目录

[1. 摘要 2](#_Toc57556237)

[Abstract 3](#_Toc57556238)

[2. 立论依据 4](#_Toc57556239)

[2.1. 研究问题 4](#_Toc57556240)

[2.2. 研究现状 4](#_Toc57556241)

[2.3. 研究意义 6](#_Toc57556242)

[3. 研究内容 7](#_Toc57556243)

[4. 材料、研究方法和技术路线 7](#_Toc57556244)

[4.1. 实验材料 7](#_Toc57556245)

[4.2. 实验方法 7](#_Toc57556246)

[4.2.1. AOM/DSS小鼠模型构建 7](#_Toc57556247)

[4.2.2. RNA-Seq数据分析 7](#_Toc57556248)

[4.2.3. LncRAN基因位置分类 7](#_Toc57556249)

[4.2.4. 时间序列分析 8](#_Toc57556250)

[4.2.5. 构建mRNA-lncRAN共表达网络 8](#_Toc57556251)

[4.2.6. LncRNA的系统性分析 8](#_Toc57556252)

[4.3. 技术路线 9](#_Toc57556253)

[5. 工作基础和已有进展 9](#_Toc57556254)

[5.1. 二代测序质控测评及比对率 9](#_Toc57556255)

[5.2. 样本相关性分析 12](#_Toc57556256)

[5.3. 差异表达分析 13](#_Toc57556257)

[6. 计划研究进度 15](#_Toc57556258)

[7. 预计目标和创新点 15](#_Toc57556259)

[7.1. 研究预期目标 15](#_Toc57556260)

[7.2. 研究创新点 16](#_Toc57556261)

[8. 参考文献 16](#_Toc57556262)

[9. 经费计算 16](#_Toc57556263)

# 摘要

结直肠癌是世界第三大最常被诊断为恶性的肿瘤疾病，并且是由癌症导致死亡的第二大主要原因，其发展机制是一个复杂的过程，而广泛参与生物发育和癌症发展的lncRNA这一类非编码的转录物近些年逐渐进人们的视野，lncRNA可以作为优质的预后标志物使其更具现实意义的研究价值。本研究通过挖掘结直肠癌发展过程中动态表达的lncRNA，结合公共数据库及整合多类分析工具，探索和结直肠癌发生相关的lncRNA，及其具体参与的方式，主要通过构建mRNA-lncRNA共表达网络，ceRNA及lncRNA与蛋白质结合等角度出发，为更好理解结直肠癌的发病机制及lncRNA分子的功能。

**关键词**：结直肠癌；AOM/DSS小鼠模型；动态表达lncRNA；CeRNA

# Abstract

Colorectal cancer is the third most diagnosed malignant tumor disease in the world and the second reason of death caused by cancer. The mechanism of colorectal cancer is so complicated that we haven’t totally understand. lncRNAs are widely involved in the biological development and cancer development which are considered as ‘junk’ transcripts previously have gradually came into our sight recent years because of Comprehensive understanding of lncRNA function and its can be used as diagnostic biomarkers in clinical. This study attempts to explore the relationship between the development of colorectal cancer and the lncRNAs by mining the dynamic expression lncRNAs during the development of colorectal cancer, combing public databases and integrating multiple analysis tools in bioinformatic. We are going do mRNA-lncRNA co-expression analysis, CeRNA and the interaction between lncRNAs and protein to explore molecular mechanism of lncRNA.

**Keyword:** Colorectal cancer; dynamic expression lncRNAs; CeRNA

# 立论依据

## 研究问题

结直肠癌(Colorectal cancer)，又称大肠癌，直肠癌，大肠直肠癌或肠癌。根据国际癌症研究机构2018年的估计，结直肠癌在全球范围内每年构成约180万例新病例，900000例死亡病例，使其成为是第三大最常被诊断为恶性肿瘤，并且是由癌症导致死亡的第二大主要原因(Bray et al., 2018)。大多数结直肠癌（>90%）是腺癌，是一种从结肠和直肠的腺上皮细胞发展而来的恶性肿瘤(Fleming et al., 2012)，结直肠癌是一种涉及到遗传，环境，生活方式等多风险因素影响复杂疾病(Walsh and Terdiman, 2003)，其发展机制是一个复杂的过程，

几年前，基因组主流的生物学功能主要局限于蛋白质编码基因和少数几种非编码RNA(rRNA，tRNA等)(ncRNA)。而然，随着基因组和转录组技术的发展和研究结果表明，大约有人类基因组大约有87%是活跃转录的，更令人惊讶的是，其中70%的转录本是非编码的(Djebali et al., 2012)。大于200个碱基的ncRNA称为lncRNA(long noncoding RNA)(Morris and Mattick, 2014)lncRNA可能在细胞中具有重要的功能，很多研究已经证实它们参与了发育和疾病等生物。

## 研究现状

（1）结直肠癌的发病过程和相关信号通路：

结直肠癌的自然病史可分为四个主要阶段：起始(initiation)，促进(promotion)，发展(promotion)和转移(metastasis)(Pitot, 1993)。起始阶段涉及不可逆的基因损伤，使受损细胞易于转化为随后的肿瘤。在促进阶段，起始细胞增殖，诱导异常生长（肿瘤），在随后的发展阶段，通过进行进一步的遗传和表观遗传学改变，这些改变可以导致细胞选择性的生长优势，良性肿瘤细胞转化为恶性癌细胞并获得侵袭性特征和转移的潜力。转移的特征是癌细胞通过血液或者淋巴系统从原发器官扩散到其他器官或组织。结直肠癌的每个阶段的持续时间很难估计，范围很广，这些过程需要很长时间，并且要在结直肠癌完成所有阶段都需要数十年的时间(Carethers and Jung, 2015)。结直肠癌发展的机制是一个复杂的过程，涉及随着癌症发展而发生的连续突变事件(Fearon, 2011)。一系列研究表明，不同的信号通路是无序的，并且有作为结直肠癌治疗靶点的潜力。Wnt，P13K/Akt，hedgehog，ErbB，RHOA，Notch，BMP，Hippo，AMPK，NF-κB ，MAPK和JNK，这些信号通路都是参与结直肠癌发展，且作用是精确而复杂(Wan et al., 2020)。

（2）AOM/DSS结直肠癌小鼠模型：

已经开发了许多和炎症相关的结直肠癌小鼠模型，包括化学诱导的小鼠模型，已证明单次使用甲氧基甲烷(AOM)与啮齿动物发炎剂葡聚糖硫酸钠(DSS)结合可显著缩短化学诱导结直肠癌模型，由于这种方法可重复性高和效率高，AOM/DSS小鼠模型可以高度可靠的概括人类肿瘤的发生和发展阶段，使其成为研究结直肠癌发生的杰出模型。

（3）lncRNA可以作为肿瘤发生中的癌症生物标志物：

lncRNA在肿瘤发生过程中具有高特异性和敏感性，因为与蛋白质编码基因相比，lncRNA表达水平根据有组织特异性，是肿瘤状态的指标(Li and Chen, 2013)，据研究，lncRNA可以作为肿瘤发育的肿瘤基因，也可以抑制肿瘤的发育(Qi and Du, 2013)，例如，被广泛研究的抑癌基因p53的表达与很多lncRNA相关，如在小鼠中观察到的lincRNA-p21受p53调控(Huarte et al., 2010)以及MEG3(Maternally Expressed gene 3 imprinted lncRNA)在肺癌中可以激活p53(Lu et al., 2013)等等。此外，在一些信号通路中，lncRNA对维持细胞平衡和预防致癌至关重要(Sahu et al., 2015)，例如lncRNA-ATB是经过TGF-β促进IL-11从而导致肝癌细胞的转移(HCC)(Yuan et al., 2014)。(2)一些lncRNA可相对稳定的从体液中取样，研究者对它们兴趣的一个重要的原因，lncRNA不仅在细胞内水平限制较低，在细胞外液也可以检测到lncRNA水平的变化，这使得对患者取样更加方便。这一点是非常关键的，一个良好标志物的特点在于其易于获得，可实现非侵入性采样。LncRNA可以分泌到体液中的机制目前还在研究，有研究表明这个过程通过囊泡或微泡介导的，这些囊泡直径50-100nm由脂蛋白形成(Arita et al., 2013)。也有研究表明囊泡中的RNA序列中含有3.36%的lncRNA，并且这个比例和lncRNA在血浆中的含量非常类似，这就表明大多数血浆lncRNA都在囊泡内部(Ramalingam and Belani, 2008)。也有学者表明，肿瘤细胞或者其临近细胞可以直接释放lncRNA进入循环(Pickl et al., 2014)，这些研究表明lncRNA是非常优秀潜在的生物标志物，并且lncRNA参与大量的生物学过程，这些分子反应了肿瘤的动态状态，说明lncRNA可作为预后工具有望阐明癌症患者病情的阶段。除了蛋白质编码基因的突变或异常表达之外，一些lncRNA的突变和失调在癌症中也起着重要作用，作为各种细胞生物学过程相关的新发现转录物，它们与多种癌症相关，与肿瘤发生，转移和肿瘤分期密切相关(Vitiello et al., 2015)。H19是最早发现的lncRNA之一，在胚胎发育和肿瘤发生中起着关键作用(Ariel et al., 2000)。H19充当miRNA的诱饵，调节其可用性和活性，也可与转录阻遏物相互作用例如EZH2和MBD1，并通过它们募集到靶基因。P53抑制H19基因，H19衍生的miR-675抑制P53和其依赖性蛋白的表达(Terracciano et al., 2017)。H19的表达涉及血管生成，细胞存活和增殖的基因激活，从而引发包括结直肠癌在内的多种恶性肿瘤。再例如CCAT1通过充当miR-155竞争性内源RNA抑制c-Myc表达从而表观遗传的下调c-Myc，同时也参与HOXB13和SPRY4的调控，CCAT1与结直肠癌，食道癌等癌症有关(Guo and Hua, 2017)。

## 研究意义

今年多有越来越多的研究揭示了lncRNA在生命过程中的分子功能，及其在临床上的潜在利用价值，但这是远远不够的，需要更进一步探索lncRNA潜在的功能揭示这一分子在生命过程中扮演的角色，本研究以结直肠癌为研究背景，探索动态表达的lncRNA在癌症发展过程中的作用，试图寻找更为优质的预后标志物。

# 研究内容

本研究以AOM/DSS结直肠癌小鼠模型为研究对象，二代的转录组测序技术，探索在结直肠癌发生过程中，动态表达的蛋白编码基因和lncRNA，构建mRNA-lncRNA共表达网络，ceRNA网络，lncRNA与TF等蛋白质的互作。

# 材料、研究方法和技术路线

## 实验材料

（1）对照，第二周，第四周，第七周，第十周的AOM/DSS小鼠模型肠上皮组织的二代illumina的RNA-seq测序数据，每个时期3个重复

（2）本地intel(R) Core(TM) i5-10210U CPU @1.6GHz，RAM:12G，华中农业大学二总集群

## 实验方法

### AOM/DSS小鼠模型构建

### RNA-Seq数据分析

使用illumina二代测序技术，使用标准的RNA-seq分析流程：fastp质控，hisat2比对到参考基因组，samtools对比对的bam文件排序并压缩，featurecount定量，使用的基因注释文件为Ensemblv98和NONCODEv5。

### LncRAN基因位置分类

使用FEELnc v0.1.1软件根据lncRNA与编码基因的位置进行分类，本研究中使用的皆为默认参数，将分类结果整合位置(重叠，上游，下游)和链(有义，反义)，在基因上游且转录方向相反距离不超过2 kb的lncRNA基因被分类为divergent，距离其最近编码基因大于100 kb的lncRNA基因被分类为基因间区lncRNA。

### 时间序列分析

本研究使用的时间序列分析的方法为maSigPro，首先使用edger v3.30.3首先对基因表达矩阵进行CPM标准化，计算拟合优度(R2)度量，R2>0.35的基因被归类为直肠癌发生过程中动态表达的基因。

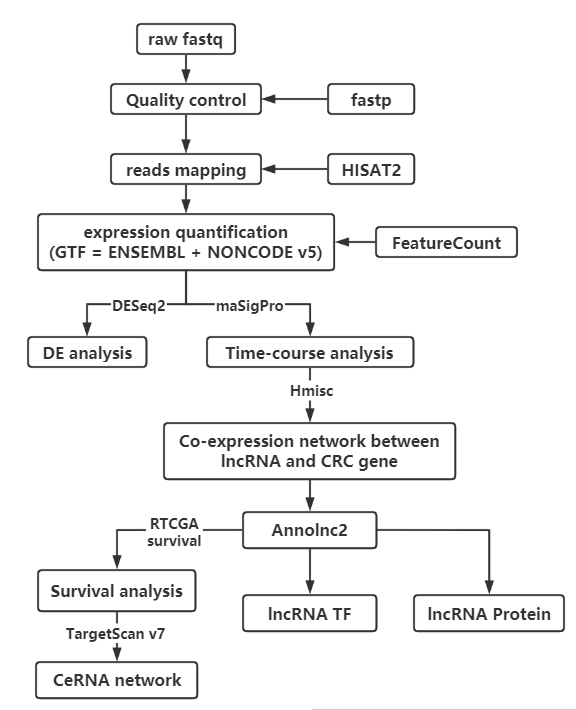
### 构建mRNA-lncRAN共表达网络

使用Himsc计算关键基因之间的相关性，R2绝对值大于等于0.7，pvalue小于0.05的基因对被认为是存在共表达，通过Cytoscape可视化动态mRNA-lncRNA共表达网络。

### LncRNA的系统性分析

将Annolnc2部署到本地，系统性分析关键的lncRNA，对其结构，表达模式，调控遗传，演化等做出系统性分析，之后整合TargetScan，miTarbase等数据，寻找lncRNA可能的生物功能。

## 技术路线

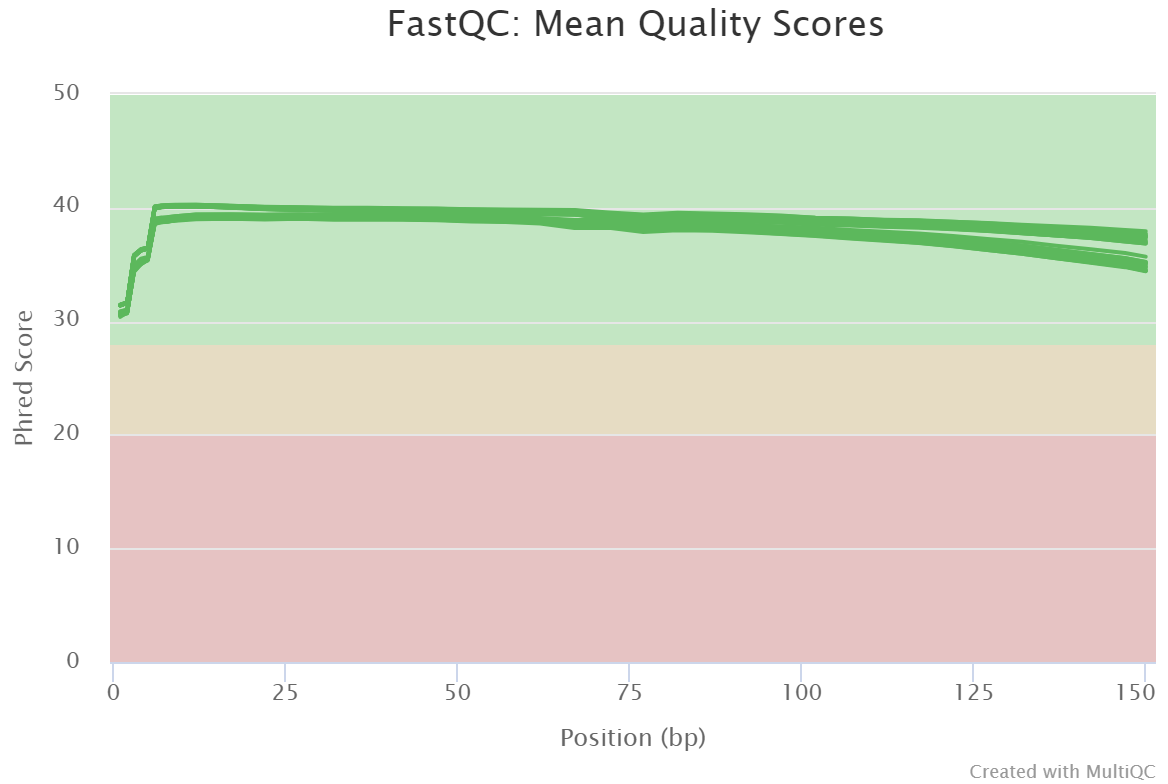


**图 1技术路线**

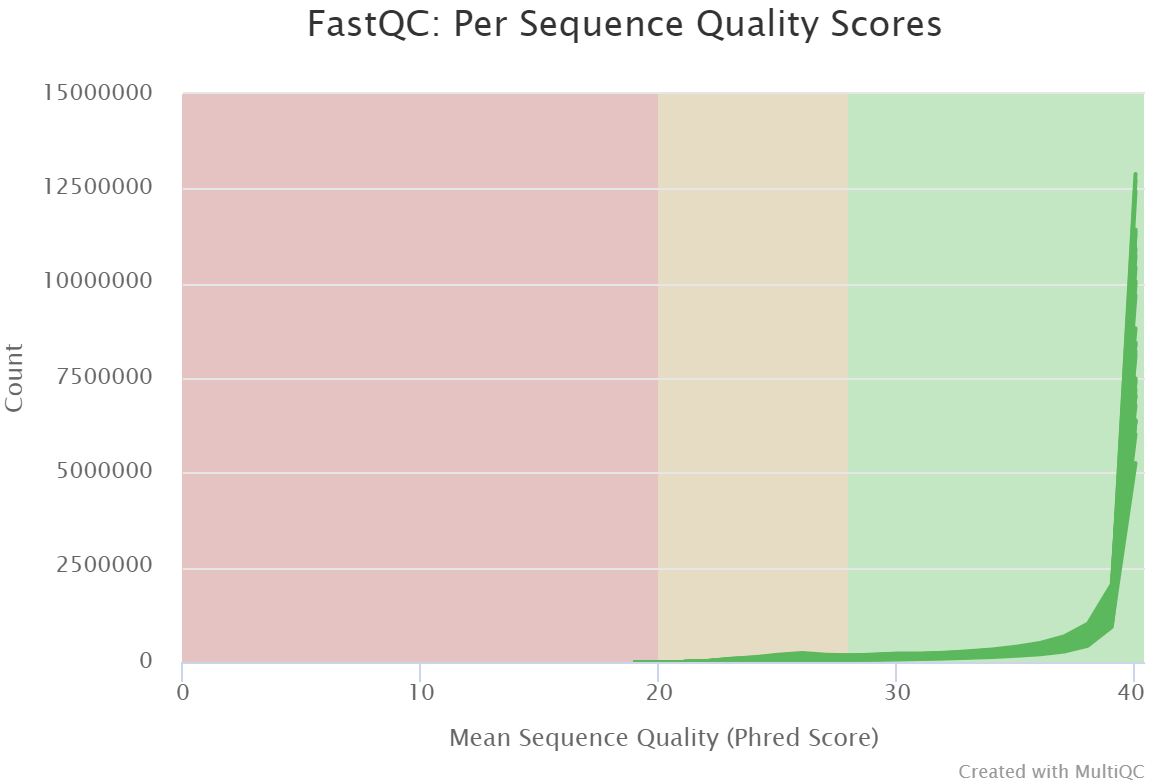
# 工作基础和已有进展

## 二代测序质控测评及比对率

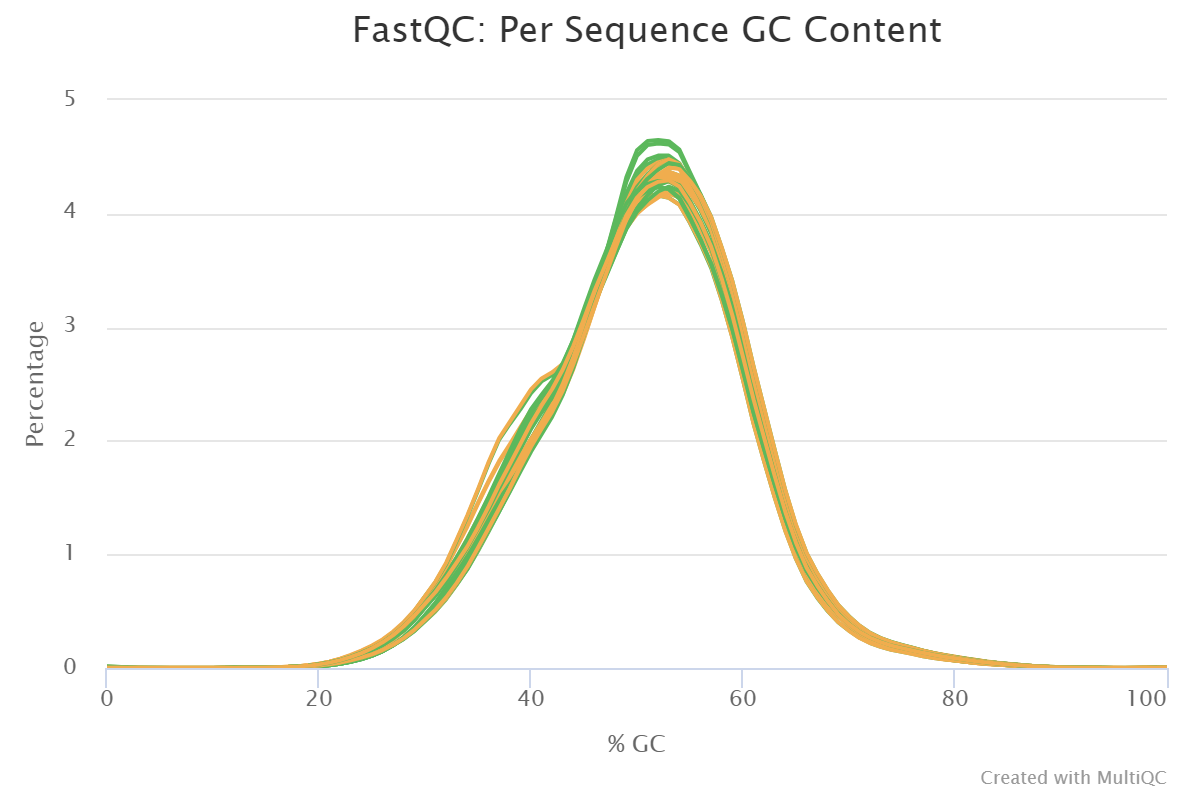
对测序得到的Fastq文件进行质量检测后发现需要进一步过滤，使用fastp对测序数据进行过滤后再次质量检测，最后用MultiQC进行可视化，结果如下图1，2，3，4所示，每个reads上的碱基数质量分数都在30以上，绝大多数reads的质量分数在40，表明测序数据准确度高，而每个reads上的GC含量的分布基本符合正态分布，说明建库过程中污染较小，且PCR偏差较小。如图5所示，所有样本的比对率均在80%以上，进一步说明测序质量很好，总而言之，质控后的测序数据基本符合预期，可以进一步实验。



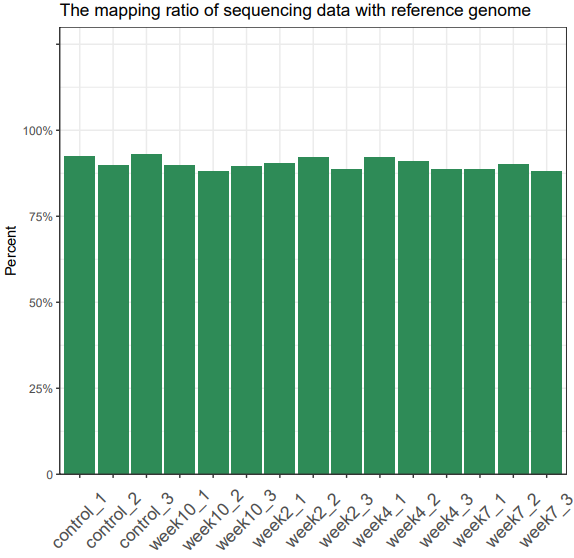
**图 2每个reads上所有碱基的平均质量分数**



**图 3 每个序列的质量得分**



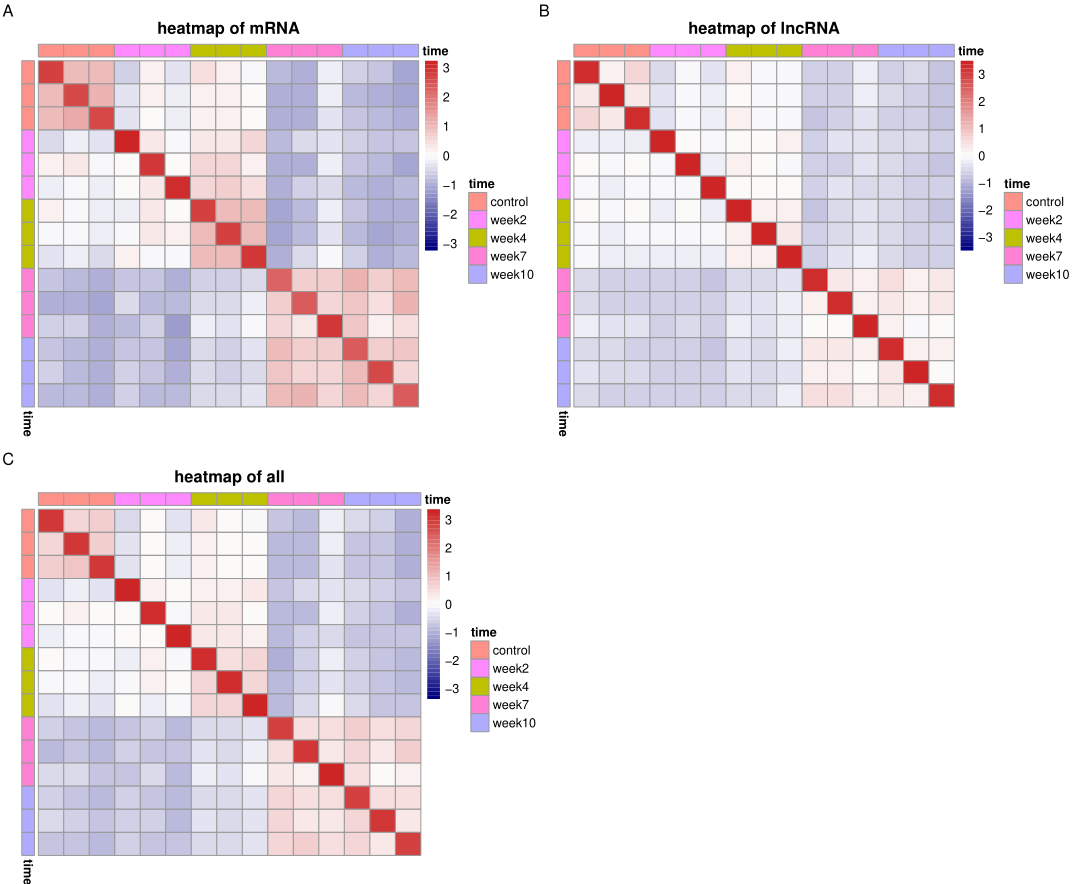
**图 4 每个序列的GC含量**



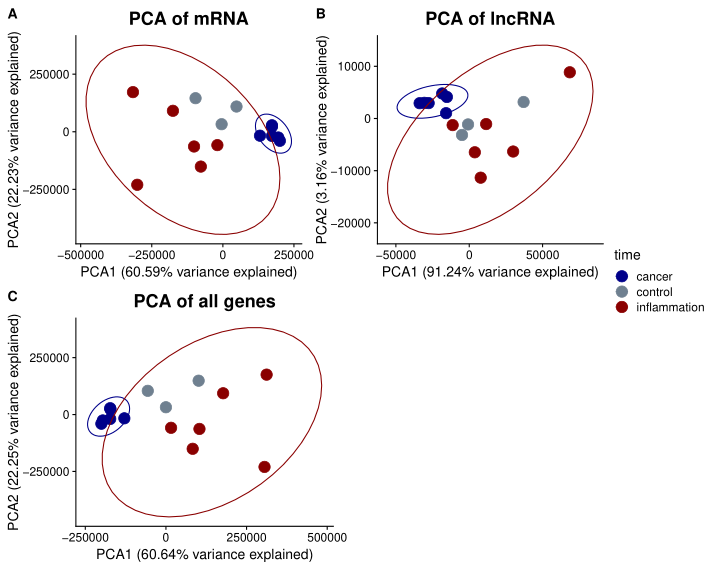
**图 5 每个样本比对到参考基因组上的比对率**

## 样本相关性分析

用于评估两个变量之间的关联强度及关系方向，如图6所示，同一时期的样本之间相关性很高，control之间相关性很高，week4之间相关性很高，week7和week10之间相关性很高，基本符合实验预期，week2到week4在AOM/DSS结直肠癌小鼠模型中数据炎症时期，week7到week10属于癌症时期。为了进一步验证，将week2和week4作为炎症期，week7和week10作为癌症期做样本间主成分分析，如图7所示，基本癌症时期聚在一起，与control和炎症时期分开，基本符合实验预期。



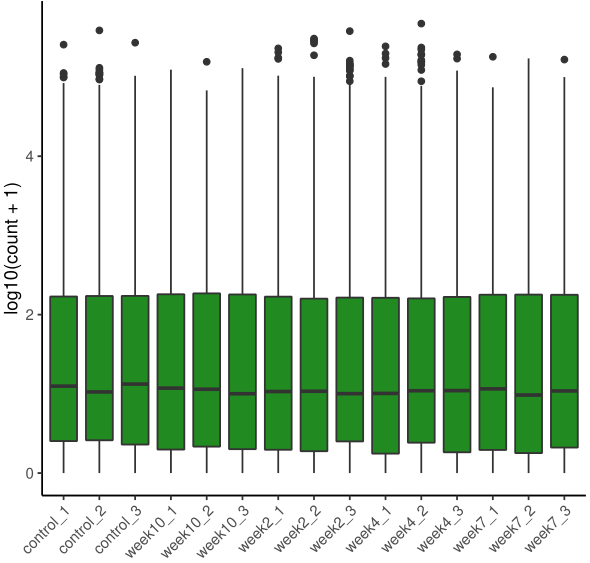
**图 6样本间相关性分析**



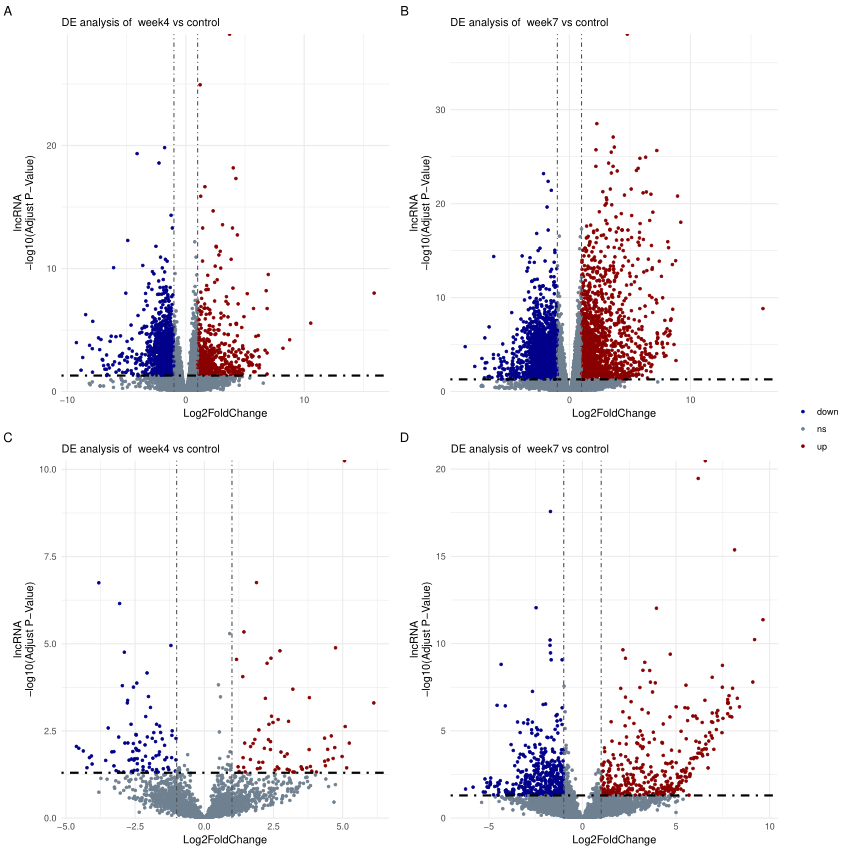
**图 7样本间主成分分析**

## 差异表达分析

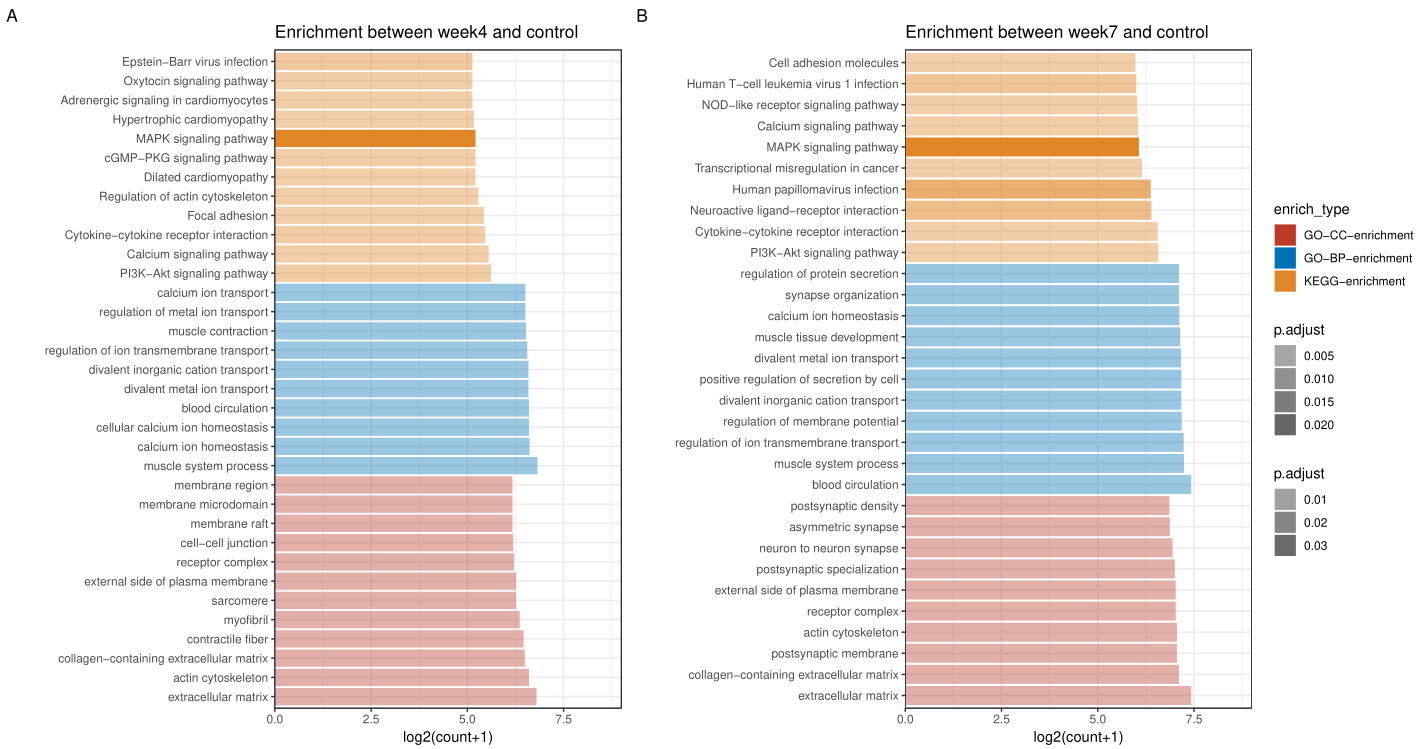
为了排除批次效应的影响，首先统计所有样本的表达量，如图8所示，所有样本表达水平基本一致，表明批次效应对本实验影响较小。使用DESeq对第4周和第7周做差异表达分析，p.adjust小于0.05且LogFoldChange绝对值大于2判定为差异表达基因，图9火山图展示了差异表达的lncRNA和mRNA上调和下调的分布，可以看出不论是mRNA还是lncRNA上调的基因个数都大于基因下调的个数。接着，对差异表达的蛋白编码基因进行功能富集分析，如图9展示了部分的富集分析结果，本研究后期会做时间序列分析探究与结直肠癌在发生过程相关的动态表达的蛋白编码基因和lncRNA，并探究他们之间的关系，这里只是简单观测一下，但是可以看到week4和week7差异表达的蛋白编码基因富集到癌症相关的信号通路MAPK等等，符合实验预期。



**图 8标准化后样本表达量**



**图 9 第四周和第七周差异表达基因上下调分布**



**图 10 第四周和第七周差异表达基因的GO和KEGG功能富集分析**

# 计划研究进度

|  |  |
| --- | --- |
| 时间 | 研究计划 |
| 2020年10月之前 | mRNA-lncRNA共表达网络 |
| 2020.年11月之前 | 构建ceRNA |
| 2020年12月之前 | 探究lncRNA与蛋白分子之间互作 |

# 预计目标和创新点

## 研究预期目标

（1）寻找和结直肠癌发展相关的动态lncRNA

（2）探究这些lncRNA具体是作用机制

## 研究创新点

今年来研究报道了各式各样的lncRNA，它们在生命过程中展现出丰富的功能的，但这对于lncRNA的研究是远远不够的，急需探索更多lncRNA及其参与复杂生命过程的方式，一是解密这一类分子的在生命体中的角色，甚至是从生物演化的角度去探索这些分子从何而来，现实的是利用lncRNA在临床上的特性探索更为优质的预后标志物。本研究试图从动态表达的lncRNA为出发点，通过挖掘公共数据库并整合多类分析方法，探索这些lncRNA和miRNA潜在的ceRNA网络，以及与转录因子，蛋白质互作方式等，希望能找到某些和结直肠癌发展相关的动态表达lncRNA以及动态表达的lncRNA在癌症中的特征，希望为揭秘lncRNA这一曾被广泛认定为基因组产生的junk这一类有趣的转录物贡献一份力。

# 参考文献

# 经费计算

应该花费不多。

1. ARIEL I, SUGHAYER M, FELLIG Y, et al. 2000. The imprinted H19 gene is a marker of early recurrence in human bladder carcinoma. Molecular Pathology [J], 53: 320.
2. ARITA T, ICHIKAWA D, KONISHI H, et al. 2013. Circulating long non-coding RNAs in plasma of patients with gastric cancer. Anticancer Res [J], 33: 3185-3193.
3. BRAY F, FERLAY J, SOERJOMATARAM I, et al. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin [J], 68: 394-424.
4. CARETHERS J M, JUNG B H 2015. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. Gastroenterology [J], 149: 1177-1190 e1173.
5. DJEBALI S, DAVIS C A, MERKEL A, et al. 2012. Landscape of transcription in human cells. Nature [J], 489: 101-108.
6. FEARON E R 2011. Molecular genetics of colorectal cancer. Annu Rev Pathol [J], 6: 479-507.
7. FLEMING M, RAVULA S, TATISHCHEV S F, et al. 2012. Colorectal carcinoma: Pathologic aspects. J Gastrointest Oncol [J], 3: 153-173.
8. GUO X, HUA Y 2017. CCAT1: an oncogenic long noncoding RNA in human cancers. Journal of cancer research and clinical oncology [J], 143: 555-562.
9. HUARTE M, GUTTMAN M, FELDSER D, et al. 2010. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell [J], 142: 409-419.
10. LI C H, CHEN Y 2013. Targeting long non-coding RNAs in cancers: progress and prospects. Int J Biochem Cell Biol [J], 45: 1895-1910.
11. LU K H, LI W, LIU X H, et al. 2013. Long non-coding RNA MEG3 inhibits NSCLC cells proliferation and induces apoptosis by affecting p53 expression. BMC Cancer [J], 13: 461.
12. MORRIS K V, MATTICK J S 2014. The rise of regulatory RNA. Nat Rev Genet [J], 15: 423-437.
13. PICKL J M, HECKMANN D, RATZ L, et al. 2014. Novel RNA markers in prostate cancer: functional considerations and clinical translation. Biomed Res Int [J], 2014: 765207.
14. PITOT H C 1993. The molecular biology of carcinogenesis. Cancer [J], 72: 962-970.
15. QI P, DU X 2013. The long non-coding RNAs, a new cancer diagnostic and therapeutic gold mine. Mod Pathol [J], 26: 155-165.
16. RAMALINGAM S, BELANI C 2008. Systemic chemotherapy for advanced non-small cell lung cancer: recent advances and future directions. Oncologist [J], 13 Suppl 1: 5-13.
17. SAHU A, SINGHAL U, CHINNAIYAN A M 2015. Long noncoding RNAs in cancer: from function to translation. Trends Cancer [J], 1: 93-109.
18. TERRACCIANO D, FERRO M, TERRERI S, et al. 2017. Urinary long noncoding RNAs in nonmuscle-invasive bladder cancer: new architects in cancer prognostic biomarkers. Transl Res [J], 184: 108-117.
19. VITIELLO M, TUCCOLI A, POLISENO L 2015. Long non-coding RNAs in cancer: implications for personalized therapy. Cellular oncology [J], 38: 17-28.
20. WALSH J M, TERDIMAN J P 2003. Colorectal cancer screening: scientific review. JAMA [J], 289: 1288-1296.
21. WAN M L, WANG Y, ZENG Z, et al. 2020. Colorectal cancer (CRC) as a multifactorial disease and its causal correlations with multiple signaling pathways. Biosci Rep [J], 40.
22. YUAN J H, YANG F, WANG F, et al. 2014. A long noncoding RNA activated by TGF-β promotes the invasion-metastasis cascade in hepatocellular carcinoma. Cancer Cell [J], 25: 666-681.