

摘要

三阴性乳腺癌 (Triple-negative breast cancer, TNBC) 是一种常见的乳腺癌病理亚型, 患者对一般的激素治疗和靶向治疗不敏感且预后较差。目前临床仍然缺乏可靠的诊断三阴性乳腺癌及其不同预后的生物标志物。研究显示, 肿瘤的发生发展与蛋白质的糖基化修饰密切相关。然而, 截至目前有关三阴性乳腺癌的异常糖基化修饰研究较少。本研究联合高通量蛋白质组学和位点特异性糖蛋白质组学的方法, 定量检测三阴性乳腺癌患者的蛋白质及糖基化修饰水平, 深入研究癌症不同预后情况的蛋白及糖基化蛋白的分子特征及功能通路, 为三阴性乳腺癌患者提供更丰富的诊断及治疗参考。主要研究内容结果如下:

(1) 本研究分别收集了 9 例预后良好及 9 例预后不良的三阴性乳腺癌患者癌症组织和配对的非癌性邻近组织 (Non-cancerous adjacent tissues, NATs), 使用串联质谱标签 (Tandem mass tag, TMT) 定量标记酶解肽段, 完整糖肽通过亲水柱富集。

(2) 通过对差异蛋白进行功能与 KEGG 通路富集分析, 发现癌症组织中表达量高的蛋白主要涉及细胞黏附、细胞迁移及改变细胞外基质等生物通路, 促进了肿瘤细胞的侵袭。对癌症不同预后情况进行分析, 发现组蛋白 H1.2 在预后不良癌症组织中相对丰度最高, 促进了癌细胞的生长发育。

(3) 对患癌症组织和非癌性邻近组织的差异糖基化修饰进行功能与 KEGG 通路富集分析, 发现糖蛋白血液结合素和抗凝血酶 III 的糖基化修饰水平在癌症组织中较高, 推测其促进了癌症的转移和侵袭。对癌症不同预后的糖基化修饰情况进行分析发现, 仅在预后良好患者中表达量高的 25 个糖蛋白与细胞外基质-受体通路和人乳头状瘤病毒感染等通路存在关联, 这些信号通路与癌细胞分化以及转移关系紧密, 其中 PI3K/AKT 信号通路为三阴性乳腺癌常见失调信号通路之一。

(4) 根据三阴性乳腺癌患者组织的差异蛋白糖基化修饰情况, 发现糖蛋白 CO6A6、DCN 和 COCA1 分别为预后良好患者、预后不良患者和三阴性乳腺癌患者的潜在预测生物标志物。值得注意的是, 我们发现部分糖蛋白的同一糖基化位点上的糖基化修饰在不同预后三阴性乳腺癌组织的修饰水平存在差异。

本研究从糖基化修饰水平解释了三阴性乳腺癌及其不同预后情况下的糖基化修饰的分子特征和相关生物功能, 为寻找三阴性乳腺癌的生物标志物提供有效的支持。

关键词: 三阴性乳腺癌; 肿瘤预后; 蛋白质组学; 糖蛋白质组学; 生物标志物

Abstract

Triple-negative breast cancer (TNBC) is a pathological subtype of breast cancer, which is insensitive to common hormone therapy and targeted therapy and has a poor prognosis. At present, there is still a lack of reliable biomarkers that can distinguish triple-negative breast cancer and its different prognosis. It was shown that the occurrence and development of tumors are closely related to the glycosylation modification of proteins, and there are abnormal glycosylation modifications in cancer tissues of patients. However, up to now, there are few studies on abnormal glycosylation modification of triple-negative breast cancer. Therefore, we combined proteomics with high-throughput and site-specific glycoproteomics to quantitatively detect the level of protein and glycosylation modification in patients with triple-negative breast cancer. The study of the molecular characteristics and functional pathways of differential proteins and glycosylated proteins in the different prognosis of cancer provided more abundant references for diagnosis and treatment of triple-negative breast cancer patients. The main research results are summarized as follows:

(1) In this study, cancer tissues and paired non-cancerous adjacent tissues (NATs) of 9 patients with good prognosis and 9 patients with poor prognosis were collected. The hydrolyzed peptides were quantitatively analyzed by tandem mass tag (TMT) labeling, and glycopeptides were enriched by hydrophilic column.

(2) Through the functional and KEGG pathway enrichment analysis of differential proteins, it was found that the highly expressed proteins in cancer tissues were mainly involved in biological pathways such as cell adhesion, cell migration, and change of extracellular matrix, which promoted the invasion of tumor cells. Through the analysis of different prognosis of cancer, it was found that the expression of histone H1.2 was the highest in cancer tissues with poor prognosis, which promoted the growth and development of cancer cells.

(3) Through the functional and KEGG pathway enrichment analysis of differential glycosylation modification in cancer tissues and non-cancerous adjacent tissues showed that glycosylation of glycoprotein hemoglobin and antithrombin III may promote cancer metastasis and invasion. Based on the analysis of glycosylation modification with different prognosis, it was found that 25 glycoproteins with high expression in patients with good prognosis were associated with extracellular matrix-receptor pathway and human papillomavirus infection by KEGG pathway analysis. These pathways are closely related to cancer cell differentiation and metastasis, among which PI3K/AKT pathway is one of the common maladjusted signal pathways in triple-negative breast cancer.

(4) According to the differential protein glycosylation modification of triple-negative breast cancer tissues, glycoprotein CO6A6, DCN, and COCA1 are predictive biomarkers for patients with good prognosis, poor prognosis, and triple-negative breast cancer patients, respectively. It is worth noting that there are differences in the glycosylation on the same glycosites of glycoproteins in cancer tissues with different prognosis.

This study explained the molecular characteristics and functions of triple-negative breast cancer and its different prognosis from the differential glycosylation modification level and provided effective therapeutic strategy support for the search for biomarkers of triple-negative breast cancer.

Keywords: Triple-negative breast cancer (TNBC); cancer prognosis; proteomics; glycoproteomics; biomarkers

目 录

第一章 绪论.....	1
1.1 三阴性乳腺癌概述.....	1
1.1.1 三阴性乳腺癌现状.....	1
1.1.2 三阴性乳腺癌分型.....	1
1.2 蛋白质翻译后修饰概述.....	2
1.2.1 蛋白质翻译后修饰.....	2
1.2.2 蛋白质 N-糖基化	3
1.3 基于质谱的蛋白质组学及糖蛋白质组学概述	4
1.3.1 蛋白质组学概述.....	4
1.3.2 糖蛋白质组学概述.....	4
1.3.3 蛋白质组学及糖蛋白质组学数据分析软件	6
1.4 肿瘤生物标志物.....	7
1.4.1 蛋白质肿瘤生物标志物.....	7
1.4.2 糖蛋白肿瘤生物标志物.....	7
1.5 论文立题依据及研究意义	8
1.6 论文研究内容.....	8
第二章 材料与amp;方法	10
2.1 实验材料.....	10
2.1.1 实验试剂.....	10
2.1.2 实验仪器.....	10
2.1.3 实验样本采集.....	11
2.2 实验方法.....	11
2.2.1 HE 染色实验筛选乳腺癌组织	11
2.2.2 免疫组化实验验证 ER,PR,HER-2 及 Ki67	12
2.2.3 患者组织匀浆和蛋白质提取.....	12
2.2.4 Bradford 法测定标准蛋白曲线	12
2.2.5 蛋白质酶解和肽段获取.....	13
2.2.6 TMT 定量标记和糖肽富集	13
2.2.7 LC-MS/MS 条件的设置.....	14
2.3 实验数据处理及可视化.....	14
2.3.1 蛋白质和糖蛋白质的数据处理.....	14
2.3.2 实验数据可视化处理.....	15
第三章 结果与amp;讨论	16
3.1 组织采集及样品质控.....	16

3.1.1 患者组织样本采集及相关临床信息.....	16
3.1.2 Bradford 法检测蛋白量	18
3.2 蛋白质组学及完整糖肽的数据筛选及质控	18
3.2.1 蛋白质组学的数据筛选及质控.....	19
3.2.2 完整糖肽的数据筛选及质控.....	20
3.3 三阴性乳腺癌及不同预后情况患者组织样品的蛋白质表达水平差异	21
3.3.1 癌症组织和非癌性邻近组织中蛋白质的表达差异	21
3.3.2 不同预后乳腺癌组织间蛋白质的差异表达	24
3.4 三阴性乳腺癌及不同预后状况患者组织的蛋白质糖基化修饰水平差异	25
3.4.1 蛋白质完整糖肽糖基化修饰水平分析	25
3.4.2 癌症组织和非癌性邻近组织中完整糖肽修饰水平差异	26
3.4.3 癌症组织和非癌性邻近组织中糖蛋白修饰水平差异	27
3.4.4 不同预后乳腺癌组织完整糖肽修饰水平差异	28
3.4.5 乳腺癌组织糖蛋白糖基化修饰水平差异	31
3.5 蛋白质组学与糖蛋白质组学的差异	31
3.5.1 糖蛋白质与蛋白质之间呈负相关性.....	31
3.5.2 完整糖肽与蛋白质之间的差异分析.....	32
主要结论与展望	34
主要结论.....	34
展望.....	34
参考文献.....	36
附录 A: 附图.....	41

本论文专用术语注释表

缩写词	英文全称	中文全称
BRCA	Breast cancer	乳腺癌
TNBC	Triple-negative breast cancer	三阴性乳腺癌
NATs	Non-cancerous adjacent tissues	非癌性邻近组织
MS	Mass spectrometry	质谱
HCD	High collision-induced dissociation	高能碰撞解离
FDR	False discovery rate	错误发现率
PSM	Peptide spectrum match	肽段匹配谱图
DTT	Dithiothreitol	二硫苏糖醇
IAM	Iodoacetamide	碘乙酰胺
ER	Estrogen receptor	雌激素受体
PR	Progesterone receptor	孕激素受体
HER-2	Human epidermal growth factor receptor-2	人表皮生长因子受体-2
HE	Hematoxylin-eosin staining	苏木精-伊红染色法
Ki67	Proliferation marker protein Ki67	增殖标记蛋白 Ki67
PTM	Post-translational modifications	翻译后修饰
IHC	Immunohistochemistry	免疫组织化学
TMT	Tandem mass tag	串联质谱标签
FC	Fold change	差异倍数变化
IGPs	Intact glycopeptides	完整糖肽
GO	Gene ontology	基因本体
BP	Biological process	生物过程
CC	Cellular component	细胞组分
MF	Molecular function	分子功能
KEGG	Kyoto Encyclopedia of Genes and Genomes	京都基因与基因组百科全书

第一章 绪论

1.1 三阴性乳腺癌概述

1.1.1 三阴性乳腺癌现状

乳腺癌是乳腺组织细胞在多种致癌因素诱导下发生癌变的恶性肿瘤，是女性最常见的肿瘤之一，也是全世界人类癌症死亡率最高的肿瘤之一^[1]。《全球癌症报告》统计结果显示，乳腺癌已正式取代肺癌成为全球第一大癌症，2020 年新增乳腺癌患者数达 226 万^[2]。在中国，乳腺癌现况也十分严峻，仅 2020 年新增确诊乳腺癌患者数达 41.6 万，死亡患者数达 11.7 万^[3]。我国每年新增确诊乳腺癌患者中有 3%-10% 的患者出现远端转移，约 30% 的乳腺癌患者最终会发展成为晚期^[4-5]。

三阴性乳腺癌（Triple-negative breast cancer, TNBC）是乳腺癌的病理亚型之一，患者癌症复发率高，且预后较差。三阴性乳腺癌患者体内缺乏雌激素受体（Estrogen, ER），孕激素受体（Progesterone receptors, PR）和人表皮生长因子受体-2（Human epidermal growth factor receptor-2, HER-2）的表达，其病理特征为较高的转移率和较差的恶性程度，然而传统的激素治疗和靶向治疗效果较差^[6]。近期统计研究显示，三阴性乳腺癌患者约占所有乳腺癌患者的 15%-20%^[7]，且生存时间短，确诊后 5 年内的死亡率为 40%^[8]。三阴性乳腺癌具有高度的侵袭性，约 46% 的患者发生肿瘤转移后平均生存时间仅为 13.3 个月。三阴性乳腺癌患者术后复发率高达 25%，平均复发时间为 19-40 个月，患者复发后 3 个月内的死亡率高达 75%^[9-10]，生存期仅为 12-18 个月^[11]。临床观察结果显示，患者癌症复发后发生远端转移（包括脑转移）和死亡的风险更高，生存期更短。相较于其它乳腺癌亚型，三阴性乳腺癌的治疗策略进度远远落后^[12]。三阴性乳腺癌早期患者的治疗手段主要为手术切除肿瘤，然而患者术后复发频繁^[13]；晚期患者由于癌症的广泛转移，因此无法通过手术治疗，通常采取放疗及化疗来延长病人的生存期。从预后情况来看，三阴性乳腺癌相较其它乳腺癌亚型的表现更差，患者的复发转移率更高，并且肿瘤复发转移后将无法治愈。

1.1.2 三阴性乳腺癌分型

乳腺癌是具有高度复杂性和显著异质性的恶性肿瘤之一。2013 年圣加仑国际乳腺癌大会基于乳腺癌临床病理表征确定了乳腺癌的亚型分类，如下表所示^[14]：

表 1-1 乳腺癌的亚型
Table 1 Subtypes of breast cancer

亚型分类	临床病理分类
Luminal B (HER 阳性)	ER 阳性, HER-2 过表达
Luminal B (HER 阴性)	ER 阳性, HER-2 阴性 同时至少具备以下条件之一： Ki67≥14% PR<14%
HER-2 阳性	ER 与 PR 阴性, HER-2 过表达

表 1-1 乳腺癌的亚型 (续)

Table 1 Subtypes of breast cancer (continued)

亚型分类	临床病理分类
Luminal A	ER 与 PR 阳性, HER-2 阴性, Ki67<14%
三阴性乳腺癌	ER 与 PR 阴性, HER-2 阴性

注：百分比表示患者样本的免疫组化染色结果。

三阴性乳腺癌的临床病理表征主要由肿瘤大小、腋窝淋巴结状态和组织学分级等因素界定，然而单独依靠它们不足以指导临床治疗，无法评估患者的预后情况^[15]。2020年，江一舟等人首次将三阴性乳腺癌划分为四种分子亚型：免疫调节型（IM），腔面雄激素受体型（LAR），基底样免疫抑制型（BLIS）和间质型（MES）^[16]，为三阴性乳腺癌患者提供分子亚型治疗方案，提高治疗效率。但尚未对三阴性乳腺癌患者的预后情况进行评估。

目前针对三阴性乳腺癌患者的预后情况研究较少，因此鉴定癌症预后情况的可靠生物标志物将准确识别对治疗有反应的三阴性乳腺癌患者^[17-18]。

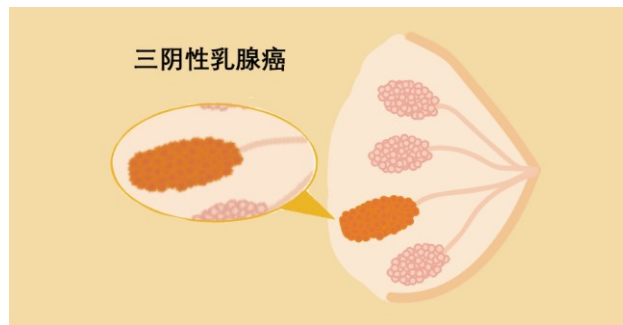


图 1-1 乳腺癌结构简化示意图

Figure 1-1 Simplified schematic diagram of breast cancer structure

1.2 蛋白质翻译后修饰概述

1.2.1 蛋白质翻译后修饰

蛋白质翻译后修饰（Protein post-translational modifications, PTMs）是对翻译结束后的前体蛋白进行后续加工以调节其活性的过程，修饰方式有功能基团的添加或亚基的水解等^[19]。人体中存在广泛的蛋白质翻译后修饰，包括 N-糖基化、甲基化、磷酸化、泛素化和乙酰化修饰等^[20]。蛋白质的结构与功能随着生物体的需求和分子环境的变化持续进行调节，蛋白质翻译后修饰是调节蛋白质功能的重要机制之一，并为其提供了结构多样性和功能多样性^[21]。

生物体内众多酶参与调节蛋白质翻译后修饰过程，若酶出现失调可能导致如癌症等疾病的发生^[22]。与正常组织细胞相比，肿瘤组织细胞的蛋白质合成及分解代谢能力较强，蛋白质表达水平特别是翻译后修饰较易出现异常^[23-24]。有研究显示，部分蛋白质翻译后修饰过程如 N-糖基化、甲基化、磷酸化和乙酰化等在乳腺癌中起着重要作用，参与上述翻译后修饰的蛋白质主要参与调节细胞活性及稳定性、响应细胞外刺激及细胞内信号

等^[25]。据报道,异常的蛋白质翻译后修饰对三阴性乳腺癌的病理过程影响较大^[27]。糖基化修饰在其他肿瘤发生发展中起关键作用,如细胞增殖分化及周期调节、血管生成、免疫应答、信号转导、氧化应激、代谢重编程、上皮间充质转化以及侵袭转移等^[27]。解析蛋白质糖基化修饰在三阴性乳腺癌中的作用可以为疾病的分子模式提供新的观点,并可能开发针对这些糖基化修饰的新型癌症疗法。

1.2.2 蛋白质 N-糖基化

蛋白质 N-糖基化是癌症细胞中最常见的蛋白质翻译后修饰之一,在真核细胞中普遍存在,指复杂的脂连接寡糖与蛋白质序列中特定的天冬氨酸位点(N-X-T/S,其中 X 是除脯氨酸外任何氨基酸)的共价结合。蛋白质 N-糖基化过程起始于内质网,通过寡糖转移酶复合体将预先组装的寡糖前体转移到新生蛋白质上,这种前体被称为脂连接寡糖。寡糖转移酶可以识别特定的天冬酰胺残基,并将脂连接寡糖转移到上面。在此之后,蛋白质被运送到高尔基体,与其共价结合的寡糖会被额外修饰,这些修饰包括剪接与添加末端糖,形成复杂的糖链结构。N-糖基化影响蛋白质的折叠、稳定性及功能^[28]。N-聚糖的存在可以影响蛋白质的构象,调节蛋白质之间的相互作用,防止蛋白质的降解。此外,N-聚糖还可以作为受体接收其他细胞上配体信号,完成跨膜信号转导,有助于细胞间的识别和交流^[29]。

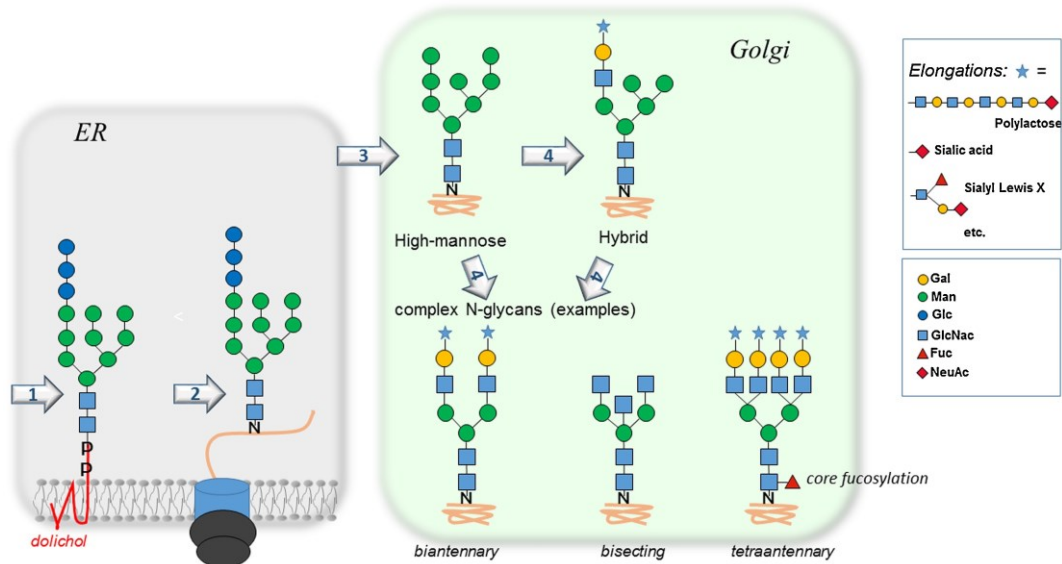


图 1-2 N-聚糖的生物合成途径^[30]

Figure 1-2 Biosynthesis pathway of N-glycans^[30]

蛋白质糖基化的改变可以影响许多细胞过程,包括细胞黏附、迁移和侵袭、血管生成、免疫逃逸和抗凋亡等^[31]。异常蛋白质 N-糖基化修饰与多种疾病有关,包括癌症、糖尿病和先天性糖基化障碍等^[32]。在癌症中,N-糖基化的改变可以产生与肿瘤相关的聚糖或糖蛋白,导致细胞出现黏附、迁移和侵袭,从而促进肿瘤的发展及转移^[33]。癌症中糖基化修饰的增加是由于携带某些特定聚糖的糖蛋白过度表达,核苷酸糖供体以及糖基转移酶和糖苷酶的表达改变引起的^[34]。蛋白质 N-糖基化修饰异常与多种癌症的发生和发展有关,例如乳腺癌中黏蛋白异常糖基化的表达。黏蛋白是一种高度糖基化的蛋白质,

参与细胞信号传递和细胞黏附，其异常糖基化修饰在乳腺癌中已被观察到，并与肿瘤侵袭性增加和患者预后不良有关^[35]。因此，肿瘤异常糖蛋白可被用作生物标志物。此外，在癌症中还观察到了用于添加和修饰 N-聚糖的糖基转移酶表达和活性的变化。例如，岩藻糖基转移酶的表达增加与乳腺癌、肝癌及结直肠癌的发生和发展有关^[36]。

总而言之，蛋白质 N-糖基化的调节是一个复杂的动态过程，在蛋白质功能和细胞通讯中起着关键作用，从而影响到人类的健康和疾病。蛋白质 N-糖基化在癌症中的改变及作用如今已成为热门研究领域。人们对癌细胞中异常糖基化蛋白质生物标志物越来越感兴趣，生物标志物可确定癌症类型，从而对病人采取高效且针对性的治疗方法。

1.3 基于质谱的蛋白质组学及糖蛋白质组学概述

质谱仪 (Mass spectrometry, MS) 作为一种强大的分析检测 (糖) 蛋白质的工具，可根据 (糖) 蛋白质离子碎片的不同质荷比进行分离，利用蛋白组学分析软件 MaxQuant 或糖蛋白质组学分析软件 GPQuest 等解析质谱结果，以定量分析复杂混合物中的蛋白质和其他分子。

1.3.1 蛋白质组学概述

蛋白质组学技术作为一种高通量分析技术，可以同时多种样本进行深度解析，包括组织、细胞、血液、尿液和脑脊液等。生物标志物是蛋白质组学的重要发现与应用，也是患者生理状态的指标。通过鉴定在疾病组织和健康组织中差异表达的蛋白质，可以开发出应用于各种疾病 (癌症、心血管疾病和神经疾病等) 及预后情况诊断的生物标志物^[37]。蛋白质组学还用于研究蛋白质翻译后修饰和蛋白质之间的相互作用，如糖基化以及蛋白质在细胞和组织中的功能解析^[38]。通过研究蛋白质及其修饰之间的相互作用，深入了解正常生理组织和疾病组织的信号通路和细胞过程的差异，对于更有效的疾病诊断与治疗策略的开发有重要研究意义。

基于质谱的蛋白质组学方法有自上而下法 (Top down) 和自下而上法 (Bottom up) 共两种，其中自上而下法是使用高分辨率质谱仪直接分析完整的蛋白质，这种方法提供了蛋白质序列、翻译后修饰和蛋白质变体的完整信息^[39]。然而，蛋白质的结构及功能较为复杂，且溶解性差，导致自上而下法对蛋白质鉴定的灵敏性较低，存在技术困难。本研究的蛋白质组学采用自下而上法，将蛋白质酶解为多肽，通过液相色谱分离得到离子化多肽后利用质谱仪进行分析，使用软件将谱图信息与蛋白质序列数据库进行匹配，以鉴定样品多肽，最终推测蛋白质信息^[40]。自下而上法具有效率高且耗时较短的优点。相较于自上而下法，自下而上法被更广泛应用于蛋白质的鉴定和表征。

1.3.2 糖蛋白质组学概述

1.3.2.1 糖蛋白质组学意义

糖蛋白质组学作为蛋白质组学的一个分支，利用质谱对糖蛋白质进行全面解析。糖蛋白质是与多糖共价连接的蛋白质，糖蛋白质组学包含鉴定和定量生物样品中存在的所有糖蛋白以及表征附着在蛋白上的糖链组成及结构。糖蛋白质在许多生物过程中发挥重

要作用，包括细胞信号、细胞黏附和免疫识别。蛋白质糖基化的改变与许多疾病有关，包括癌症、自身免疫性疾病和传染病等。因此，了解糖蛋白质的复杂交互网络对其在健康和疾病人群中的功能差异分析具有重要意义。由上可知，检测健康和疾病人群中差异表达的糖基化蛋白可以提高人们理解其与疾病发生发展的关系。然而，由于糖蛋白质在蛋白质中相对丰度较低以及糖链结构的异质性较高等特点，其在质谱中的鉴定覆盖率远低于没有糖链修饰的多肽^[41]。因此，蛋白质组学不能用于捕捉糖肽信息，无法分析复杂生物样品中糖蛋白质信息。糖蛋白质组学的研究具有许多潜在的后续应用，例如新的疾病生物标志物，开发针对糖链的治疗方法，以及了解生物系统中糖蛋白与其他分子之间的复杂交互作用等^[42-44]。糖蛋白质组学是新兴且发展迅速的领域，其有助于理解蛋白质糖基化在健康和疾病人群中的差异。

1.3.2.2 糖蛋白质组学制备样品方法

在复杂样品中分离及浓缩糖基化蛋白是糖蛋白质组学研究的关键。目前常见的分离富集完整糖肽的研究方法，包括凝集素亲和层析、胍化学富集法和亲水色谱法等。其中，凝集素亲和技术利用了凝集素与糖链结构的特异性结合这一性质，对糖肽进行分离纯化从而富集^[45]。然而，此方法存在非特异性结合、选择性有限与结合能力弱等缺点，例如一些凝集素可以与未被糖基化的肽段进行非特异性结合，导致捕获不需要的信息，降低了方法的特异性；部分糖链结构没有合适的凝集素用于识别，使得凝集素的选择性受到限制；凝集素与糖肽的结合能力有限，这可能导致糖肽捕获不完全。胍化学富集法通过酰肼试剂选择性地修饰氧化糖链并将其与酰肼共价结合，从而捕获糖肽^[46]。胍化学富集法也有潜在的缺点，例如酰肼化学试剂依赖于被氧化的糖链，然而部分糖链结构不容易被氧化，导致糖蛋白质的覆盖率有限。本研究使用亲水色谱法，利用糖肽单糖上的羟基具有亲水性这个特性，从而富集糖肽。将复杂样品溶于高有机相溶剂中，当样品溶液经过亲水性 MAX 柱时，糖链上的亲水性羟基会与填料结合，其余样品则同溶剂一起流出。最后逐步提升洗脱液的极性百分比以提高糖肽洗脱强度，得到富集糖肽^[47]。亲水色谱法富集糖肽具有以下优点：高灵敏度，可以富集低丰度糖肽，为分析糖基化修饰水平较低是糖蛋白提供帮助；与质谱仪兼容，从而能够识别和表征附着在多肽上的糖链组成；高复现性，能够可靠地检测和定量生物样品中的糖肽；所需样品量少，为稀有样品的分析提供帮助；可检测丰富且广泛的糖链结构。综上所述，亲水色谱法是糖蛋白质组学分析检测糖肽的有力工具。

基于质谱的糖蛋白质定量分析可采用同位素试剂等量标记方法或非标记方法。非标记方法定量糖蛋白质较为简便，但质谱仪每次的工作状态不一定相同，使得其电离效率不稳定，检测精确性较差^[48]。然而，上述问题可以通过同位素试剂等量标记方法解决。本研究引入串联质量标签（Tandem mass tag, TMT），可同时标记多个样品，被广泛应用于疾病研究中^[49-50]。同位素试剂的反应基团用于多肽共价修饰，并且每个试剂包含一个具有不同质量同位素的报告基团，然而不同的同位素试剂具有相同的总质量。通过解离前体离子，二级质谱中产生了不同质量的报告基团，其强度的差异代表所标记多肽的

相对丰度。因此引入同位素试剂可以减少不同样品批次之间的差异，确保定量结果的准确性^[51]。

1.3.2.3 糖蛋白质组学的质谱方法

糖基化修饰蛋白分析中常用的两种不同的质谱学方法有数据依赖采集模式 (Data dependent acquisition, DDA) 和数据非依赖采集模式 (Data independent acquisition, DIA)。DIA 主要通过一级质谱和离子碎片交替采集信息，不筛选母离子，而 DDA 方法为一级质谱分析后，根据提前设置的筛选条件进行母离子的筛选，从而有选择地分析二级质谱，获得更多的碎片信息^[52]。与 DIA 相比，DDA 有很多的优点：因其对高丰度的前体离子有选择性地碎裂，从而更加灵敏和准确地检测糖肽；适用于靶向分析，从而能够检测特定糖肽；因其可提供详细的多肽序列和蛋白修饰信息，可以用于样品的糖肽检测与表征。总之，DDA 相较于 DIA 更加灵敏，适用于样品的靶向检测分析以及表征。因此，本研究使用 DDA 作为分析糖基化修饰蛋白的质谱学方法。

质谱仪有多种离子碎裂方式用于检测糖肽，包括电子转移解离 (Electron transfer dissociation, ETD)、碰撞诱导解离 (Collision-induced dissociation, CID) 以及高能碰撞解离 (High energy collision dissociation, HCD) 等。其中 ETD 是将阴离子基团上的一个电子转移到带正电荷的肽段上，产生互补性的 c、z 离子的技术，常用于断裂肽键。CID 是样品与气体中性分子发生碰撞导致 b、y 离子及碎片产生的技术^[53]。然而，样品裂解过程中得到的有效碎片峰值普遍较小，因此在质谱图中杂峰较多且有效峰值较少，难以对糖肽进行精确分析^[54]。HCD 与 CID 裂解方式相似，但 HCD 可以改善 CID 丢失的低质量碎片效应，因此 HCD 碎裂方式更高效^[55]，故本研究使用 HCD 碎裂方式。

1.3.3 蛋白质组学及糖蛋白质组学数据分析软件

质谱数据搜库软件用于将质谱中碎裂的离子信息与已知数据库进行匹配，最终获得样品蛋白质以及完整糖肽信息。由于样品通过高通量质谱仪分析后将会获得相对庞大且复杂的数据，因此蛋白质组学以及糖蛋白质组学的数据分析软件的开发与优化也在不断完善。

对于蛋白质组学数据的解析，MaxQuant 软件^[56]被广泛应用于蛋白质组学的大规模复杂质谱数据，并为数据分析提供了全面的工作流程。该软件使用先进的算法来识别和量化肽段信息，并为数据可视化和统计分析提供支持，支持定量标记 (如 TMT 标记)，鉴定蛋白质准确性高。

人体病理过程中，部分糖基化蛋白的变化不仅体现于蛋白质的丰度，还表现在不同糖型与糖基化位点的占有上。质谱仪分析技术的发展使得复杂生物样品中完整糖肽的解析成为可能，然而识别其详细信息仍具有挑战性。因此，本文引入 GPQuest 分析软件^[57]以识别复杂样品 HCD 模式碎裂的完整糖肽。通过 LC-MS/MS 对富集的完整糖肽进行分析，并以氧鎓离子作为糖肽的特征离子来确定完整糖肽。最后将糖肽谱图与糖肽数据库进行比较，从而确定每个完整糖肽信息。此外，谱库匹配使研究者有可能在特定糖位上识别可能从预定的糖链数据库中缺失的新的或修饰的糖链。

1.4 肿瘤生物标志物

1.4.1 蛋白质肿瘤生物标志物

蛋白质作为肿瘤的生物标志物 (Biomarker) 在疾病检测等方面具有特异性优势, 为医疗人员提供更准确的诊断方案。同时, 部分蛋白质作为生物标志物可以为患者提供癌症复发率等信息, 具有预后意义。

临床研究显示, 蛋白质酪氨酸激酶 2 (JAK2) 在三阴性乳腺癌患者的癌症组织中相对丰度增加, 通常被认为是不良预后情况因素^[58]; 据报道, 组蛋白去甲基酶 4B (KDM4B) 和组蛋白 H1.2 是三阴性乳腺癌的重要靶标, 其中组蛋白去甲基酶 4B 的异常表达可导致细胞凋亡^[59], 组蛋白 H1.2 可以促进癌细胞的生长^[60]; 有研究显示, 蛋白 IF1AX (真核细胞翻译起始因子) 通过促进 G1/S 细胞周期转换从而导致乳腺癌细胞增殖^[61]; 蛋白 LUM 可以抑制乳腺癌细胞中上皮-间充质转化 (Epithelial-to-mesenchymal transition, EMT) 等癌症转移途径^[62]。

1.4.2 糖蛋白肿瘤生物标志物

糖蛋白作为肿瘤生物标志物是指用于检测生物体内是否出现疾病或异常生理状态的物质。其中糖蛋白肿瘤生物标志物主要被用来指示生物体细胞是否出现癌变。糖蛋白质差异表达和糖基化异常改变与癌症、神经疾病和心血管疾病等多种疾病有关^[63-65]。利用质谱对糖蛋白质进行检测, 可以提供对糖蛋白及其糖链组成全面的解析。使用基于质谱学的糖蛋白质组学分析技术检测和定量临床样品中特定糖蛋白肿瘤生物标志物的存在与含量, 以鉴定患者机体是否出现癌变。

作为肿瘤的生物标志物, 糖蛋白相较于非糖基化蛋白的优点在于糖蛋白通常对疾病或生理状态更具特异性, 且多糖组成及结构可以提供关于疾病进展和治疗的额外信息。然而, 糖蛋白质在生物个体之间糖基化模式可能存在差异, 因此我们开发了独特的分析方法检测和定量不同机体的糖蛋白相对丰度。

据报道, 乳腺癌的糖蛋白生物标志物包括唾液酸化的多糖和岩藻糖化的葡聚糖。例如, α -1-酸性糖蛋白 (AGP) 上岩藻糖修饰的多糖水平的升高与乳腺癌风险增加和预后不良有关^[66]; 黏蛋白 (MUC1) 是一种跨膜糖蛋白, 与正常乳腺组织相比, 黏蛋白通常在乳腺癌细胞中过表达, 与肿瘤的发生、转移和预后不良有关^[67]; 乳腺癌患者的免疫球蛋白 A1 (IgA1) 的聚糖唾液酸化相比于健康个体有所升高^[68]。此外, 有研究显示异常的糖基化修饰在乳腺癌组织中也有所表征, 例如质谱检测到的高甘露糖型 N2H8 和 N2H9 的数量在三阴性乳腺癌组织中多于 HER-2 阳性乳腺癌组织^[69]。

总体而言, 三阴性乳腺癌糖蛋白生物标志物的识别和验证可以为诊断、预后和治疗疾病决策提供有价值的信息。然而, 需要进一步的研究来充分了解三阴性乳腺癌糖基化模式的复杂性, 并开发出更灵敏和可靠的糖蛋白生物标志物。

1.5 论文立题依据及研究意义

乳腺癌已对人类的健康造成了严重的危害。在中国，乳腺癌是一个严峻的健康问题，在过去的几十年里，乳腺癌患者的发病率迅速上升。据最新数据统计，我国绝经前女性确诊乳腺癌的比例较高，而且患者确诊的年龄还在不断减小，并且女性获得乳腺癌筛查和治疗仍然是一个巨大的挑战。三阴性乳腺癌是乳腺癌的一种病理亚型，具有较高的侵袭性和较差的预后情况，临床上更难治疗。与其他乳腺癌亚型相比，三阴性乳腺癌患者在治疗后的 2-5 年内更有可能复发。然而，部分预后良好三阴性乳腺癌的女性在治疗后确实取得良好的结果。早期三阴性乳腺癌患者的预后一般好于晚期疾病，晚期或转移性三阴性乳腺癌患者的预后一般较差。因此，研究可用于鉴定三阴性乳腺癌患者预后的生物标志物对于寻找对治疗有良好效果的三阴性乳腺癌患者至关重要。

糖蛋白是一种被多糖分子修饰的蛋白质，其已经成为包括癌症在内的几种疾病的生物标志物。糖蛋白肿瘤生物标志物具有一些潜在优势：

(1) 高灵敏度：机体内糖蛋白通常浓度较低，但使用质谱仪分析方法以高灵敏度进行检测，使其成为癌症疾病的早期检测生物标志物。

(2) 特异性：糖蛋白对癌症或其他疾病具有高度特异性，这使它们易于区分不同疾病类型。

(3) 稳定性：糖蛋白通常比其他类型的生物标志物更稳定，如 mRNA 或 DNA。这意味着可以更容易地从长期储存的生物样本中检测糖蛋白。

(4) 靶向治疗的可能性：一些糖蛋白可能仅参与某种疾病途径，这使它们成为治疗的新潜在靶点，代表着使用糖蛋白作为肿瘤生物标志物将有助于癌症的诊断。

截至目前，有关三阴性乳腺癌预后情况的研究多停滞在医学统计水平。据报道，大多数研究仅限于三阴性乳腺癌相关的基因水平分析。然而，三阴性乳腺癌的产生与发展是患者的个人遗传信息、身体状况及生存环境等多种因素共同作用的病理结果，若仅从基因水平进行分析，忽视蛋白质翻译及翻译后修饰等水平的调控过程，具有一定的局限性。大量研究显示，机体组织的 mRNA 和蛋白质的相对丰度并不存在“一一对应”的线性关系，因此不能根据 mRNA 的表达水平来确定蛋白质的表达水平^[70]。因此，三阴性乳腺癌的异常蛋白质糖基化修饰的研究具有重要科学意义，对临床诊断及治疗具有应用价值^[71]。但目前的三阴性乳腺癌不同预后情况的差异蛋白质表达及糖基化修饰水平的研究进程缓慢。

综上所述，糖蛋白作为肿瘤生物标志物可以改善疾病的检测、诊断和治疗。目前的研究重点是识别适用于三阴性乳腺癌不同预后的新型糖蛋白生物标志物。

1.6 论文研究内容

本项研究中，从被确诊为三阴性乳腺癌的 18 位患者中分别采集癌症组织和非癌性邻近组织 (Noncancerous adjacent tissues, NATs)，其中 9 位为预后良好患者，9 位为预后不良患者。对 18 位患者组织进行全面的 TMT 试剂定量标记的蛋白质组学和糖蛋白质组学分析，通过蛋白质糖基化修饰水平和蛋白质表达水平，分子模式和功能富集对癌症

两种预后情况进行详细的解析。这些研究或许会有助于解释不同预后情况的三阴性乳腺癌分子模式及相关的调控途径，并阐明二者之间存在的差异，以提高鉴定不同预后患者的准确性和可靠性。

论文研究内容如下：

(1) 三阴性乳腺癌患者组织样本的采集及筛选。收集并确认三阴性乳腺癌患者组织，通过苏木精-伊红染色以及免疫组化染色的方法从已收集的乳腺癌患者组织样本中筛选出合格的三阴性乳腺癌样本，即 ER、PR 和 HER-2 的染色结果均为阴性。采集样本时确认患者信息与样本信息一一对应。采集样本后，从组织块中提取蛋白质并使用 Bradford 法测定浓度，制作标准蛋白曲线并计算每个样本蛋白含量，记录并保存信息。

(2) 肽段富集及数据质控。将从组织提取的蛋白酶解获得的肽段进行 TMT 试剂等量标记后，分别除盐富集多肽和完整糖肽，使用高分辨率液相质谱分析癌症组织以及配对患者的非癌性邻近组织的糖蛋白质及蛋白质，进行全面的 TMT 标记糖蛋白质组学和蛋白质组学分析，利用生物信息学工具对糖蛋白质及蛋白质联合处理分析。将软件搜库结果进行初步数据质控以确保后续分析数据质量合格。

(3) 蛋白质糖基化修饰水平和蛋白质表达水平进行对比分析。将糖蛋白质组学和蛋白质组学的结果分别进行定性分析，统计检测到的糖蛋白质数目以及蛋白质数目，N-糖基化多肽以及 N-糖基化位点等信息。

(4) 蛋白质糖基化修饰水平和蛋白质表达水平的定量分析。对糖基化修饰水平及蛋白质的表达水平进行统计及分析，对比癌症样品与非癌性邻近组织之间蛋白质糖基化修饰差异和蛋白质差异表达水平，寻找能够通过糖蛋白质糖基化修饰水平区分预后良好与预后不良的特殊糖蛋白。此外，通过功能注释分析探究癌症不同预后情况的分子模式及功能通路差异，为三阴性乳腺癌不同预后情况的糖蛋白生物标志物提供数据支持。

第二章 材料与方法

2.1 实验材料

2.1.1 实验试剂

表 2-1 实验所用试剂

Table 2-1 The experimental reagents

试剂名称	生产厂商		
盐酸 (HCl)	国药集团化学试剂有限公司		
乙醇 (C ₂ H ₅ OH)			
二甲苯 (C ₆ H ₄ (CH ₃) ₂)			
DAB 染色液			
柠檬酸			
柠檬酸钠			
磷酸盐缓冲溶液		生工生物工程 (上海) 股份有限公司	
羟乙基哌嗪乙磺酸 (HEPES)			
胰蛋白酶 (Trypsin)			北京生夏蛋白技术有限公司
组织固定液			上海碧云天生物技术有限公司
组织分化液			
碳酸氢铵 (NH ₄ HCO ₃)	美国 Sigma-Aldrich 生物试剂公司		
氯化钠 (NaCl)			
尿素 (Urea)			
DTT			
IAM			
乙腈 (Acetonitrile, ACN)			
三氟乙酸 (Trifluoroacetic acid, TFA)			
甲酸 (Formic acid, FA)			
Bradford 蛋白检测试剂盒		上海碧云天生物技术有限公司	
C-18 柱 (Sep-Pak 1cc)		美国 Waters 公司	
MAX 柱 (Oasis 1cc)			
30kDa 超滤膜	美国 Millipore 公司		
Ziptip C-18 柱			

2.1.2 实验仪器

表 2-2 实验所用仪器设备

Table 2-2 The experimental instruments

仪器名称	仪器厂商
超声波细胞粉碎机	宁波新芝生物科技股份有限公司
多功能酶标仪	美国 Bio-Rad 公司
HYL-A 全温摇瓶柜	强乐实验设备有限公司
电子天平	Mettler Toledo 仪器有限公司
涡旋仪	海门市其林贝尔仪器制造有限公司
CT-5RE 离心机	日本日立公司

表 2-2 实验所用仪器设备 (续)

Table 2-2 The experimental instruments (continued)

仪器名称	仪器厂商
干式恒温器	杭州奥盛仪器有限公司
移液枪	
NanoDrop 微量分光光度计	美国 Thermo Fisher Scientific 公司
Orbitrap fusion lumos 质谱仪	

2.1.3 实验样本采集

(1) 实验样本: 本研究采集的 18 例三阴性乳腺癌患者组织样本均来自福建医科大学协和医院 (全乳房切除术或保乳手术), 每位患者的组织样本包括癌症组织和其配对的非癌性邻近组织。福建医科大学协和医院伦理委员会 (中国福建省) 已批准本项研究, 且参与研究者均提供书面知情同意书。

(2) 患者临床信息: 18 位患者均为女性, 且每位患者在手术前没有接受化疗或放疗。采集患者的原发性癌症组织与配对的非癌性邻近组织, 二者之间距离大于等于 3 cm。所有患者均被组织病理学检查为三阴性乳腺癌亚型, 癌症组织中 ER、PR 与 HER-2 均呈阴性。记录患者的多项临床信息, 包括无病生存期、癌症组织分期、分子亚型和年龄。

(3) 样品保存: 手术切除的每个患者的组织块迅速转移至无菌管, 液氮冷冻并转移保存在 -80°C 环境。

2.2 实验方法

2.2.1 HE 染色实验筛选乳腺癌组织

福尔马林固定石蜡包埋 (Formalin fixed paraffin embedded, FFPE) 组织切片的脱蜡: 首先将 FFPE 组织切片在盛有二甲苯溶液的玻璃皿中充分浸泡两次以清洗石蜡, 每次持续 10 min。接着使用无水乙醇使切片脱蜡时的二甲苯被洗脱, 持续 10 min。组织切片表面的二甲苯被洗脱后, 将其依次置于浓度递减的乙醇溶液中, 每次各 5 min, 乙醇浓度分别为 95% (v: v)、90% (v: v)、80% (v: v) 和 70% (v: v), 接着使用蒸馏水缓慢清洗组织切片并于室温干燥。细胞核染色实验步骤: 组织切片通过苏木素染色液充分染色 10 min 后, 使用蒸馏水清洗多余染色液。接着使用 1% (v: v) 盐酸乙醇溶液对切片进行分化处理, 持续 5 min, 除去与细胞核和细胞质结合过多的染液。蒸馏水清洗并使用 0.5% (v: v) 氨水处理, 使细胞核于弱碱性溶液中, 被染成蓝色后蒸馏水清洗组织切片。细胞质染色实验步骤: 组织切片通过伊红染色液染色持续 3 min 后, 使用蒸馏水清洗。使用浓度递增的乙醇溶液进行切片的脱色, 每次各 3 min, 重复两次, 乙醇浓度分别为 80% (v: v)、85% (v: v) 和 95% (v: v)。接着将组织切片放入二甲苯溶液中浸泡, 每次持续 10 min, 重复两次后干燥切片。最后向切片缓慢滴加中性树脂进行封片, 防止出现气泡。使用显微镜对通风干燥后的组织切片进行测定并收集其图像, 记录数据用于后续分析。

乳腺癌组织的评判指标主要有腺管形成程度、细胞核形态以及细胞核有丝分裂。与患者正常组织细胞相比，乳腺癌腺管分化程度较高，且细胞形状多为片块或条索状；细胞核大小具有明显差异且形态多样，核仁明显；染色质增多，且分布不均。

2.2.2 免疫组化实验验证 ER, PR, HER-2 及 Ki67

FFPE 组织切片完成脱蜡后，使用蒸馏水清洗并晾干，放入盛有 pH 值为 6.0 的柠檬酸盐缓冲液的玻璃皿中浸泡组织切片，每次持续 2 min，重复 4 次。取出切片，使用蒸馏水冲洗，每次持续 3 min，重复 2 次。最后使用磷酸盐缓冲液轻柔地冲洗切片，每次持续 3 min，重复 2 次。待组织切片干燥后，放入干燥玻璃皿中，缓慢滴加 4% (v: v) H_2O_2 溶液，确保其充分覆盖组织切片，室温孵育 10 min。接着使用磷酸盐缓冲液冲洗组织切片，每次持续 2 分钟，重复 4 次。组织切片于室温干燥后，向其表面缓慢滴加一抗试剂，倾斜切片使一抗全部均匀覆盖在切片上。然后将切片平缓地移入保湿盒中，置于 4°C 冰箱孵育 24 h。待一抗孵育完成后将切片取出，放入干燥的玻璃皿中，使用磷酸盐缓冲液缓慢清洗切片 4 次。待组织切片干燥后缓慢滴加二抗溶液，确保其充分覆盖组织切片，于室温孵育 2.5 h。组织切片孵育二抗完成后，使用磷酸盐缓冲液缓慢冲洗 4 次后，室温静置干燥。接着将 DAB 染色液在组织切片上缓慢滴加，静置时观察组织切片的染色情况，若切片显色，使用蒸馏水缓慢清洗组织切片 4 次以停止反应。接着将切片转移至装有 75% (v: v) 乙醇溶液的玻璃容器中，室温静置 3 分钟后再次使用蒸馏水清洗组织切片 3 次。接着配制不同浓度梯度的乙醇溶液，其中乙醇溶液浓度分别为 75% (v: v)、85% (v: v) 和 95% (v: v)，将组织切片依次置于装有乙醇溶液的玻璃皿中进行脱水。最后将组织切片浸入二甲苯溶液中以脱水切片，每次持续 10 min，重复 2 次。待切片于静置干燥后，向其表面缓慢滴加中性树胶，使树胶均匀覆盖切片表面，室温静置 3 h 后，通过显微镜观察组织切片，采集图像并记录实验结果。

2.2.3 患者组织匀浆和蛋白质提取

首先将 -80°C 冷冻保存的组织块样品取出于冰上消融。使用酒精消毒手术剪刀将组织剪成小块放入新 EP 管，标记样品名称后加入 1 mL 磷酸盐缓冲液，涡旋 20 s，5000 rpm 离心 4 min 后去除上清液。重复上述清洗步骤直至溶液无色。接着使用组织研磨仪研磨组织块，研磨仪的模块提前放在 -20°C 冰箱预冷 20 min。然后将剪碎组织及研磨钢珠 7 粒转入新 EP 管中，标记样品名称后加入 1 mL 磷酸盐缓冲液，使用组织研磨仪器进行蛋白质提取。运行条件：运行时间 60 s，中断时间 60 s，频率 60 Hz，温度 -30°C，重复研磨 2 次。将研磨完成的组织以 12000 rpm 离心 20 min 后吸取上清液，加入尿素使溶液终浓度为 8 mol·L⁻¹。最后使用 Bradford 法测定蛋白质浓度，记录并保存数据。

2.2.4 Bradford 法测定标准蛋白曲线

首先配制 1 mg·mL⁻¹ 的牛血清蛋白 (Bovine serum albumin, BSA) 溶液，然后加入磷酸盐缓冲液配制一组浓度梯度的 BSA 溶液，BSA 浓度分别为 1.5 mg·mL⁻¹，1.0 mg·mL⁻¹，0.75 mg·mL⁻¹，0.5 mg·mL⁻¹，0.25 mg·mL⁻¹ 和 0.125 mg·mL⁻¹，上述溶液终体积

均为 100 μL 。接着，每个浓度梯度的蛋白标准溶液各取 15 μL ，加入 500 μL Bradford 工作液中，移液枪吹吸混匀，室温静置反应 5 min。标准蛋白曲线如图 2-1 所示。

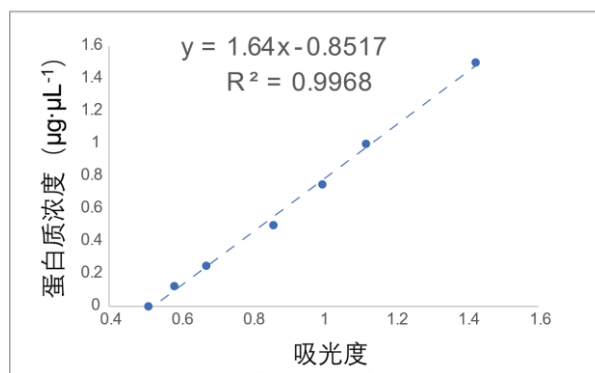


图 2-1 标准蛋白定量曲线

Figure 2-1 Standard protein quantitative curve

图注：Bradford 标准蛋白定量曲线用于检测样品蛋白浓度，横轴代表吸光度，纵轴代表蛋白质浓度。

接着将反应完成的溶液取 300 μL 加入 96 孔板中。使用酶标仪测定在 A595nm 时所有样品的吸光度，记录数据。根据标准蛋白浓度对应的吸光值绘制标准曲线，并计算样品蛋白浓度。

2.2.5 蛋白质酶解和肽段获取

首先取蛋白样品转移至新 EP 管，加入 DTT 溶液使得其终浓度为 10 $\text{mmol}\cdot\text{L}^{-1}$ ，反应条件为 56 $^{\circ}\text{C}$ ，45 min。接着加入 IAM 溶液使 IAM 最终浓度为 20 $\text{mmol}\cdot\text{L}^{-1}$ ，反应条件为室温避光，35 min。将样品混合液转移至 30kDa 的超滤膜上，10000 g 离心 3 min，减小样品混合液体积至 200 μL ，换新 EP 管并将超滤膜倒置，2800 g 离心 3 min。收集溶液后，向 EP 管中加入 1.2 mL 40 $\text{mmol}\cdot\text{L}^{-1}$ NH_4HCO_3 溶液，使原溶液中尿素浓度降至 1.5 $\text{mol}\cdot\text{L}^{-1}$ 。然后加入胰蛋白酶以终浓度 1:40 (m: m) 的比例分批加入，37 $^{\circ}\text{C}$ 过夜孵育 12 h。最后使用 NanoDrop 检测酶切肽段浓度并记录，TFA 调节溶液至 pH 为 2.0 以终止酶解反应。多肽的获取需 C-18 柱除盐富集，首先向 C-18 柱中加入 1 mL 100% (v: v) ACN，重复三次；接着上柱 1 mL 0.1% (v: v) TFA / 75% (v: v) ACN，重复三次；最后上柱 1 mL 0.1% (v: v) TFA，重复四次。上述步骤用于活化 C-18 柱。接着将酸化后的多肽样品加入 C-18 柱中，重复三次。然后向 C-18 柱中加入 1 mL 0.1% (v: v) TFA，去除样品中小分子盐，重复三次。除盐后上柱 1 mL 1% (v: v) FA 溶液更换 C-18 柱溶液体系，重复三次。最后向 C-18 柱中加入 1 mL 0.1% (v: v) FA / 50% (v: v) ACN 溶液，洗脱与 C-18 柱结合的肽段。

2.2.6 TMT 定量标记和糖肽富集

TMT 定量标记组学试验：真空浓缩仪干燥溶液得到肽段干粉，并使用 1 $\text{mol}\cdot\text{L}^{-1}$ HEPES 溶液复溶，NanoDrop 测定浓度。18 个病人共 36 个样品分为 4 组，通过十种 TMT 标记试剂标记样品，分别为 126、127-C、127-N、128-C、128-N、129-C、129-N、130-C、130-N 和 131^[72]。各取每个通道的标记样品 5 μg ，Ziptip-C18 柱除盐以检测标记效率。

使用亲水材料 MAX 柱对完整糖肽 (Intact glycopeptides, IGP) 进行富集, 其试验步骤参考文献^[73]中具体内容。使用 NanoDrop 测定完整糖肽浓度并记录, 真空浓缩仪干燥溶液后用 2% (v:v) ACN / 0.1% (v:v) FA 复溶样品, 调节糖肽浓度至 $0.5 \text{ mg}\cdot\text{mL}^{-1}$, 取 $3 \mu\text{g}$ 糖肽用于糖蛋白质组学分析。

2.2.7 LC-MS/MS 条件的设置

将肽段与糖肽干粉分别重新悬浮于 2% (v:v) ACN / 0.1% (v:v) FA 溶液中, 调节肽段溶液终浓度为 $0.5 \mu\text{g}\cdot\mu\text{L}^{-1}$, 糖肽溶液终浓度为 $0.7 \mu\text{g}\cdot\mu\text{L}^{-1}$, 为确保有足够的样品用于分析, 每次进样体积均为 $3 \mu\text{L}$ 。样品均通过 Orbitrap Fusion Lumos 质谱仪进行高分辨率分析, 且每个样品重复两次质谱分析。首先, 样品通过 EASY-nLC 1200 系统内部配备的 RSLC-C-18 色谱柱 ($75 \mu\text{m}\times 25 \text{ cm}$) 进行分离。本实验采用 DDA 质谱方法和离子碎裂模式为 HCD。流动相流速维持在 $650 \text{ nL}\cdot\text{min}^{-1}$, 且由 A 相: 0.1% (v:v) FA 和 B 相: 90% (v:v) ACN / 0.1% (v:v) FA 共同组成。设置样品的洗脱梯度: 0 min-88 min, 0%-4% (v:v); 88 min-110 min, 4%-22% (v:v); 110 min-115 min, 22%-32% (v:v); 115 min-120 min, 32%-80% (v:v), 其中洗脱液浓度为流动相 B 占比。一级质谱参数设置如下: AGC 目标为 4×10^5 , 质量容忍度为 10.00 ppm, 离子最大注入时间为 50 ms, 电荷数为 2-5, 分辨率为 60000 进行数据采集。二级质谱参数设置如下: 分辨率为 30000, AGC 目标为 5×10^4 。其中二级质谱 HCD 碎裂模式下的完整糖肽碰撞能量为 35%, 肽段碰撞能量为 32%, 阶梯碰撞能量为 $\pm 5.00\%$, 离子最大注入时间设为 70 ms。

2.3 实验数据处理及可视化

2.3.1 蛋白质和糖蛋白质的数据处理

肽段质谱图鉴定使用 MaxQuant 1.6.1.0 软件以及 Homo sapiens UniProt 数据库(2019 年 6 月更新) 搜索数据, 此数据库包括 20432 种蛋白质^[56]。软件搜库参数设置如下: 添加 “TMT” 标记方法; 还原模式下使用正反库方法鉴定肽段; 碎片离子的质量容忍度为 0.02 Da, 母离子容忍度为 10.0 ppm; 酶切方法为 Trypsin/P (赖氨酸或精氨酸的羧基端允许在脯氨酸处切割); 半胱氨酸的碘乙酰胺化被设置为固定修饰; 完整糖肽的酶切后遗漏位点的最大数量为两个, 肽段的酶切后遗漏位点的最大数量为一个; 肽段匹配谱图 (Peptide spectrum match, PSM) 的错误发现率 (False discovery rate, FDR) 均为 1%。其余搜库条件未另作说明, 则默认参数不改变。对肽段搜库数据进行筛选: 将名称中有 “CON-” 及 “REV-” 的肽段删除; Trypsin 酶切遗漏位点的最大数量为一个; 使用肽段强度 (Intensity) 数据进行定量分析, 因此其强度值需大于零。接着对样品一次实验的筛选数据进行归一化处理: 计算 TMT131 通道 (Mix) 相同肽段强度的中位数, 并取 Log 值; PSM 在其余通道的强度值均取 Log; 每个 PSM 的相对强度值等于其强度 Log 值与其在 Mix 中的相同肽段 Log 值之差; 独特肽段 (Unique peptides) 强度值为相同肽段 PSM 的相对强度值的中位数。最后对所有样品的实验数据进行归一化处理: 首先将所有样品的 Mix 中相同肽段的 Log 值求中位数, 然后与一次实验的相同独特肽段求和, 得样品组

间归一化数据。将归一化数据导入 Perseus 软件，每种肽段在所有样品中被鉴定到的次数占样品总数的百分比应大于 70%。接着将数据空缺处填补替换值，进行列归一化后即得可用于定量分析的高置信度数据。

糖肽质谱图鉴定使用 GPQuest 2.017^[57]软件和包含 45491 种糖肽及 252 种 N-聚糖的数据库^[74]进行搜索处理。糖肽数据筛选方法与肽段数据筛选方法有区别，其余数据处理方法基本一致。区别如下：每个 PSM 的 FDR 小于 1%；每种去糖肽至少有两个 PSM。

2.3.2 实验数据可视化处理

对本研究筛选数据进行分析处理，主要涉及的工具如下：

(1) 使用 Excel (2021) 软件对数据进行整理、统计及可视化，对鉴定到的蛋白质、糖蛋白以及完整糖肽进行定性分析，以获取其数量及种类等信息。依据所有样本的组间归一化后数据进行定量分析。对筛选的多肽及完整糖肽在所有样品中被鉴定到的次数应大于样品总数的 70%，其中缺失值不包括在定量分析中（18 对样品各 2 次平行质谱实验，共 72 次实验中，51 次实验及以上能够同时鉴定到）。同时对数据进行双尾学生 T-成对检验，以确定多肽、完整糖肽及糖蛋白质在癌症组织与非癌性邻近组织（Non-cancerous adjacent tissues, NATs）和不同预后情况之间的差异表达是否符合统计学意义，筛选应满足差异倍数（Fold change, FC） >1.5 或 $FC < 0.67$ ，P 值（p-Value） < 0.05 视为差异显著。

(2) 针对差异表达蛋白利用 DAVID 在线网页工具^[75-76]进行蛋白功能注释与通路富集分析，包括基因本体论（Gene ontology, GO）和京都基因及基因组百科全书（Kyoto Encyclopedia of Genes and Genomes, KEGG）分析。GO 分析包括生物过程（Biological process, BP）、细胞组成（Cellular component, CC）与分子功能（Molecular function, MF）。数据筛选标准为 $FDR \leq 0.05$ ，P 值 ≤ 0.05 。

(3) 通过 GraphPad Prism (8.4.3 版本) 绘制多肽与完整糖肽在癌症组织及非癌性邻近组织的差异倍数值的频率分布直方图，目的是直观显示出样本的频率分布规律。R 语言（pheatmap 包，ComplexHeatmap 包，Mfuzz 包）分析癌症组织、非癌性邻近组织及不同预后情况差异表达的多肽、糖蛋白质及完整糖肽的聚类分析（FC 值大于 1.5 或小于 0.67）。线上的 OmicStudio (<https://www.omicstudio.cn/tool>) 数据分析工具对蛋白质组和糖蛋白质组数据进行差异表达分析（FC 值大于 1.5 或小于 0.67，P 值小于 0.05）绘制火山图用来将数据可视化，方便观察数据整体分布情况，并标记具有显著差异的多肽或完整糖肽。使用线上 bioinformatics (<https://www.bioinformatics.com.cn/>) 数据分析工具进行主成分分析，目的是通过蛋白或完整糖肽在不同样本的表达量差异进行区分。利用 Adobe Illustrator (2021) 对数据分析工具分析图进行进一步绘制修改。

第三章 结果与讨论

3.1 组织采集及样品质控

3.1.1 患者组织样本采集及相关临床信息

本研究为阐明三阴性乳腺癌患者的癌症组织与非癌性邻近组织的蛋白质及糖基化修饰和组成，同时分析其在预后良好情况和预后不良情况两种预后类型的表征及差异，最终挑选了 18 位三阴性乳腺癌患者进行实验，其中 9 位患者预后情况良好，9 位患者预后情况不良（预后不良患者手术后均出现复发情况）。每位患者均采集原发性癌症组织和其配对的非癌性邻近组织。

使用苏木精-伊红染色法（Hematoxylin-Eosin staining, HE）对患者肿瘤组织切片按照病理指标检验其是否属于乳腺癌，然后通过免疫组织化学法（Immunohistochemistry, IHC）检测 ER、PR、HER-2 以及 Ki67 的表达水平，并选取 ER、PR 以及 HER-2 表达呈阴性（低表达）的样本，患者组织样本中 ER、PR 与 HER-2 表达均为阴性即视为合格样本，可进行后续实验。其中 Ki67 是一种癌症生物标志物，用来评价乳腺癌细胞增殖水平。采集患者临床信息见表 3-1，因患者 P2-P9 和 P18 共 9 位患者术后未出现复发情况，则这 9 位患者的预后情况为良好，然而其余 9 位患者经过摘除手术后癌症复发，则将其定义为预后不良情况。所有患者均为女性，且癌症等级、淋巴结转移及癌症分子亚型信息如表所示。

表 3-1 18 位三阴性乳腺癌患者的临床信息

Table 3-1 Summary of clinical information on the eighteen patients

患者编号	年龄	无病生存期	癌症组织分期	分子亚型
P2	51	无	T1, N0	IM
P3	63	无	T1, N1	LAR
P4	53	无	T1, N1	IM
P5	47	无	T2, N1	IM
P6	61	无	T2, N3	IM
P7	50	无	T2, N1	IM
P8	59	无	T1, N1	IM
P9	58	无	T3, N0	LAR
P18	55	无	T2, N3	LAR
P10	49	6	T2, N3	IM
P11	40	13	T2, N1	BLIS
P12	56	16	T2, N2	IM
P13	47	20	T2, N0	IM
P14	63	28	T1, N0	LAR
P15	43	21	T2, N0	BLIS
P16	56	13	T2, N3	IM
P17	61	18	T3, N2	LAR

表 3-1 18 位三阴性乳腺癌患者的临床信息（续）

Table 3-1 Summary of clinical information on the eighteen patients (continued)

患者编号	年龄	无病生存期	癌症组织分期	分子亚型
P1	67	45	T2, N1	IM

注：无病生存期表示患者从手术切除到局部复发的时间（月）。癌症组织分期中 T 指癌症等级，根据病灶大小以及浸润深度的不同可分为 T1、T2 和 T3 等；N 指根据是否有淋巴结转移以及淋巴结转移的数目等情况，可分为 N0、N1、N2 和 N3。

三阴性乳腺癌组织切片的 HE 染色结果显示（如图 3-1a），与患者非癌性邻近组织相比，癌症组织细胞多为无规律且不规则形状，类似片状排列，细胞核较大且多数处于有丝分裂时期，对比非癌性邻近组织腺管形态规则，分化程度低。将患者癌症组织进行 ER、PR 和 HER-2 的免疫组化染色，结果显示肿瘤细胞均未被染色，表明其相关蛋白表达量较低，均呈阴性（如图 3-1b, c, d）。

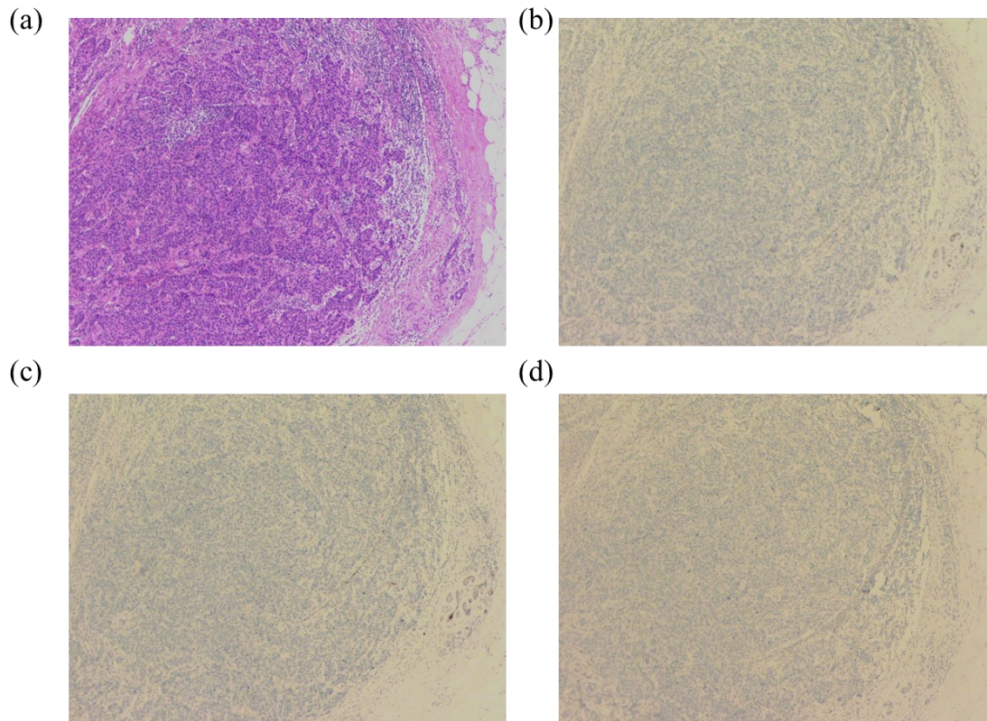


图 3-1 三阴性乳腺癌组织样品的确认

Figure 3-1 Confirmation of triple-negative breast cancer tissues

图注：上图为三阴性乳腺癌患者组织样品的 HE 染色图 (a) 和 IHC 染色图 (b-d)，分别检测 ER (b)、PR (c) 和 HER-2 (d) 表达状态。

据报道，核蛋白质 Ki67 通常用来评估免疫组织化学染色中患者乳腺癌组织细胞的增殖程度，其高表达量通常与乳腺癌患者较高的复发率及较差的生存率相关^[77]。与其他乳腺癌亚型组织相比，三阴性乳腺癌组织的蛋白 Ki67 表达量最高，平均有 $50.9 \pm 23.7\%$ 的细胞核被染色^[78]。因此我们对三阴性乳腺癌组织样品进行抗 Ki67 的免疫组化染色实验，与不同程度 Ki67 的免疫组化结果相比^[78]，在癌症组织中其表达量较高（如图 3-2）。根据以上实验所得结论，可以筛选出参与后续实验的合格样本。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/808046072003006042>