

·基础研究·

空肠弯曲杆菌诱导的结直肠癌的转录组学分析

李 燕, 毛玉丹, 张幸鼎, 牟相宇
(中山大学医学院, 广东 深圳 518107)

摘 要:【目的】探究空肠弯曲杆菌(*C. jejuni*)的致癌机制。【方法】将18只雌性的C57BL/6 *Apc*^{Min/+}小鼠随机分为*C. jejuni*感染组9只和空白对照组9只。采用灌胃*C. jejuni*和葡聚糖硫酸盐共同诱导*Apc*^{Min/+}小鼠产生结直肠癌。实验终点时,取各组小鼠结直肠组织解剖计数肿瘤个数,分别提取*C. jejuni*组4只小鼠和空白对照组3只小鼠的结直肠肿瘤和癌旁组织RNA,采用RNA-seq技术进行转录组学测序,筛选出差异表达基因后进行GO富集分析和KEGG通路富集分析。【结果】与对照组相比,*C. jejuni*感染组肿瘤发生率显著增高($P < 0.01$),显示造模成功。RNA-seq测序结果显示,*C. jejuni*感染组的肿瘤组织跟两个对照组相比,分别有394个和501个差异表达基因(变化 >4 倍; $P < 0.05$)。GO富集分析差异基因主要富集在多种免疫反应的调控和激活途径、多种蛋白质转运和受体结合途径。KEGG通路富集分析癌症相关通路和代谢途径较显著。上述差异表达基因中,有17个基因在两次对照中均有出现,经过筛选,得到14个“核心”差异表达基因。在进一步的qRT-PCR验证中,有9个基因表达差异显著,其中有3个基因(*Gm1987*、*Saxo1*和*Plekhs1*)表达上调,6个基因(*Lrp2*、*Serpina3c*、*Fabp4*、*Tmem52*、*Lrrn4*和*Upk3b*)表达下调。【结论】本研究指出了9个可能在*C. jejuni*诱导的结直肠癌的发生发展中起到了重要且独特作用的宿主基因,为后续研究*C. jejuni*导致结直肠癌的机制提供了新的方向和思路。

关键词:空肠弯曲杆菌;结直肠癌;转录组

中图分类号:Q78

文献标志码:A

文章编号:1672-3554(2022)04-0548-15

DOI: 10.13471/j.cnki.j.sun.yat-sen.univ(med.sci).2022.0405

Transcriptome Analysis of *Campylobacter jejuni*-Induced Colorectal Cancer

LI Yan, MAO Yu-dan, ZHANG Xing-ding, MOU Xiang-yu

(School of Medicine, Sun Yat-sen University, Shenzhen 518107, China)

Correspondence to: ZHANG Xing-ding; E-mail: zhangxd39@ms.sysu.edu.cn; MOU Xiang-yu;

E-mail: mouxy5@ms.sysu.edu.cn

Abstract:【Objective】To explore the carcinogenic mechanism of *Campylobacter jejuni*.【Methods】Eighteen female C57BL/6 *Apc*^{Min/+} mice were randomly divided into the *C. jejuni*-infected group and the non-infection control group, each group with nine mice. Colorectal cancer of *Apc*^{Min/+} mice was induced by dextran sulfate sodium and gavage of *C. jejuni* (or PBS as a control). At the end of the experiment, the number of tumors in colorectal tissues of mice in each group was counted, and RNA was extracted from colorectal tumors, along with para-cancer tissues as controls. Transcriptome sequencing was performed by RNA-Seq technology, and data were analyzed for differentially expressed genes (DEGs). Further, selected DEGs were subjected to GO (gene ontology) enrichment analysis and KEGG pathway enrichment analysis.【Results】Compared with that of the non-infection control group, the incidence of tumor in *C. jejuni*-infected group was significantly higher ($P < 0.01$), which indicated the success of recreating the *C. jejuni*-induced CRC model. RNA-seq results showed

收稿日期:2022-04-29

基金项目:国家自然科学基金(32000096)

作者简介:李燕,硕士生,研究方向:分子医学,E-mail:liyan298@mail2.sysu.edu.cn;张幸鼎,通信作者,E-mail:zhangxd39@ms.sysu.edu.cn;牟相宇,通信作者,E-mail:mouxy5@ms.sysu.edu.cn

that there were 394 and 501 DEGs (fold change > 4 and $P < 0.05$) in the *C. jejuni*-infected group compared with the two control groups, respectively. In GO enrichment analysis, DEGs were mainly enriched in immune response regulation and activation pathways, multiple protein transport pathways and receptor binding pathways. Cancer-related pathways and metabolic pathways were significant enriched in KEGG pathway enrichment analysis. Among these DEGs, 17 genes were found in comparisons with both control groups. The 17 genes were further selected, resulting in 14 “core” DEGs. In further validation of qRT-PCR, 9 genes were significantly differentially expressed, among which 3 genes were up-regulated (*Gm1987*, *Saxo1* and *Plekhs1*) and 6 were down-regulated (*Lrp2*, *Serpina3c*, *Fabp4*, *Tmem52*, *Lrrn4* and *Upk3b*).【Conclusion】 This study emphasizes 9 host genes that may play important and unique roles in the occurrence and development of colorectal cancer induced by *C. jejuni*, which provides new insights for further studies on the carcinogenic mechanism of *C. jejuni*.

Key words: *Campylobacter jejuni*; colorectal cancer; transcriptomics

[J SUN Yat-sen Univ (Med Sci), 2022, 43(4): 548-562]

结直肠癌(colorectal cancer, CRC)发病率在癌症中居第三位,其发生与肠道微生物有着密切关系。已有研究发现,拟杆菌属(*Bacteroides*)、梭杆菌属(*Fusobacterium*)、沙门氏菌属(*Salmonella*)、埃希氏杆菌属(*Escherichia*)和弯曲杆菌属(*Campylobacter*)等多种肠道病原体与结肠癌的发生发展密切相关^[1-4]。空肠弯曲杆菌(*Campylobacter jejuni*, *C. jejuni*)是一种带荚膜的革兰氏阴性杆菌,是常见的人畜共患食源性病原菌之一。人类感染 *C. jejuni* 后,会产生肠胃炎、腹泻、腹部绞痛以及发热的症状,并可能会伴有反应性关节炎、肝炎、格林-巴利综合征和瑞特氏病等免疫性损伤性疾病^[5-6]。*C. jejuni* 81-176会通过产生细胞致死性膨胀毒素进而诱导DNA损伤,在小鼠模型中促进结直肠癌的发生^[7]。因此,*C. jejuni*感染不仅会导致患者产生肠胃炎,还有增加患者罹患结直肠癌风险的可能。然而,在 *C. jejuni* 诱导的结直肠癌中,宿主基因是如何参与癌症发生发展的,这一点仍不清楚。转录组学测序(RNA sequencing, RNA-seq)技术可以从整体水平上反映细胞中基因转录情况^[8]。为了更好的了解空肠弯曲菌诱导的结直肠癌中的基因表达情况,本研究使用了 *C. jejuni* 诱导的 *Apc^{Min/+}* 小鼠结直肠癌模型,在试验终点分离小鼠结直肠肿瘤组织以及癌旁组织,分别提取组织RNA后进行转录组学测序分析,探索、筛选出差异基因,以提高对 *C. jejuni* 致癌的作用机制的认识,为结直肠癌的防治提供新的探索方向。

1 材料与方法

1.1 *C. jejuni* 81 - 176 的培养

人类临床分离株 *C. jejuni* 81 - 176 用弯曲杆菌属选择性培养基在 37 °C 下微需氧条件下培养 48 h。

1.2 动物实验

本实验共选用 18 只体质量约为 (18.0 ± 3.0) g, 雌性 C57 BL/6 *Apc^{Min/+}* 小鼠 (5-8 周龄), 小鼠购买自江苏集萃药康公司, 于中国科学院广州生物医药与健康研究院动物房 SPF 级环境中分笼饲养, 实验经过中国科学院广州生物医药与健康研究院实验动物福利与伦理委员会批准 (伦理编号: N2022003)。随机分为两组, 计为 Mock (空白对照组, $n=9$) 和 Campy (*C. jejuni* 处理组, $n=9$)。分别放入 4 笼小鼠笼中 (每笼 4~5 只)。

检疫合格的小鼠在动物房观察一周后, 进行打耳标记。为了使菌在肠道内更容易定植, 第 1~7 天, 给小鼠饮用含有 4 种混合抗生素的饮用水 (100 mg/L 万古霉素、200 mg/L 甲硝唑、200 mg/L 氨苄青霉素和 200 mg/L 新霉素)。抗生素处理后, 第 8 天以 1×10^8 CFU/只的量给 Campy 组小鼠灌胃 *C. jejuni*, Mock 组以等量 PBS 作为对照灌胃。随后, 在第 15~20 天, 在饮用水中加入 25 g/L 的葡聚糖硫酸钠 (DSS)。在第 60 天, 安乐死小鼠, 并进行实验材料取材。关于肿瘤部位与癌旁组织的取材实验时是以肉眼可见的隆起为肿瘤组织部位, 以肿瘤中心为圆心距离 3 mm 的地方为癌旁, 癌旁组织所取的

面积约为0.3 cm×0.3 cm。肠道组织 HE 染色由赛维尔生物科技有限公司完成。因 Mock 组一只小鼠在予以 DSS 水处理期间发生死亡,本文选用 Mock 组 3 只小鼠和 Campy 组 4 只小鼠的肿瘤和癌旁部位进行转录组测序。

1.3 RNA 提取及文库构建与高通量测序

分别取对照组(3个生物学重复)和实验组小鼠(4个生物学重复)的结肠肿瘤和癌旁组织,放入装有 1 mL RNA later 的 2 mL 离心管中,液氮速冻后转移至-80 °C 超低温冰箱保存。提取 RNA 时,先向 2 mL EP 管中加入 1.5 mL TRIzol 裂解液,后取适量组织于液氮保护下将组织样品快速研磨成粉末,转入裂解液,使组织细胞充分裂解。按 Invitrogen TRIzol 试剂盒(赛默飞有限公司)说明分别提取组织的总 RNA,Fragment analyzer(安捷伦生物有限公司)对 RNA 的质量和浓度进行评估,质量检测合格(28 S/18 S > 1.0, RQN 值 > 7.0)的样本用 Promega 反转录试剂盒(普洛麦格生物技术有限公司)进行反转录(操作步骤按照说明书进行),按照 SMART cDNA 文库构建试剂盒(宝生物工程有限公)说明书分别对样本进行文库构建。构建好的文库在 BGISEQ 测序平台(华大基因,中国深圳)进行转录组测序。

1.4 测序数据过滤及参考基因组比对

本实验委托华大基因进行,具体步骤如下:利用过滤软件 SOA Pnake 软件^[9]过滤低质量 reads,即去除测序片段中包含接头的 reads、去除未知碱基 N 含量大于 5% 的 reads、去除低质量(质量值低于 15 的碱基占该 reads 总碱基数的比例大于 20%) 的 reads,筛选获得 clean reads,利用 HISAT2 2.0.4 软件^[10]将 clean reads 比对到参考基因组序列,使用 Bowtie2 软件^[11]将 clean reads 比对到参考基因序列上得到比对结果。比对完,通过统计比对率、reads 在参考序列上的分布情况等,判断比对结果是否通过第二次质控(QC of alignment)。若通过,则进行后续的数据分析。使用 RSEM 软件^[12]计算各个样品的基因表达水平。

1.5 生物信息学分析

进行基因定量分析、基于基因表达水平的各项分析(主成分、相关性、差异基因筛选等等),并对筛选出的样品间差异表达基因通过 DAVID 平台进行 gene ontology(GO)富集分析和 KEGG 通路显著性富集分析等更深入的挖掘分析。

参考物种信息如下,物种名:Mus_musculu,参考基因组版本 GCF_000001635.26_GRCh38.p6 (NCBI)。

1.6 qRT-PCR 验证

根据反转录试剂盒(宝生物工程有限公)进行反转录获取 cDNA,用 qPCR 验证筛选出的差异表达基因。反应条件为 95 °C 30 s,95 °C 10 s,60 °C 30 s,40 个循环。采用 2^{-ΔΔCt} 法分析 mRNA 表达水平。引物由擎科生物科技有限公司合成,详见二维码 1。



1.7 统计学处理

使用 GraphPad Prism 8.0 和 SPSS 25.0 软件处理。小鼠肠道肿瘤的数目的统计学差异比较经 K-S 检验数据正态分布和单因素 ANOVA 分析中的方差齐性检验方差齐性后,使用双尾未配对 *t* 检验计算。差异基因的筛选和后续分析基于 R 软件中的 DESeq2 程序包进行分析,利用 Benjamini 法校正后的 *P* 值以控制错误发现率,校正后的 *P* 值 < 0.05 以及 |log₂fold change| > 2 作为显著差异表达的阈值。

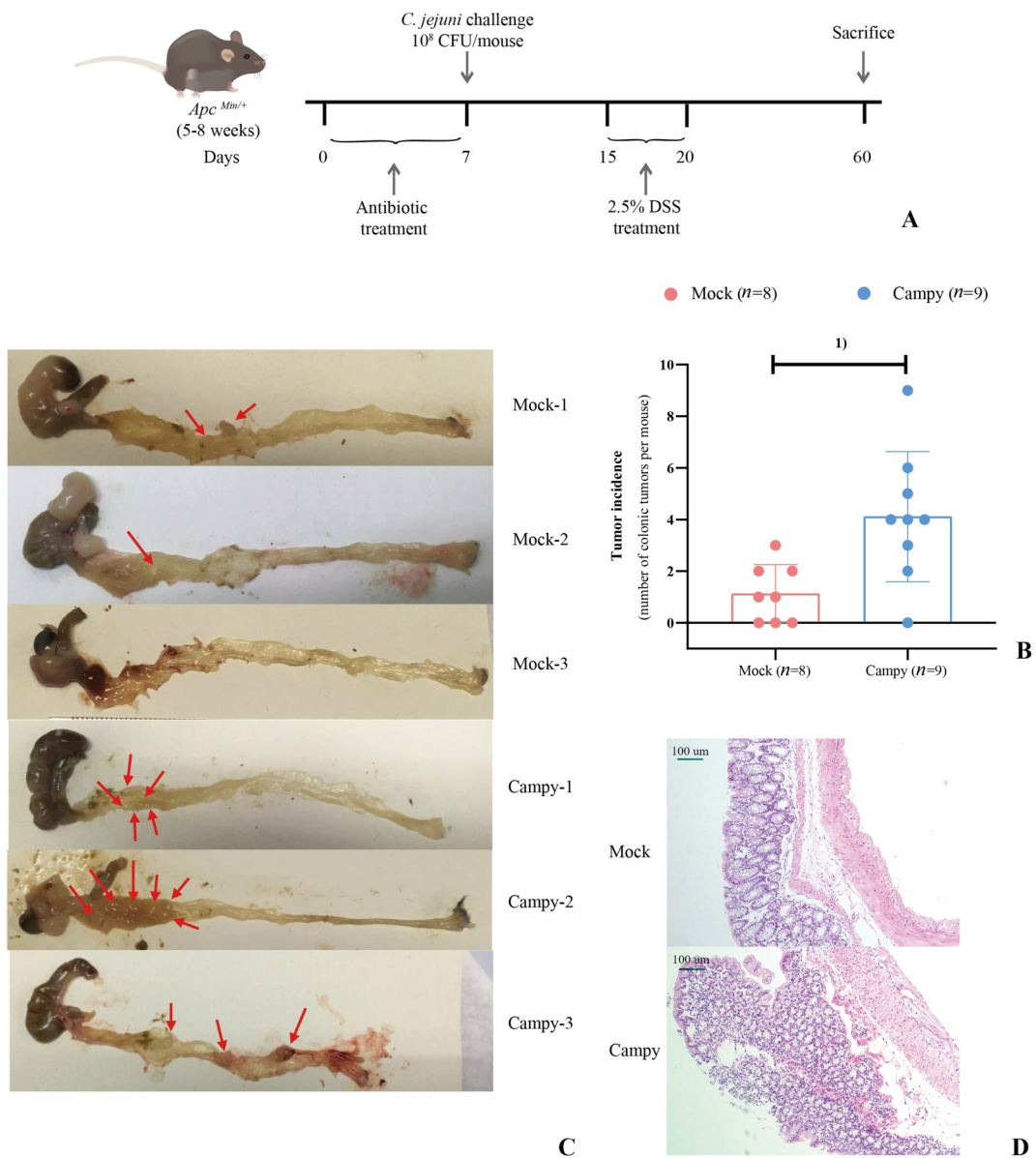
2 结果

2.1 C. jejuni 诱导小鼠结直肠癌造模情况

如图 1 所示,Campy 组小鼠的成瘤数量比 Mock 组小鼠的成瘤数显著性增多,差异具有统计学意义,统计数据两组均符合数据正态分布和方差齐性,经独立样本 *t* 检验, ($t = -3.079, P = 0.008 < 0.05$)。D 图展示的肠道 HE 染色结果,Campy 组的肠道黏膜层增厚,中央见坏死区,上皮细胞边缘不清,细胞核浆比增大,见不少核染色深,核仁明显,有见核分裂象。Mock 组相较之下结肠腺体排列规则,细胞核比例大小正常。

2.2 转录组测序数据质量情况

2.2.1 测序数据质控概况 每个实验组小鼠的结直肠肿瘤以及相对应的癌旁组织样本进行转录组分析,本项目使用 BGISEQ 平台一共测了 14 个样



A: Schematic diagram of animal experimental model. B: Tumor incidence in the colon. Each dot represents one individual animal. Bars represent the geometric mean \pm SD. *P* values were calculated by unpaired, two-tailed Student's *t* test on log-transformed data. 1) *P* < 0.01 indicate significant differences compared to the Mock group. Specific n numbers are indicated in the figure. C: Representative macroscopic morphologies of mice in Mock and Campy groups. D: HE-stained colon sections of mice in Mock and Campy groups.

图1 *C. jejuni* 促进小鼠肿瘤发生

Fig. 1 *C. jejuni* promotes tumorigenesis in mice

品,每个样品平均产出6.67G数据。样品比对基因组的平均比率为90.89%,比对基因的平均比率为68.71%,一共检测到18 964个基因。测序质量值大于30(Q30)的碱基占有所有碱基的89.52%以上,测序比对效率均值约在90.25%。测序数据质量良好,满足后续分析要求(表1)。

2.2.2 基因表达水平分布 经过预处理和过滤后,基因表达量用FPKM(Fragments Per Kilobase of ex-

on model per Million mapped fragments)值归一化表示,即每千个碱基的转录每百万映射读取的碎片大小,是对基因长度和测序深度进行校正后的值。本研究使用箱线图展示各样品基因表达水平的分布情况,评估各样本中基因表达数据是否具有对称性,分布的分散程度等信息。如图2所示,基因表达于各样本中差异不大,分散程度相似。

表1 小鼠结直肠样本RNA-seq测序数据质量

Table 1 Quality of RNA-seq data from mouse colorectal samples

Sample name	Total Raw Reads	Total Clean Reads	Total Clean Bases	Clean Reads Q30	Clean Reads Ratio
Mock-para-1	49.08	44.89	6.73	90.58	91.46
Mock-para-2	47.33	43.83	6.57	92.07	92.62
Mock-para-3	49.08	44.06	6.61	90.36	89.77
Mock-tumor-1	49.08	45.07	6.76	90.61	91.83
Mock-tumor-2	48.41	42.96	6.44	89.79	88.75
Mock-tumor-3	49.08	44.25	6.64	90.52	90.15
Campy-para-1	49.08	44.58	6.69	90.45	90.84
Campy-para-2	49.08	44.66	6.70	90.73	90.99
Campy-para-3	52.59	44.60	6.69	90.19	84.82
Campy-para-4	47.33	43.85	6.58	92.00	92.66
Campy-tumor-1	49.08	44.83	6.72	89.52	91.33
Campy-tumor-2	50.83	45.34	6.80	89.54	89.20
Campy-tumor-3	49.08	44.35	6.65	89.48	90.37
Campy-tumor-4	50.83	45.12	6.77	92.14	88.75

Q30: The percentage of bases with a quality greater than 30 in the total base of reads after filtration; Mock-tumor: the tumor of the control group; Mock-para: the para-tumor tissue of the control group; Campy-tumor: the tumor of the *C. jejuni* challenge group; Campy-para: the para-tumor tissue of the *C. jejuni* challenge group.

2.3 基因表达差异分析

经DEseq2分析,本试验在筛选的过程中,将差异倍数(fold change, FC)的绝对值 > 4 ,即 $\log_2FCI > 2$, P 值 < 0.05 作为筛选标准。通过热图可以发现,经过聚类,同一组内表达相似的基因聚类后几乎都出现在同一簇中。其中,FC表示的是两个组之间基因表达量的比值。差异表达基因(differentially expressed genes, DEGs)分布情况用火山图表示。如图3所示,通过测序结果显示,比较Campy组肿瘤和癌旁组织,即Campy-tumor与Campy-para,共发现394个显著差异基因,其中341个基因上调,53个基因下调。比较Campy-tumor与Mock-tumor,共有501个显著差异基因,其中126个基因表达上调,375个基因表达下调。比较Campy-para与Mock-para,共有316个显著差异基因,其中上调基因215个,下调基因101个。

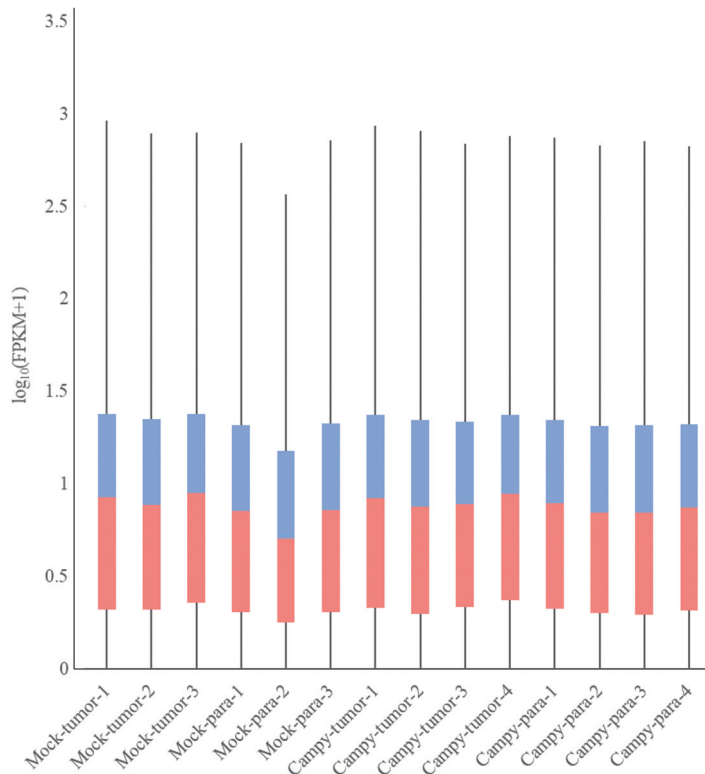
2.4 差异表达基因GO富集分析

对Campy组与Mock组肿瘤与癌旁组织三种比较组筛选出来的显著性差异基因进行GO富集分析,以 $P < 0.05$ 作为筛选标准。GO富集分析主要包括三个部分,分子功能(molecular function, FM)、

生物学过程(biological process, BP)以及细胞组分(cellular component, CC)。通过GO富集分析结果可以看出,对比Campy组肿瘤部位与癌旁部位,如图4-1所示,差异基因中的上调基因在生物学过程模块中,主要富集在肽酶活性的调节、白细胞的迁移、伤口愈合、白细胞细胞-细胞黏附等,分子功能模块中,细胞因子活性、受体配体活性、细胞因子受体结合通路富集程度较高;下调基因在生物学过程中表达量显著性差异较大的基因主要富集在脂肪酸代谢通路,其余还有脂肪细胞分化、棕色脂肪细胞分化通路等。

可以看出,*C. jejuni*作用下的小鼠结直肠癌的发生、发展借助表面受体与细胞外基质的黏附因子特异性结合,进而诱导受体激活,并通过一系列的蛋白水解、跨膜运输等影响肿瘤细胞的生长、侵袭和转移。

对比Campy组与Mock组的肿瘤组织,如图4-2所示,上调基因中除了主要富集在膜、刷状缘相关的细胞组分通路外,在生物学过程中主要富集在阴离子转运、调节其他生物、抗菌肽介导的抗菌体液免疫反应、胆固醇平衡和粘膜固有免疫途径以及分



The boxplot of each region corresponds to five statistics (upper limit, upper quartile, median, lower quartile and lower limit respectively from top to bottom, where the upper limit and lower limit are not considered outliers).

图2 样品基因表达水平分布图

Fig. 2 Distribution of gene expression in various samples

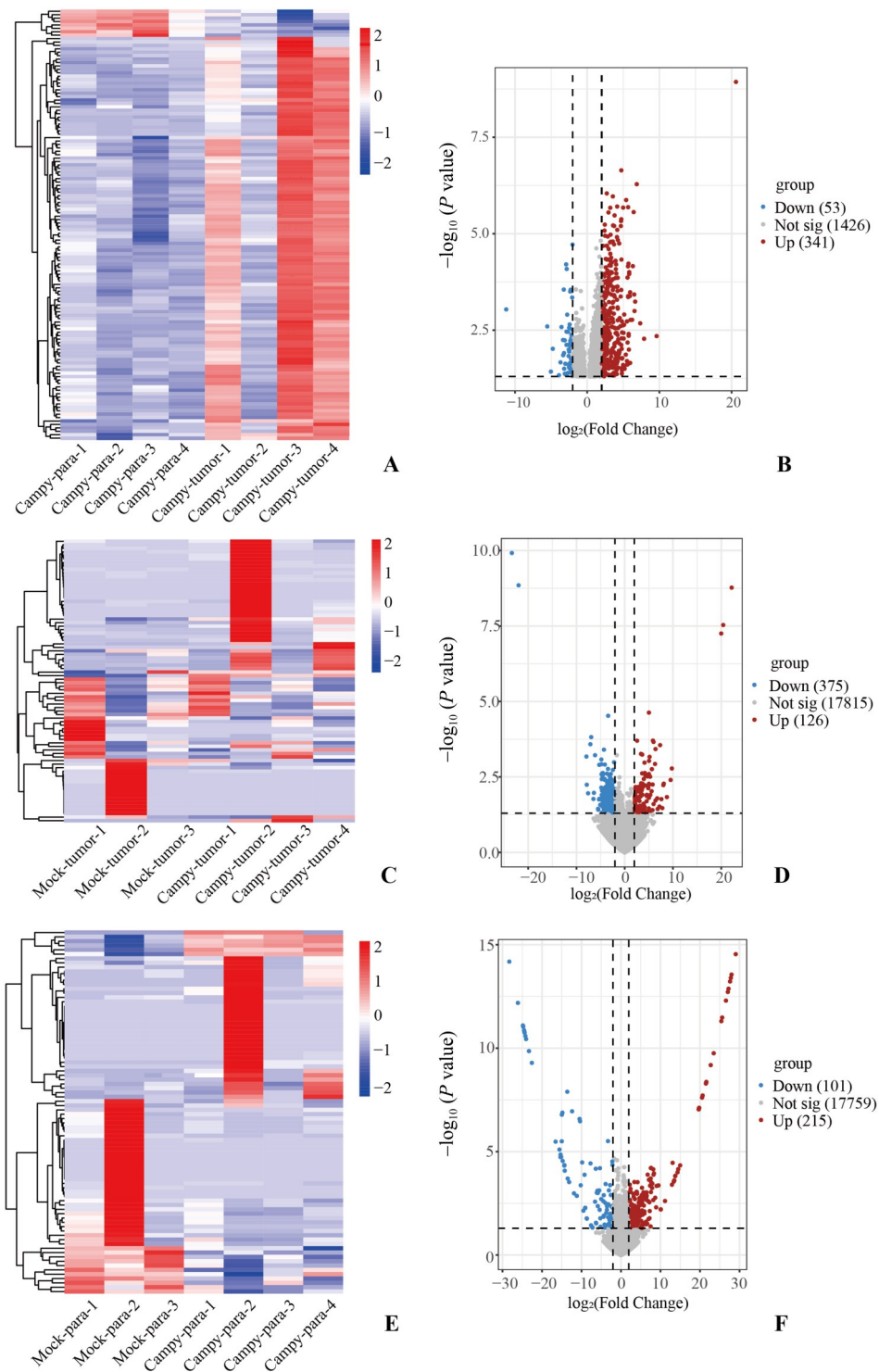
子功能通路的脂质转运蛋白、血红素结合以及跨膜转运蛋白活性等;下调基因中主要富集在生物学过程中的多种免疫相关通路包括单核细胞、淋巴细胞、B细胞、T细胞等的增殖和活化通路调节,分子功能通路上主要富集到的通路是GTP酶、碳水化合物、细胞因子等多种蛋白质的结合和调节通路。

对比 Campy 组和 Mock 组的癌旁部位,如图 4-3 所示,上调基因中显著性差异较大的主要富集在生物学过程中的其他生物膜破裂、抗菌肽介导的抗菌体液免疫反应、抗菌体液反应、防御细菌反应和粘膜固有免疫途径以及分子功能中的多种物质的跨膜转运活性,与 Campy 与 Mock 组的肿瘤部位相似;下调基因中基因表达差异较大的主要富集在生物学过程中的消化途径和分子功能的丝氨酸类肽酶活性和丝氨酸水解酶活性途径。

2.5 KEGG 通路分析

对三种对比组的差异基因进行 KEGG 通路富集分析,判断基因富集的阈值为 $P < 0.05$ 。我们分别挑选了基因上调和下调的富集最显著的前 10 条

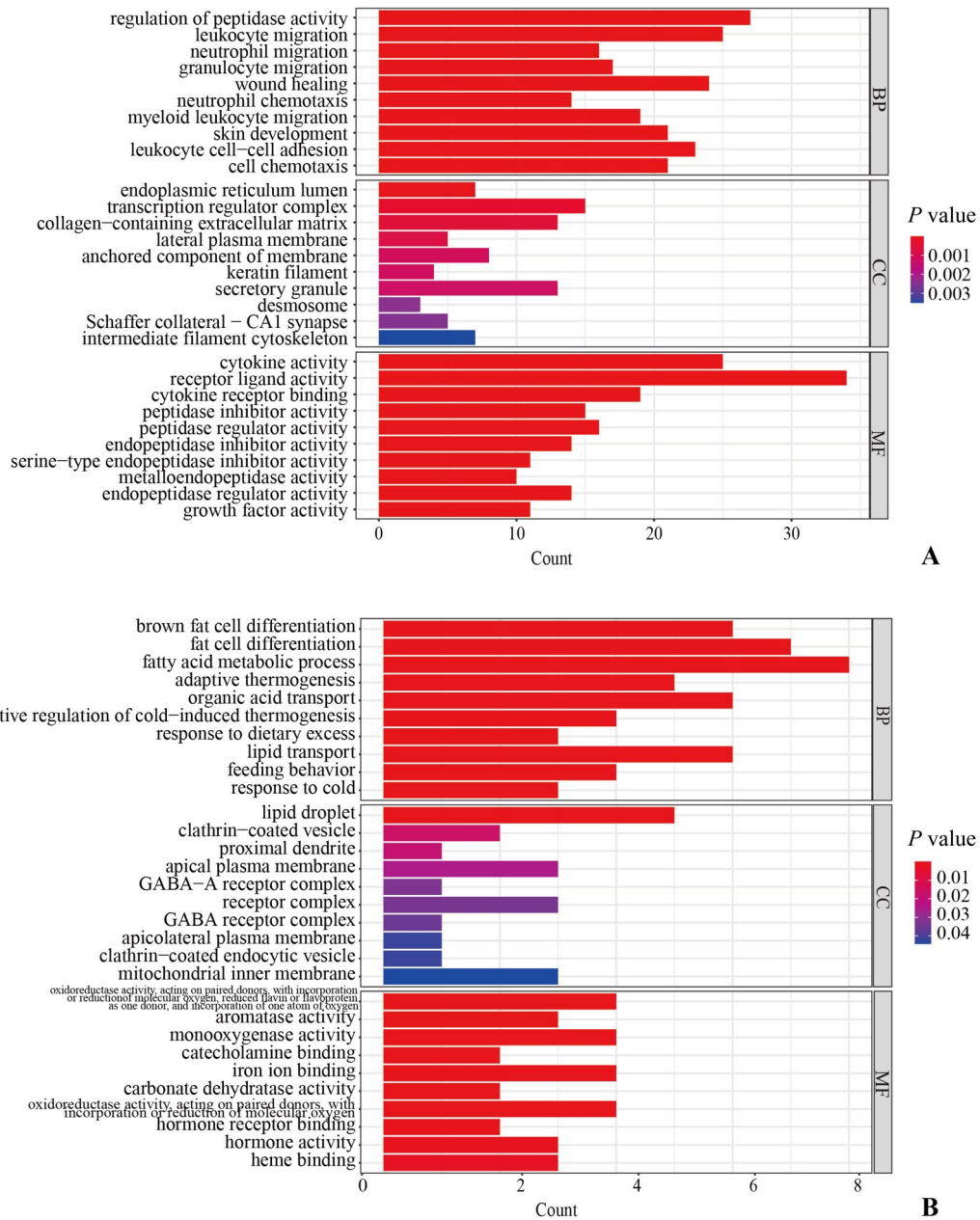
通路进行展示(其中 Campy-tumor 组对比 Campy-para 组的下调基因只富集在 6 条通路),如图 5 所示。对于 Campy-tumor 组和 Campy-para 组,细胞因子和细胞因子受体相互作用通路是上调基因中差异最显著且富集最多的通路,其余还有 IL-17 信号通路和 Wnt 信号通路等;富集最多显著性最高的下调基因通路是 PPAR 信号通路,其次是神经活性配体-受体相互作用。对于 Campy-tumor 组和 Mock-tumor 组,上调基因中富集最多的是代谢途径,富集最显著的是脂肪的消化和吸收;下调基因中富集最多且最显著的是细胞黏附分子,其余还有细胞因子与细胞因子受体的互作、癌症中的通路和 NF- κ B 信号通路等。关于 Campy-para 组和 Mock-para 组,上调基因富集最多的为代谢途径,其余还有癌症中的转录失调、金黄色葡萄球菌感染和 NOD 样受体信号通路等;下调基因富集最多且最显著的是胰腺分泌,其余还有代谢途径、蛋白质的消化和吸收等。



A, C and E: Heat map of cluster of differentially expressed genes between groups (A: Campy-tumor group vs. Campy-para group, C: Campy-tumor group vs. Mock-tumor group, E: Campy-para group vs. Mock-para group); B, D and F: Volcano map of differentially expressed genes between groups (B: Campy-tumor group vs. Campy-para group, D: Campy-tumor group vs. Mock-tumor group, F: Campy-para group vs. Mock-para group). Red spots represent significantly high expression genes with fold change greater than \log_2 , and blue spots represent significantly low expression genes with fold change less than \log_2 (-2). The horizontal dashed line indicates the P value threshold of 0.05.

图3 样本间基因表达差异分析

Fig. 3 Analysis of gene expression differences between samples



A: up-regulated genes; B: down-regulated genes.

图4-1 Campy-tumor 组对比 Campy-para 组差异基因GO富集分析

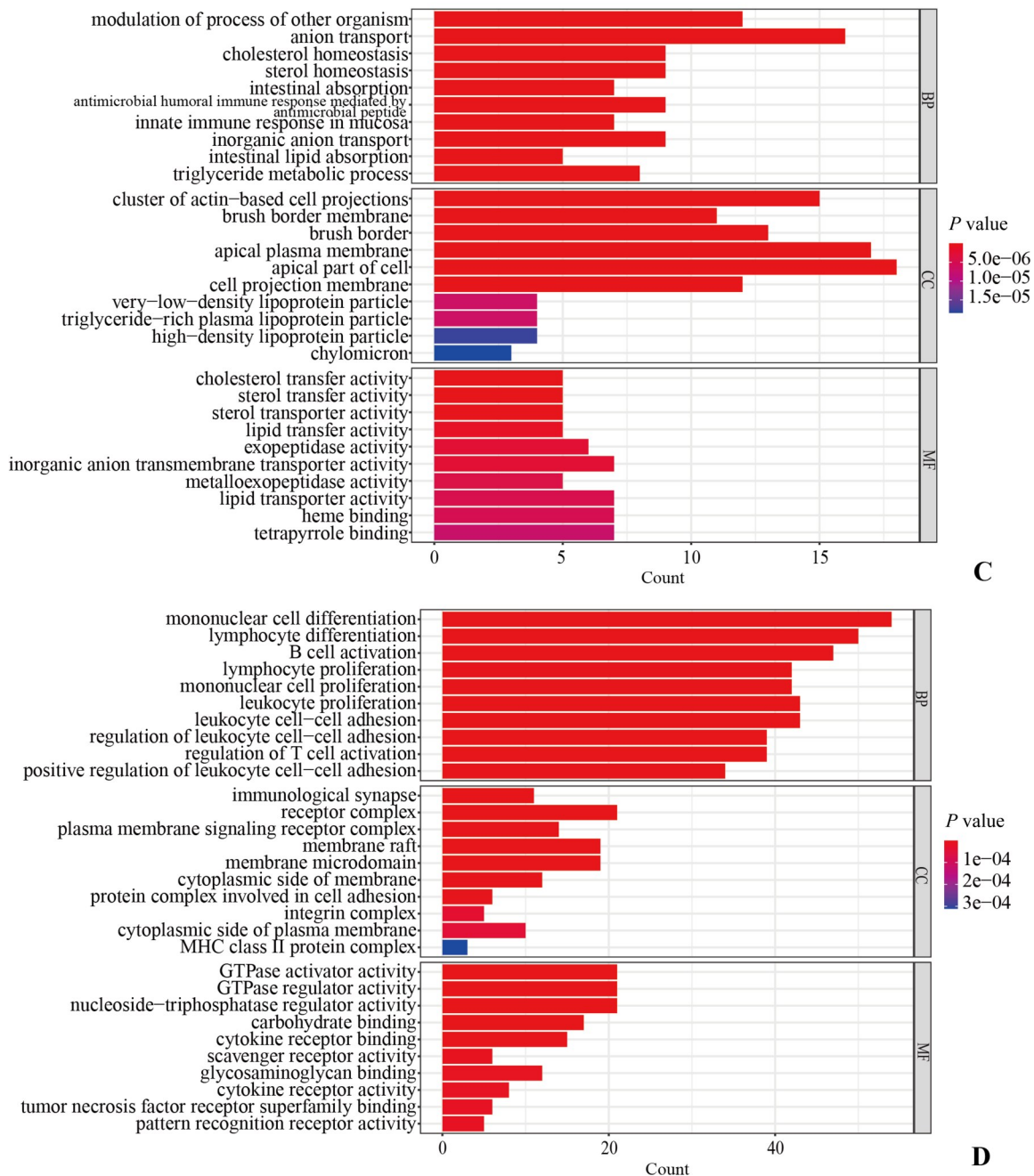
Fig. 4-1 GO enrichment of DEGs from Campy-tumor group vs. Campy-para group

2.6 韦恩图法分析

为了明确 *C. jejuni* 诱导的结直肠癌(Campy-tumor组)的基因表达特点,本研究设计了两个对照,分别为 Campy-para 组和 Mock-tumor 组。前述分析可知, Campy-tumor 与对照组 Campy-para 相比,共发现 394 个差异表达基因;而 Campy-tumor 与另一个对照组 Mock-tumor 相比,共有 501 个差异表达基

因;结果显示,有 17 个差异表达基因(图 6A 红蓝交汇处),在两次对照中均有出现,意味着它们在 *C. jejuni* 诱导的结直肠癌的发生发展可能起到重要作用。

两个对照组之间的比较,即 Mock-tumor 与 Mock-para 相比,得到 1 008 个差异表达基因(黄色区域)。这些基因表达变化是 *Apc* 基因缺失导致的



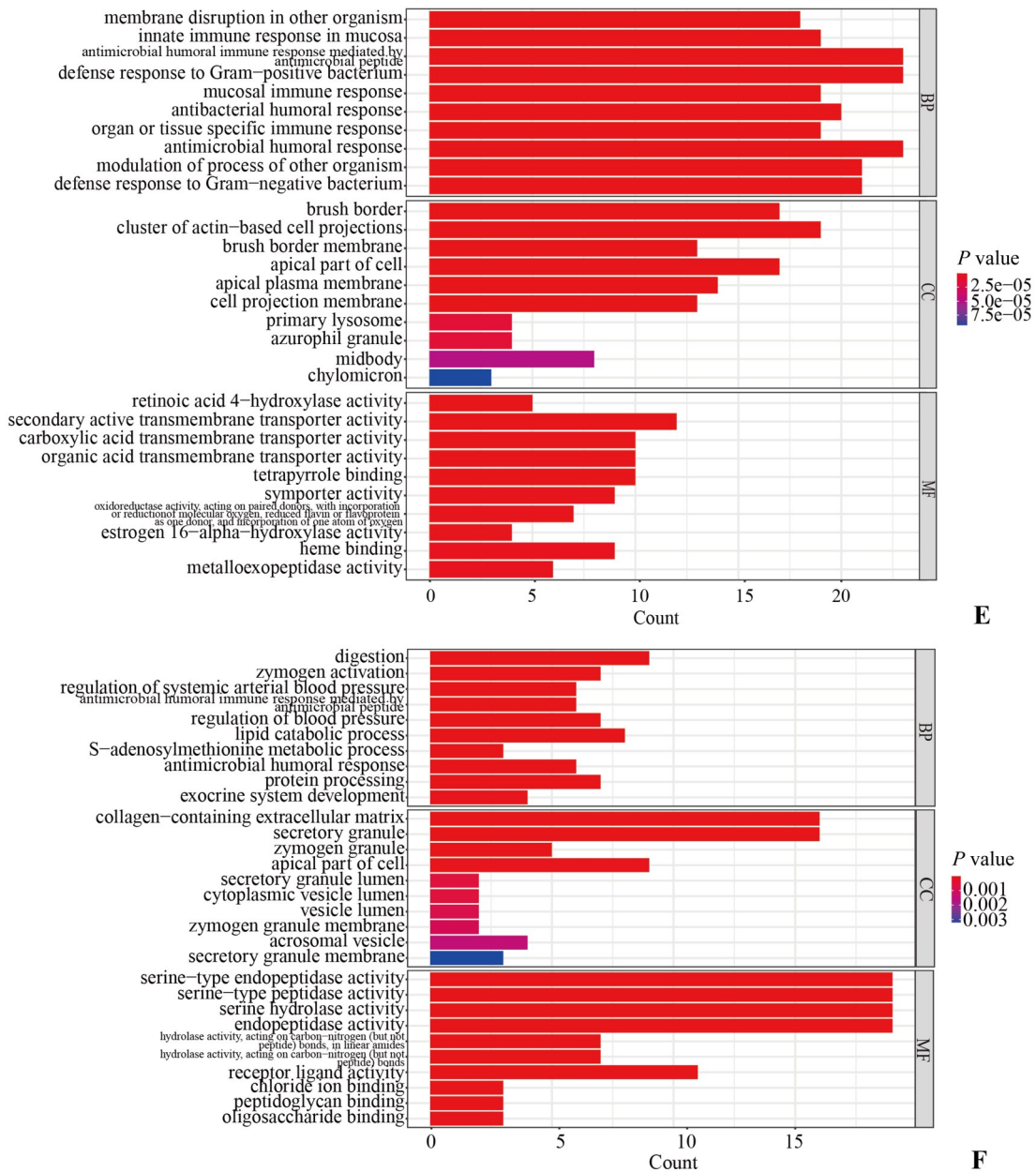
C: up-regulated genes; D: down-regulated genes.

图4-2 Campy-tumor组对比Mock-tumor组差异基因GO富集分析

Fig. 4-2 GO enrichment of DEGs from Campy-tumor group vs. Mock-tumor group

癌变本身即具有的,它们的表达变化不依赖于 *C. jejuni* 的诱导。另一方面, Campy-para 与 Mock-para 相比,得到316个差异表达基因(绿色区域)。这些基因表达变化与 *C. jejuni* 感染有关,但与癌变这一结果不相关。因此,为了得到更加严谨的结果,我们从上述红蓝交汇处的17个基因减去了与黄色区域交汇的1个基因(红蓝黄交汇区)和与绿色区域

交汇的2个基因(红蓝绿交汇区),剩下14个基因。这14个基因表达的变化在 *C. jejuni* 诱导的结直肠癌的发生发展中可能起到了重要且独特的作用,因此我们称其为“核心差异表达基因”。其中,有5个基因(*Lipf*, *Gm1987*, *Ascl5*, *Saxo1* 和 *Plekhs1*)表达上调,9个基因(*Lrp2*, *Phex*, *Serpina3c*, *Fabp4*, *Vwa3a*, *Tmem52*, *Lrrn4*, *Upk3* 和 *Crb2*)表达下调(表2)。其



E: up-regulated genes; F: down-regulated genes.

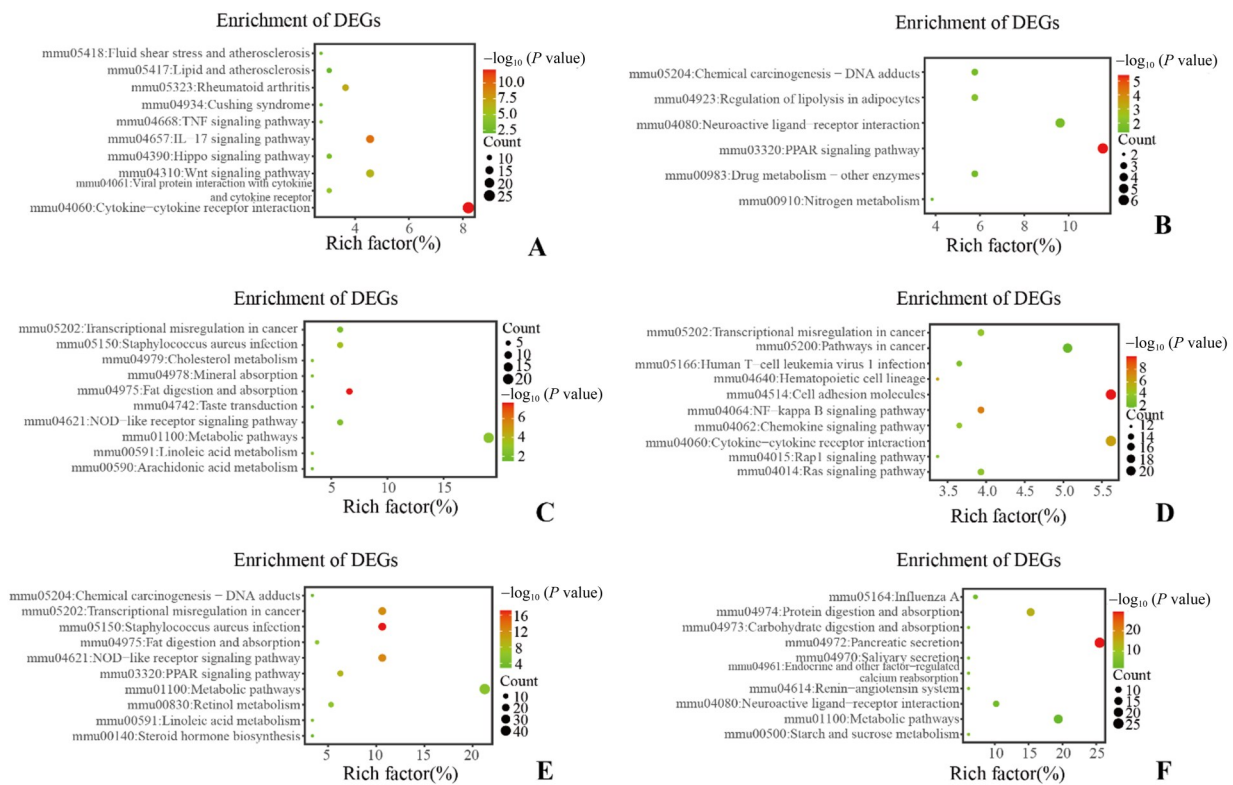
图 4-3 Campy-para 组对比 Mock-para 组差异基因 GO 富集分析

Fig. 4-3 GO enrichment of DEGs from Campy-para group vs. Mock-para group

中表达量差异超过 30 倍的有：*Lipf* (1×10^6 倍)、*Gm1987* (1×10^3 或 1×10^6 倍, 取决于不同的对照组)、*Ascl5* (约 32 倍)、*Phex* (约 32 倍)。

上述 14 个基因的组间表达量聚类图、GO 注释信息和 KEGG 注释信息如图 6(B-F) 所示。可以看到, 在 GO 注释信息里的生物学过程模块里这些差异表达基因主要参与了细胞过程、生物调节、生物过程的调控、对刺激的反应、代谢通路等过程; 在分子功能模块, 主要参与了结合功能; 在细胞组分模

块, 主要与细胞和膜相关。在 KEGG 通路注释信息里, 这些核心差异表达基因主要参与了代谢相关通路和癌症发生相关通路, 代谢相关通路有: 代谢途径、胆固醇代谢、脂肪消化吸收、脂肪细胞里的脂肪分解调节、甲状腺激素的合成和甘油酯类代谢通路, 与癌症相关的通路包括了细胞因子-细胞因子受体互作、趋化因子信号通路、河马信号通路、NF- κ B 信号通路、PPAR 信号通路、刺猬信号通路。



A and B: Campy-tumor group vs. Campy-para group; C and D: Campy-tumor group vs. Mock-tumor group; E and F: Campy-para group vs. Mock-para group; A, C and E: up-regulated genes; B, D and F: down-regulated genes. X-axis is enrichment ratio (the ratio of the number of genes annotated to an item in the selected gene set to the total number of genes annotated to the item in the species, calculated by Rich factor= Term Candidate Gene Num/Term Gene Num, the Y-axis is KEGG Pathway term, the bubble size represents the number of differentially annotated genes on a term, and the color represents the enrichment significance value P value, and the redder the color represents the smaller significance value.

图5 KEGG富集分析

Fig. 5 KEGG enrichment analysis

2.7 qRT-PCR对RNA-seq结果的验证

我们用qRT-PCR对上述14个差异表达基因进行验证。与Campy-para相比, Campy-tumor中 *Lipf*、*Gm1987*、*Ascl5*、*Saxo1*、*Plekhs1* mRNA表达上调, *Lrp2*、*Phex*、*Serpina3c*、*Fabp4*、*Vva3a*、*Tmem52*、*Lrrn4*、*Upk3b*、*Cr2*表达下调,升降趋势与RNA-seq结果完全一致,其中有9个基因(*Gm1987*、*Saxo1*、*Plekhs1*、*Lrp2*、*Serpina3c*、*Fabp4*、*Tmem52*、*Lrrn4*、*Upk3b*)的表达两组间具有显著性差异($P < 0.05$;图7),表明测序结果重复性较好,可信度较高。

3 讨论

已知 *C. jejuni* 对宿主造成疾病主要是依赖运动、黏附、侵袭和产毒素这些毒力因子^[13],其产生的细胞致死性膨胀毒素(cytolethal distending toxin,

CDT)是目前已知的促进结直肠癌发生的关键因素^[7],然而宿主基因是如何参与 *C. jejuni* 诱导癌症发生的,这一点仍不清楚。因此,本研究通过转录组学分析,以揭示在 *C. jejuni* 诱导的肿瘤中表达改变的宿主基因。

*Apc^{Min/+}*小鼠是研究肠道肿瘤发生、发展过程的经典模型^[14],通过给 *Apc^{Min/+}*小鼠灌胃 *C. jejuni* 81-176株细菌诱导小鼠结直肠癌的发生,再用葡聚糖硫酸钠(dextran sulfate sodium, DSS)破坏肠上皮屏障加快建模。通过评估小鼠的成瘤情况,我们发现 *C. jejuni* 诱导下的小鼠肿瘤数量显著高于对照组,表明 *C. jejuni* 可以促进小鼠结直肠癌的发生,与前人的研究^[7]相符。在已报道的动物模型中^[7],观察到 *C. jejuni* 诱导的CRC有两个先决条件:即宿主的 *Apc* 基因突变和有一个致炎环境(如DSS处理)。因此,检测到的基因表达变化可以被描述为当前CRC

表2 14个核心差异表达基因
Table 2 Fourteen core DEGs

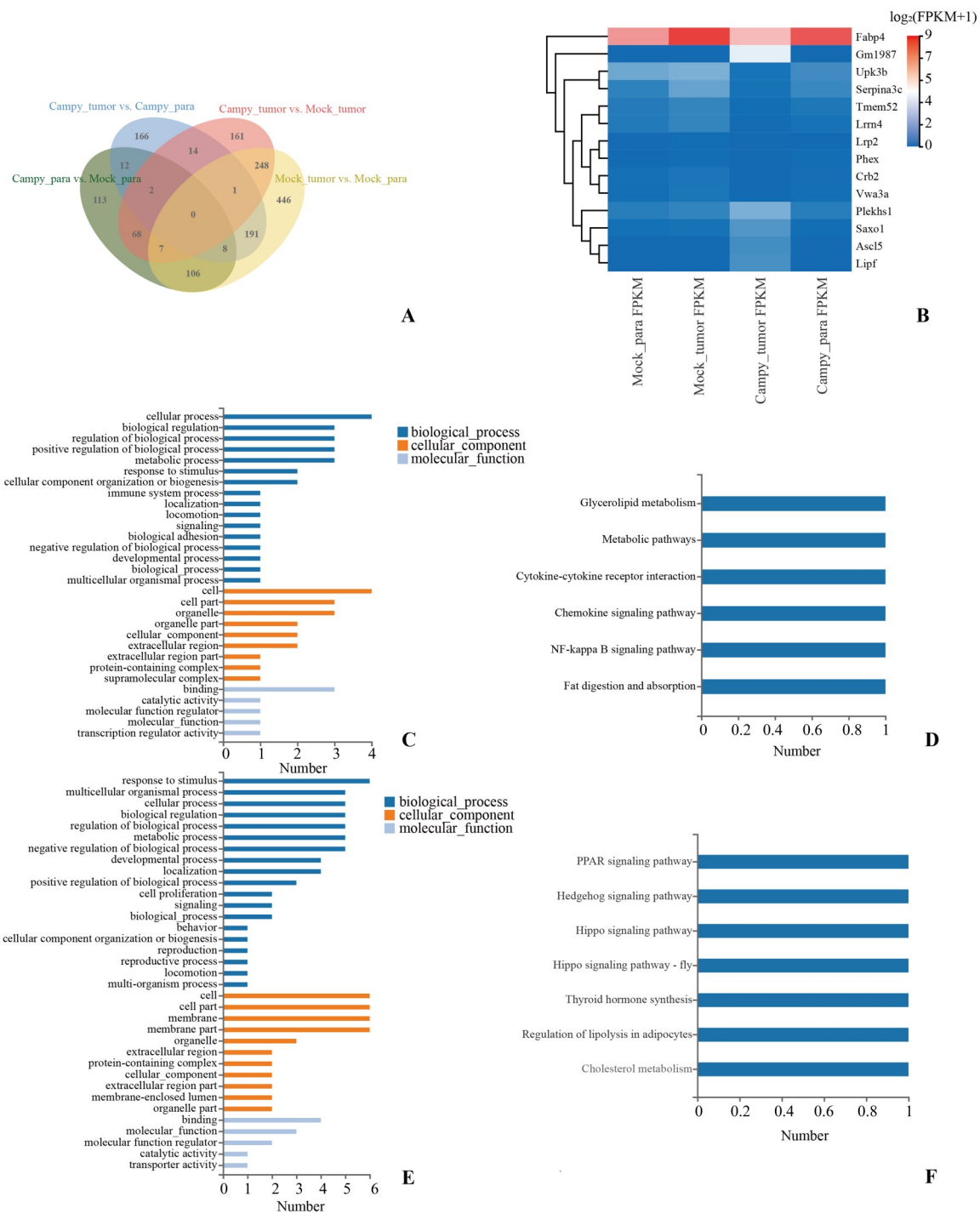
Gene name	Description	Campy_tumor vs. Campy_para		Campy_tumor vs. Mock_tumor	
		log ₂ FC	P value	log ₂ FC	P value
<i>Lipf</i>	lipase, gastric	20.61	0.000	20.41	0.000
<i>Gm1987</i>	predicted gene 1987	9.62	0.004	22.17	0.000
<i>Ascl5</i>	achaete-scute family bHLH transcription factor 5	5.12	0.029	5.32	0.042
<i>Saxo1</i>	stabilizer of axonemal microtubules 1	4.22	0.002	3.17	0.029
<i>Plekhs1</i>	pleckstrin homology domain containing, family S member 1	2.51	0.000	2.17	0.036
<i>Lrp2</i>	low density lipoprotein receptor-related protein 2	-3.33	0.014	-4.66	0.002
<i>Phex</i>	phosphate regulating endopeptidase homolog, X-linked	-5.03	0.037	-5.07	0.027
<i>Serpina3c</i>	serine (or cysteine) peptidase inhibitor, clade A, member 3C	-2.36	0.000	-3.44	0.015
<i>Fabp4</i>	fatty acid binding protein 4, adipocyte	-2.04	0.000	-2.28	0.019
<i>Vwa3a</i>	von Willebrand factor A domain containing 3A	-3.34	0.006	-4.81	0.001
<i>Tmem52</i>	transmembrane protein 52	-2.68	0.041	-3.69	0.015
<i>Lrrn4</i>	leucine rich repeat neuronal 4	-2.34	0.006	-4.18	0.004
<i>Upk3b</i>	uropod protein 3B	-2.30	0.030	-3.84	0.009
<i>Crb2</i>	crumbs family member 2	-2.44	0.007	-3.91	0.019

模型(即包含 *Apc*-和 DSS 前提)下, *C. jejuni* 诱导的 CRC 基因表达变化。

通过较为严谨的对照,我们指出了 14 个可能在 *C. jejuni* 诱导的结直肠癌发生发展中起到重要且独特作用的基因。经过进一步的 qRT-PCR 验证, 14 个基因中的 9 个(*Gm1987*、*Saxo1*、*Plekhs1*、*Lrp2*、*Serpina3c*、*Fabp4*、*Tmem52*、*Lrrn4*、*Upk3b*) 在结直肠肿瘤和癌旁部位的表达具有统计学上的显著性差异。其中关于表达上调的基因, *Plekhs1* 在人类里也有同源基因, 有研究表明它的过表达增强了蛋白激酶 B 的磷酸化, 继而增加致癌活性, 促进癌细胞增殖、存活和侵袭, 该基因能促进甲状腺癌的恶化^[15], 也可以作为膀胱癌^[16]和胃癌^[17]的基因标志物, 但目前关于它的表达是如何被调控的以及和结直肠癌的关系尚不清楚。结直肠作为临近胃和膀胱的一个器官, *C. jejuni* 很可能通过某些方式引起该基因的高表达, 继而引发结直肠癌的发展。*Saxo1* 编码微管稳定蛋白, 对纤毛等有特异性, 目前关于它的报道主要关注在小鼠精子活力方面, 鉴于我们的实验动物是雌性小鼠, 所以该基因的具体作用还有待探究。还有一个值得关注的基因是

Gm1987, 目前尚没有在 Pubmed 数据库里发现有关该基因的数据, 但它在我们的研究中提示参与了多条信号通路, 包括有细胞因子-细胞因子信号通路、NF- κ B 信号通路以及趋化因子信号通路, 这些都是癌症发生发展过程中极其重要的通路, 所以该基因值得进一步探究。

关于我们筛选出的 *C. jejuni* 诱导结直肠癌有关的表达下调的 6 个核心差异表达基因, 只有 2 个基因有被报道与 CRC 有关。*Fabp4* 是一种编码脂肪酸结合蛋白 4 的基因, 人类基因组上也有同源基因。有报道表明在 CRC 患者的肿瘤细胞中 FABP4 蛋白表达降低, 推测 FABP4 参与了 PPAR γ 通路促进结直肠癌细胞的迁移和侵袭^[18]。*Lrrn4* 编码富含亮氨酸重复序列神经元 4 蛋白, 先前的一项研究发现, LRRN4 在正常结肠组织中的丰度较低, 在 CRC 组织中的水平更低^[19]。然而前不久的一项研究表明它在 CRC 中无论是 mRNA 还是蛋白水平的表达都较高, 其通过 RAS/MAPK 信号通路调节肿瘤细胞的几种恶性表型^[20]。这种矛盾的结果可能是由于方法或者样本数不同引起的。还有一篇文献报道了 *Lrrn4* 和我们筛选的另一个表达下调的基因



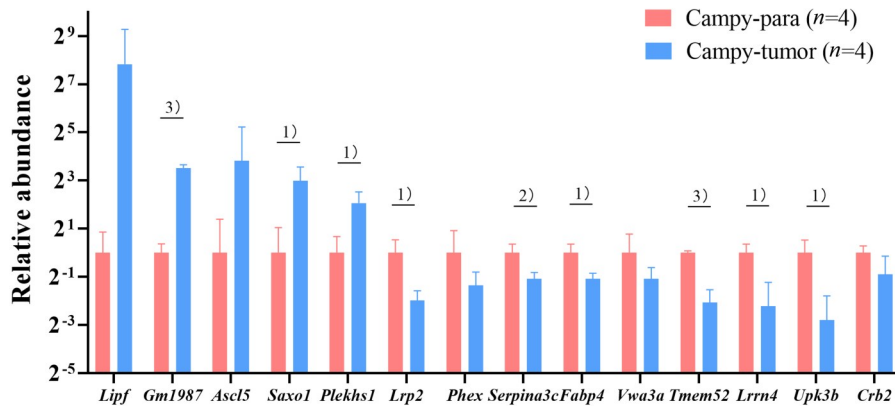
A: Venn diagram of common differentially expressed genes between different comparison groups. B: Heatmap of the core *C. jejuni* related DEGs. C and E: GO annotation of core *C. jejuni* related DEGs. D and F: KEGG annotation of core *C. jejuni* related DEGs. C and D: up-regulated genes; E and F: down-regulated genes.

图6 韦恩图法分析与14个核心DEGs的功能分析

Fig. 6 Venn diagram analysis and the resulting 14 core DEGs with functional analysis

Upk3b 它们编码的蛋白 LRRN4 和 uroplakin 相较于原代间皮细胞在间皮瘤内表达下调或缺失^[21], 但与 CRC 相关的研究目前没有任何报道。 *Lrp2* 编码低

密度脂蛋白相关蛋白 2 或维生素 D 相关转运蛋白 megalin, 它在结肠等多个上皮细胞系中表达^[22], 目前还没有研究调查 *Lrp2* 与 CRC 之间的可能关联,



Bars represent the geometric mean \pm SD. *P* values were calculated by unpaired, two-tailed Student's *t* test on log-transformed data. Specific *n* numbers are indicated in the figure. 1): *P* < 0.05 compared with Campy-para group; 2): *P* < 0.01 compared with Campy-para group; 3): *P* < 0.001 compared with Campy-para group.

图7 14个核心DEGs的qRT-PCR结果

Fig. 7 qRT-PCR results of 14 core DEGs

但与 *Lrp2* 相关的基因 *Cubn* 曾在了一项荟萃分析里报道与 CRC 存在一定的关系^[23]。*Serpina3c* 是小鼠中人类 *SERPINA4* 基因的一个分支,是调节脂肪生成过程中信号网络的关键因素,它可以通过 Wnt/ β -catenin-PPAR γ 途径参与脂肪分化的调节,在 *Serpina3c*^{-/-} 小鼠中引起脂肪细胞分化受损^[24]。虽然目前没有报道它与癌症的直接关系,但脂肪代谢的紊乱往往会促进癌症的发展。关于 *Tmem52* 基因,有少量的文献介绍其在肺腺癌组织中表达上升^[25]。总的来说,我们推测可能是这些表达下调的基因所编码的产物没有为肿瘤提供竞争优势,所以在肿瘤部位的表达会受抑制。有趣的是,在我们的研究中筛选出来的核心差异表达基因,大多数目前还没有被报道与 CRC 之间的关系,这值得将来进一步的探索。

本实验通过对筛选出来的差异基因进行 GO 富集、KEGG 通路富集分析,Campy 组与 Mock 组相比,差异基因参与了机体许多重要的生物学过程:生物膜破裂、蛋白质水解、信号转导通路以及与细菌引起的免疫反应相关的通路:抗菌肽介导的抗菌体液免疫反应、抗菌体液反应、防御革兰氏阳/阴性细菌反应和粘膜固有免疫途径。之前的研究表明,通过对 *C. jejuni* 81-176 感染的无菌 *Apc*^{Min/+} 小鼠的远端结肠组织进行转录组测序,*C. jejuni* 感染的小鼠中有

两条致癌途径富集——神经活性配体-受体相互作用和细胞因子-细胞因子受体相互作用^[7],在我们的 Campy 组与 Mock 组对比以及 Campy 组的肿瘤与癌旁组织对比中也有这两条通路的富集。此外,代谢通路在 Campy 组对比 Mock 组中富集最显著,提示着 *C. jejuni* 感染下代谢反应相当活跃。综上,差异基因参与了多条重要的生物学通路,提示结直肠癌的发生发展通过代谢、蛋白转运等相关通路进行。需要指出的是,本研究 campy-tumor-2 号小鼠的基因表达情况与该组其它小鼠差异明显,然而该鼠的肿瘤数量(4个)与该组平均水平(4个)相当。我们认为此小鼠组学数据的不同是个体差异性所致,我们将在后续研究中扩大样本量来平衡个体差异问题。

综上所述,本研究通过较为严谨的对照,指出了9个可能在 *C. jejuni* 诱导的结直肠癌的发生发展中起到重要且独特作用的基因:*Gm1987*, *Plekhs1*, *Saxo1*, *Lrp2*, *Serpina3c*, *Fabp4*, *Tmem52*, *Lrrn4*, *Upk3b*。为后续研究这些基因在 *C. jejuni* 诱导的结直肠癌中的作用提供了新的方向和思路。需要指出的是,关于这“9个宿主基因”,我们目前得出的结论是“可能”参与肿瘤的发生发展,我们将在后续的研究中扩大样本量,并对上述基因进行进一步的功能验证。

参考文献

- [1] Borges-Canha M, Portela-Cidade JP, Dinis-Ribeiro M, et al. Role of colonic microbiota in colorectal carcinogenesis: a systematic review [J]. *Rev Esp Enferm Dig*, 2015, 107(11): 659-671.
- [2] Wu N, Yang X, Zhang R, et al. Dysbiosis signature of fecal microbiota in colorectal cancer patients [J]. *Microbial Ecology*, 2013, 66(2): 462-470.
- [3] Warren RL, Freeman DJ, Pleasance S, et al. Co-occurrence of anaerobic bacteria in colorectal carcinomas [J]. *Microbiome*, 2013, 1(1): 16.
- [4] Allali I, Delgado S, Marron PI, et al. Gut microbiome compositional and functional differences between tumor and non-tumor adjacent tissues from cohorts from the US and Spain [J]. *Gut Microbes*, 2015, 6(3): 161-172.
- [5] Man SM. The clinical importance of emerging *Campylobacter* species [J]. *Nat Rev Gastroenterol Hepatol*, 2011, 8(12): 669-685.
- [6] Kaakoush NO, Castaño-Rodríguez N, Mitchell HM, et al. Global epidemiology of *campylobacter* infection [J]. *Clin Microbiol Rev*, 2015, 28(3): 687-720.
- [7] He Z, Gharaibeh RZ, Newsome RC, et al. *Campylobacter jejuni* promotes colorectal tumorigenesis through the action of cytolethal distending toxin [J]. *Gut*, 2019, 68(2): 289-300.
- [8] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics [J]. *Nat Rev Genet*, 2009, 10(1): 57-63.
- [9] Chen Y, Chen Y, Shi C, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data [J]. *Gigascience*, 2018, 7(1): 1-6.
- [10] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements [J]. *Nat Methods*, 2015, 12(4): 357-360.
- [11] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2 [J]. *Nat Methods*, 2012, 9(4): 357-359.
- [12] Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome [J]. *BMC Bioinformatics*, 2011, 12(1): 323.
- [13] 黄亚奇, 熊静禹, 沈达, 等. 空肠弯曲杆菌毒力因子及其致病机制研究进展 [J]. *中国家禽*, 2019, 41(22): 46-51.
- Huang YQ, Xiong JY, Shen D, et al. Research progress on virulence factors and pathogenesis of *Campylobacter jejuni* [J]. *China Poultry*, 2019, 41(22): 46-51.
- [14] 苗晋鑫, 宋韶鹤, 李秀敏. 结直肠癌小鼠模型研究进展 [J]. *中国实验动物学报*, 2020, 28(2): 267-272.
- Miao JX, Song SH, Li XM. Advances in mouse model of colorectal cancer [J]. *Acta Laboratorium Animalis Scientia Sinica*, 2020, 28(2): 267-272.
- [15] Xing X, Mu N, Yuan X, et al. PLEKHS1 over-expression is associated with metastases and poor outcomes in papillary thyroid carcinoma [J]. *Cancers (Basel)*, 2020, 12(8): 2133.
- [16] Gill E, Sandhu G, Ward D G, et al. The sirenic links between diabetes, obesity, and bladder cancer [J]. *Int J Mol Sci*, 2021, 22(20): 11150.
- [17] Abdel-Tawab MS, Fouad H, Othman AM, et al. Evaluation of gene expression of PLEKHS1, AADAC, and CDKN3 as novel genomic markers in gastric carcinoma [J]. *PloS one*, 2022, 17(4): e0265184.
- [18] Zhao D, Ma Y, Li X, et al. microRNA-211 promotes invasion and migration of colorectal cancer cells by targeting FABP4 via PPAR γ [J]. *J Cell Physiol*, 2019, 234(9): 15429-15437.
- [19] Lim SR, Gooi BH, Gam LH. Identification of low abundance proteins in colorectal cancer tissues [J]. *Cancer Biomark*, 2012, 12(4): 185-198.
- [20] Xu C, Chen Y, Long F, et al. Prognostic value and biological function of LRRN4 in colorectal cancer [J]. *Cancer Cell Int*, 2022, 22(1): 158.
- [21] Kanamori-Katayama M, Kaiho A, Ishizu Y, et al. LRRN4 and UPK3B are markers of primary mesothelial cells [J]. *PloS one*, 2011, 6(10): e25391.
- [22] Ternes SB, Rowling MJ. Vitamin D transport proteins megalin and disabled-2 are expressed in prostate and colon epithelial cells and are induced and activated by all-trans-retinoic acid [J]. *Nutr Cancer*, 2013, 65(6): 900-907.
- [23] Al-Tassan NA, Whiffin N, Hosking FJ, et al. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer [J]. *Scientific Reports*, 2015, 5(1): 10442.
- [24] Guo J, Qian L, Ji J, et al. Serpina3c regulates adipose differentiation via the Wnt/ β -catenin-PPAR γ pathway [J]. *Cell Signal*, 2022, 93: 110299.
- [25] Wang Z, Pei H, Liang H, et al. Construction and analysis of a circRNA-mediated ceRNA network in lung adenocarcinoma [J]. *Onco Targets Ther*, 2021, 14: 3659-3669.

(编辑 余菁)