500. proteins of MSS-type colorectal cancer by stable isotope labeling and mass spectrometry. *J proteomics*. 2012, 75(10):3050-62.
- [42] Fan NJ, Gao JL, Liu Y, et al. Label-Free Quantitative Mass Spectrometry Reveals a Panel of Differentially Expressed Proteins in Colorectal Cancer. *BioMed Res Int*. 2015, 2015:365068.
- [43] Sugihara Y, Taniguchi H, Kushima R, et al. Proteomic-based identification of the APC-binding protein EB1 as a candidate of novel tissue biomarker and therapeutic target for colorectal cancer. *J Proteomics*. 2012, 75(17), 5342-55.
- [44] Giusti L, Iacconi P, Da Valle Y, et al. A proteomic profile of washing fluid from the colorectal tract to search for potential biomarkers of colon cancer. *Mol BioSyst*. 2012,8(4),1088-99.
- [45] Bosch LJ, de Wit M, Oudgenoeg G, et al. Stool proteomics reveals new candidate biomarkers for colorectal cancer screening. *Cancer Res*. 2012, 72(8), 4523-23.
- [46] Luque - García, JL, Martínez - Torrecuadrada JL, Epifano C, et al. Differential Protein expression on the cell surface of colorectal cancer cells associated to tumor metastasis. *Proteomics*. 2010, 10(5):940-52.
- [47] Barderas R, Mendes M, Torres S, et al. In-depth characterization of the secretome of colorectal cancer metastatic cells identifies key proteins in cell adhesion, migration, and invasion. *Mol Cell Proteomics*. 2013, 12(6):1602-20.
- [48] Lin Q, Tan HT, Lim TK, et al. iTRAQ analysis of colorectal cancer cell lines suggests Drebrin (DBN1) is overexpressed during liver metastasis. *Proteomics* 2014, 14(11):1434-43.
- [49] Ghosh D, Yu H, Tan XF et al. Identification of key players for colorectal. cancer metastasis by iTRAQ quantitative proteomics profiling of isogenic SW480 and SW620 cell lines. *J proteome R*. 2011, 10(10):4373-87.
- [50] Ghosh D, Li Z, Tan XF, Lim TK, et al. iTRAQ Based Quantitative Proteomics. Approach Validated the Role of Calcyclin Binding Protein (CacyBP) in Promoting Colorectal Cancer Metastasis. *Mol Cell Proteomics*. 2013, 12(7):1865-80.

- [51] Arielly, SS, Ariel M, Yehuda, R, et al. Quantitative analysis of caveolin-rich lipid raft proteins. from primary and metastatic colorectal cancer clones. *J proteomics*. 2012, 75(9):2629-37.
- [52] Xue H, Lu B, Zhang J, et al. Identification of serum biomarkers for colorectal. cancer metastasis using a differential secretome approach. *J Proteome Res* 2010, 9(1):545-55.
- [53] Bauer KM, Lambert PA, Hummon AB. Comparative label - free LC - MS/MS analysis of colorectal adenocarcinoma and metastatic cells treated with 5 - fluorouracil. *Proteomics*. 2012, 9(1):1928-37.
- [54] Lee JG, McKinney KQ, Pavlopoulos AJ, et al. Identification of anti-metastatic drug and natural compound targets in isogenic colorectal cancer cells. *J Proteomics* 2015, 113:326-36.
- [55] Lin Q, Lim HS, Lin HL, et al. Analysis of colorectal cancer glyco - secretome identifies laminin $\beta$  - 1 (LAMB1) as a potential serological biomarker for colorectal cancer. *Proteomics*. 2015, 15(22):3905-20.
- [56] Leiphrakpam PD, Rajput A, Mathiesen M, et al. Ezrin expression and cell. survival regulation in colorectal cancer. *Cell Signal*. 2014, 26(5):868-879.
- [57] Kim HJ, Kang UB, Lee H, et al. Profiling of differentially expressed. proteins in stage IV colorectal cancers with good and poor outcomes. *J Proteomics*. 2012, 75(10): 2983-97.
- [58] O'Dwyer D, Ralton LD, O'Shea A, et al. The proteomics of colorectal cancer: identification of a protein signature associated with prognosis. *PLoS One*, 2011,6(11): e27718.
- [59] Monteleone F, Rosa R, Vitale M, et al. Increased anaerobic metabolism. is a distinctive signature in a colorectal cancer cellular model of resistance to anti-epidermal growth factor receptor antibody. *Proteomics*. 2013, 13(5), 866-77.
- [60] Sakai A, Otani M, Miyamoto A, et al. Identification of phosphorylated. serine-15 and -82 residues of HSPB1 in 5-fluorouracil-resistant colorectal cancer cells by proteomics. *J Proteomics*. 2012, 75(3), 806-18.
- [61] Gong FM, Peng XC, Tan BX, et al. Comparative proteomic analysis of. irinotecan-sensitive colorectal carcinoma cell line and its chemoresistant counterpart. *Anticancer drugs*. 2011, 22(6), 500-6.
- [62] Van Houdt WJ, Emmink BL, Pham TV, et al. Comparative proteomics of colon cancer stem. cells and differentiated tumor cells identifies BIRC6 as a potential therapeutic target. *Mol Cell Proteomics* 2011, 10(12):M111. 011353.
- [63] Tan HT, Lim TK, Chung MC, et al. *Drug Design and Discovery*. Springer. 2011, pp. 207- 224.
- [64] Tan HT, Tan S, Lin Q, et al. Quantitative and temporal proteome analysis of butyrate-treated. colorectal cancer cells. *Mol Cell Proteomics*. 2008, 7(6): 1174-85.
- [65] Norris L, Malkar A, Horner - Glister E, et al. Search for novel circulating cancer. chemopreventive biomarkers of dietary rice bran intervention in Apc(Min) mice model of colorectal carcinogenesis, using proteomic and metabolic profiling strategies. *Mol Nutr*

- Food Res. 2015,59(9):1827-36.
- [66] Sefton BM, Shenolikar S. Overview of protein phosphorylation. *Curr Protoc Mol Biol.* 2001, 18(11):11-15.
- [67] Johnson LN. The regulation of protein phosphorylation. *Biochem Soc Trans.* 2009, 37(4): 627-41.
- [68] Schmidt SR, Schweikart F, Andersson ME. Current methods for phosphoprotein isolation and enrichment. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2007, 849(1-2):154-162.
- [69] Ali NA, Molloy MP. Quantitative phosphoproteomics of transforming growth factor -  $\beta$  signaling in colon cancer cells. *Proteomics.* 2011, 11(16):3390-401.
- [70] Kim JE, Tannenbaum SR, White FM. Global phosphoproteome of HT-29 human colon adenocarcinoma cells. *J Proteome Res.* 2005, 4(4): 1339-46.
- [71] Leroy C, Fialin C, Sirvent A, et al. Quantitative phosphoproteomics reveals a cluster of tyrosine kinases that mediates SRC invasive activity in advanced colon carcinoma. *Cells Cancer Res.* 2009, 69(6), 2279-86.
- [72] Organ SL, Tong J, Taylor P, et al. Quantitative phosphoproteomics profiling of hepatocyte growth factor (HGF)-MET signaling in colorectal cancer. *J Proteome Res.* 2011, 10(7):3200-11.
- [73] Piersma SR, Knol JC, de Reus I, et al. Feasibility of label-free phosphoproteomics and application to base-line signaling of colorectal cancer cell lines. *J Proteomics.* 2015, 127(PtB):247-58.
- [74] Guo W, Lasky JL, Wu H. Cancer stem cells. *Pediatr res.* 2006, 59(4), 59R-64R.
- [75] Reya T, Morrison SJ, Clarke MF, et al. Stem cells, cancer, and cancer stem cells. *Nature* 2001, 414(6895):105-11.
- [76] Tan BT, Park CY, Ailles LE, et al. The cancer stem cell hypothesis: a work in progress. *Lab invest.* 2006, 86(12):1203-7.
- [77] Clarke MF, Dick JE, Dirks PB, et al. Cancer stem cells perspectives on current status and future directions: AACR Workshop on cancer stem cells. *Cancer Res.* 2006, 66(19):9339-44.
- [78] Dalerba P, Cho RW, Clarke MF. Cancer stem cells: models and concepts. *Annu Rev Med.* 2007, 58: 267-84.
- [79] Shibata M, Shen MM. The roots of cancer: stem cells and the basis for tumor heterogeneity. *Bioessays.* 2013, 35(3):253-60.
- [80] Yu Z, Pestell TG, Lisanti MP, et al. Cancer stem cells. *Int J Biochem Biol.* 2012, 44(12), 2144- 51.
- [81] Papailiou J, Bramis KJ, Gazouli M, et al. Stem cells in colon cancer. A new era in cancer theory begins. *Int J Colorectal Dis.* 2011, 26(1), 1-11.
- [82] Vaiopoulos AG, Kostakis ID, Koutsilieris M, et al. Colorectal cancer stem cells. *Stem cells* 2012, 30(3), 363-71.
- [83] Ricci-Vitiani L, Fabrizio E, Palio E, et al. Colon cancer stem cells. *J Mol Med.* 2009, 87(11),

1097-104.

- [84] Subramaniam D, Ramalingam S, Houchen C, et al. Cancer stem cells: a novel paradigm for cancer prevention and treatment. *Mini Rev Med Chem*. 2010, 10(5):359-71.
- [85] Langan RC, Mullinax JE, Raiji MT, et al. Colorectal cancer biomarkers and the potential role of cancer stem cells. *J Cancer*. 2013, 4(3), 241-50.
- [86] Gemei M, Mirabelli P, Di Noto R, et al. CD66c is a novel marker for colorectal cancer stem cell isolation and its silencing halts tumor growth in vivo. *Cancer*. 2013,119(4):729-38.
- [87] Corbo C, Orrù S, Salvatore F. SRp20: an overview of its role in human diseases. *Biochem Biophys Res Commun*. 2013, 436(1):1-5.
- [88] Fang DD, Kim YJ, Lee CN, et al. Expansion of CD133(+) colon cancer cultures retaining stem properties to enable cancer stem cell target discovery. *Br j Cancer*. 2010, 102(8):1265-75.
- [89] Zou J, Yu XF, Bao ZJ, et al. Proteome of human colon cancer stem cells: a comparative analysis. *World J Gastroenterol*. 2011, 17(10):1276-85.
- [90] Coghlin C, Murray GI. Biomarkers of colorectal cancer: recent advances and future challenges. *Proteomics Clin Appl*. 2015, 9(1-2):64-71.