[1. 前言 2](#_Toc57666846)

[1.1. 结直肠癌 2](#_Toc57666847)

[1.1.1. 结直肠癌发病过程和相关信号通路 3](#_Toc57666848)

[1.1.2. 结直肠癌临床诊断及预防 5](#_Toc57666849)

[1.2. 长链非编码RNA 6](#_Toc57666850)

[1.2.1. 非编码RNA 6](#_Toc57666851)

[1.2.2. LncRNA的功能 6](#_Toc57666852)

[1.2.3. lncRNA与结直肠癌 9](#_Toc57666853)

[1.3. AOM/DSS结直肠癌小鼠模型 10](#_Toc57666854)

[1.3.1. AOM/DSS小鼠模型组织病理学 10](#_Toc57666855)

[1.3.2. AOM/DSS小鼠模型分子特征 10](#_Toc57666856)

[2. 材料与方法 11](#_Toc57666857)

[2.1. 实验材料 11](#_Toc57666858)

[2.2. 主要生物信息软件 11](#_Toc57666859)

[2.3. 实验方法 11](#_Toc57666860)

[2.3.1. AOM/DSS小鼠模型构建 11](#_Toc57666861)

[2.3.2. RNA-Seq数据分析 11](#_Toc57666862)

[2.3.3. LncRNA基因位置分类 12](#_Toc57666863)

[2.3.4. 主成分分析和差异表达分析 12](#_Toc57666864)

[2.3.5. 基因功能富集分析 13](#_Toc57666865)

[2.3.6. 时间序列分析 13](#_Toc57666866)

[2.3.7. 构建mRNA-lncRNA共表达网络 13](#_Toc57666867)

[2.3.8. LncRNA的系统性分析 14](#_Toc57666868)

[2.3.9. 构建ceRNA网络 14](#_Toc57666869)

[2.4. 技术流程图 15](#_Toc57666870)

[3. 结果与分析 16](#_Toc57666871)

[3.1. 测序数据质量及比对率 16](#_Toc57666872)

[3.2. 样本间相关性分析 18](#_Toc57666873)

[3.3. 差异表达分析 20](#_Toc57666874)

[3.4. maSigPro时间序列分析 22](#_Toc57666875)

[3.4.1. 鉴定动态表达基因 22](#_Toc57666876)

[3.4.2. 动态表达的lncRNA基因特征 24](#_Toc57666877)

[3.4.3. 寻找与结直肠癌发展相关的动态基因 25](#_Toc57666878)

[3.5. 构建mRNA-lncRNA共表达网络 27](#_Toc57666879)

[3.6. ceRNA调控网络构建 28](#_Toc57666880)

[3.6.1. lncRNA与miRNA结合靶位点预测 28](#_Toc57666881)

[3.6.2. miRNA靶基因功能预测及功能分析 29](#_Toc57666882)

[3.6.3. miRNA生存分析及差异表达分析 29](#_Toc57666883)

[3.6.4. CeRNA的构建 32](#_Toc57666884)

[3.7. 动态表达lncRNA与蛋白质互作分析 33](#_Toc57666885)

[3.8. 动态表达lncRNA与转录因子互作分析 33](#_Toc57666886)

# 前言

## 结直肠癌

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)。

通常结直肠癌在经济发达和经济迅速增长的地区发病率较高，加大筛查力度会提升早期发现和改善治疗，因此结直肠癌的发病率的增加并不一定会导致死亡率上升(Arnold et al., 2017)。结直肠癌是老龄化疾病，2018年全球约90%的病例和死亡发生在50以后。男性的发病率和死亡率都高于女性，分别约为女性的1.4倍和1.5倍，在50岁以后差距更为明显(Jiao et al., 2014)，这可能与多种因素相关，男性在结直肠癌发生受环境因素的影响比遗传因素的影响更大，研究估计男性的结直肠癌遗传力为28%而女性的为45%(Graff et al., 2017)，男性也更容易接触类风险因素如内脏脂肪，酒精饮料，吸烟和不良的饮食习惯，此外有可能女性受益于内源性雌激素的保护作用(Murphy et al., 2015)。

结直肠癌是有高度异质性疾病，由不同病因和临床结果的亚型组成。一般情况下按照结直肠肿瘤解剖部位可定义为：近端结肠，远端结肠和直肠(Li and Lai, 2009)。研究表明不同的风险因素和不同的肿瘤解剖部位相关，如吸烟与近端结直肠和直肠癌的风险相关但与远端结直肠癌不相关。女性的近端结直肠癌比男性更普遍，比例随着年龄的增加而增加，亚洲人直肠癌患者的比例最高(Shin et al., 2012)，而欧美人近端结肠癌的比例最高(Yang et al., 2018)，这表明结直肠癌发病机制中环境因素比遗传因素的影响更大。此外整个结直肠癌肿瘤部位很可能部分与肠道微生物和宿主特征的变化有关。结直肠癌的发展是一个多步骤过程，基因组的不稳定性目前被确定为一种主要的分子特征，会导致潜在遗传畸变增加，大量研究表明，至少存在三种涉及结直肠癌起源和发展的途径：染色体不稳定(CIN)，微卫星不稳定(MSI)，CpG岛甲基化表型(CIMP)(Pino and Chung, 2010)。CIN途径占85%，CIN肿瘤的核型分析表明一些特定的抑癌基因和癌症基因染色体异常，而这些突变和异常基因在肿瘤发展途径的作用是公认的(Esteller et al., 2001)。MSI途径是指包含微卫星区域的反复变化，但不会造成基因组中明显的结构和数值变化。有报道道指出，由于不匹配修复系统中的种系突变或由于MLH1基因启动子甲基化导致体细胞失活，约15%结直肠癌患者中的MSI发生率很高(Singh et al., 2019)。CIPM是另一个导致结直肠癌的主要途径，约占结直肠癌患者的20%至30%，其中女性占30%至40%，近端结肠占3%至12%(Migliore et al., 2011)。CIMP肿瘤具有独特的病因，分子特征和表观遗传环境，1999年研究人员首次提出了大肠癌的CIMP现象，在MINT基因的7个基因组发现了高度甲基化，其中至少有3个与CDKN2A和MLH1高甲基化密切相关(Toyota et al., 1999)。描述分子亚型的方法具有排他性，某种分类的方法在对特定的一组数据中可能是正确的，但是有必要建立一个黄金标准，能够归一化用于数据处理。2015年有学者系统性将六个独立分类系统合并为四个具有不同特征的公有分子亚型(CMS)：CMS1(微卫星不稳定免疫)，超突变，微卫星不稳定和强免疫激活；CMS2(标准，37%)，上皮，WNT和MYC信号激活；CMS3(新陈代谢，13%)，上皮和明显的代谢失调；CMS4(间质，23%)，转化生长因子β的显著激活，基质浸润和血管生成。具有混合特征(13%)的样品可能为过渡表型或肿瘤内的异质性(Guinney et al., 2015)。结直肠癌分子亚型可改善结直肠癌的预后和诊断，但是CMS分类仍在临床试验中探索。

### 结直肠癌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)。值得注意的是，人们越来越考虑癌症干细胞的存在，癌症干细胞可能通过迅速的连续分裂成癌细胞或相同的子代细胞从而形成增值性癌细胞群，从而在结直肠癌发生的过程中起作用。

结直肠癌通过不同的致癌途径产生：腺瘤-癌序列(adenoma–carcinoma sequence)，锯齿状腺瘤途径(serrated pathway)和炎症途径(inflammatory pathway)。腺瘤-癌序列（其中腺瘤是结直肠癌的前体）是解释大多是结直肠癌的经典途径，在该模型中，遗传和表观遗传的变化是逐步积累驱动正常细胞向小腺瘤，大腺瘤，最后向癌症转化(Keum et al., 2019)。结直肠癌锯齿状腺瘤途径（主要是固着锯齿状腺瘤），该模型描述为正常细胞逐步变成增生型息肉(hyperplastic polyp)，再到无柄锯齿状腺瘤，最后形成结直肠癌(Kedrin and Gala, 2015)。另一种独特的致癌途径涉及慢性炎症，据报道在炎症性肠病(IBD)，尤其是溃疡性结直肠炎患者中，患结直肠癌的风险(95%CI2.1-2.7)估计比普通人群平均高2.4倍(Jess et al., 2012)，在这些患者中，癌变过程从无发育异常发展到不确定的发育异常，低度发育异常，高度发育异常，最后发展为结直肠癌。

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)。此外，越来越多的研究表明，遗传干扰和表观遗传的异常可能促进结直肠癌的发展，或者癌症可以引起遗传扰动或表观遗传的异常(Visone et al., 2019)。

### 结直肠癌临床诊断及预防

结直肠癌患者临床表现出广泛的体征和症状，例如隐匿性或者明显的直肠出血，排便习惯的改变，贫血或腹痛等。然而，结直肠癌在达到晚期前基本上是无症状疾病，因此需要其他的风险因素来辅助确定需要进一步肠镜检查的患者。新发的直肠出血的45岁以上的患者需要进行肠镜检查，在年轻患者中应该通过其他因素判断(有大肠癌家族病史，排便习惯改变，体重的减轻以及血液粪便混合而不是血液在粪便表面等)。内窥镜检查对于结直肠癌是首选的方法，结直肠癌晚期的肠镜鉴定相对简单，对于早期的结直肠癌很可能表现出细微的粘膜病变(Fijten et al., 1995)。粪便检测可能发现结直肠癌潜在的标志物(例如，大便中的血液或分子标志物)，两次检测为阳性的患者应进一步进行肠镜检测，愈创木脂测试(Guaiac-based Faecal Occult Blood Test, gFOBT)，这项测试通过化学试剂来检测大便隐血的方式相比十年前减少了约16%的大肠癌死亡，但此方法有对患者有药物和饮食的限制，且只根据颜色的变化判断因此有机会出现错误(Hewitson et al., 2008)。另一种常用的粪便检测的方式是免疫化学测试(Faecal Immunochemical Test, FIT)，这项测试是通过抗体来检测大便隐血，此方准确度较高，有研究表明通过此方法筛查结直肠癌死亡率降低了22%，FIT是欧洲有组织筛查的首选和最常用的方法，再荷兰的相对较高的参与率高达73%(Ventura et al., 2014)。此外，已开发出粪便DNA测试，此方法将DNA和FIT结合使用对比单独使用FIT对结直肠癌和晚期腺瘤的敏感性略高，但费用较高。用于结直肠癌筛查的非入侵方式还有血液检测，但是其灵敏度有待提高(Imperiale et al., 2014)。

越来越多证据表明，戒烟，健康饮食和定期运动可以预防结直肠癌的发展，每日活动至少30min，食用牛奶，全谷物，新鲜水果，坚果和蔬菜，以及摄入钙和纤维对预防结直肠癌(Baron et al., 2015)。

## 长链非编码RNA

### 非编码RNA

最近的研究已经发现20684个人类的编码基因，这些基因只占基因组的1.2%，但基因组中有80%是转录的(Consortium, 2012)。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)。这些数据使研究人员认为，非编码RNA可能不像之前预期的那样是无用的。非编码RNA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。

根据ncRNA的生物功能可将ncRNA分为两类：结构性ncRNA和调控型ncRNA。结构性ncRNA包括发现的tRNA，rRNA，snRNA，snoRNA，它们的特征是都为蛋白质的合成的“机器”。调控型ncRNA根据其核苷酸长度，小于200个碱基的ncRNA分为小ncRNA如：miRNA，piRNA，大于200个碱基的ncRNA称为lncRNA(long noncoding RNA)(Morris and Mattick, 2014)，它们的特征是特定的时空表达模式。这些ncRNA在生物过程中的重要性都所有提高，一些研究者认为RNA不仅作为DNA和蛋白质之间的信使，而且是细胞功能的关键参与者(Fatima et al., 2015)。高通量测序技术的发展，如微阵列和全基因组测序，使得研究者识别和鉴定基因组中非编码的部分，2015年发表的一项研究在人类基因组中已经确定了58648个lncRNA基因(Iyer et al., 2015)。

### LncRNA的功能

大多数lncRNA与mRNA相似，通过聚合酶Ⅱ转录，且大多数5’端加帽，也会发生多聚腺苷酸化和剪接。但平均来讲，lncRNA外显子数目较低，且在不同组织表达水平较mRNA更低(Derrien et al., 2012)。LncRNA的外显子基因组序列的保守性比蛋白编码基因的外显子弱但比内含子强，这表明lncRNA具有中等快速演化(Yue et al., 2014)。lncRNA在核酸序列保守性很差，尽管它们广泛存在各个物种如植物，酵母，原核生物，病毒(Guttman and Rinn, 2012)，但是lncRNA的核酸序列保守性很低，这也限制了我们对其功能的研究，尽管lncRNA序列保守性很低但是它们却可能有相同三位结构，因此它们可能会有相同的分子功能(Pang et al., 2006)。LncRNA序列保守性的一种解释是，物种越复杂那么lncRNA的数目越多，这一发现与lncRNA的高度组织特异性表达相结合，表明lncRNA很有可能是促进物种特异性特征和器官复杂性的关键分子，因此lncRNA很可能在复杂生命体的演化发挥了重要功能(Mercer et al., 2008)。LncRNA调控基因和基因组的活性。LncRNA在细胞核中最主要的功能是它们在不同水平上调控基因和基因组的活性，lncRNA通过表观遗传机制影响染色质的修饰和结构，进而影响转录或其他与可能与染色质相关的功能。

lncRNA参与组蛋白修饰：lncRNA与组蛋白修饰复合物的相互作用，例如多梳抑制复合物PRC1和PRC2，它们介导了组蛋白H3亚基上的第27位赖氨酸残基处甲基化，此组蛋白主要参与基因的下调相关。最先报道的与PRC2相互作用的lncRNA来自于哺乳动物X染色体失活的研究，此研究涉及到一个是X染色体失活的特性行转录本Xist，此转录本在雌性失活的X染色体高表达并召集PRC2使基因沉默(Plath et al., 2003 , Mercer et al., 2008)。Braveheart(Bvht)是心脏早期分化过程中被激活的lncRNA，在MESP1的上游起作用，Bvht还可以通过与PRC2的相互作用发挥功能(Klattenhoff et al., 2013)。LncRNAs是如何招募它们的蛋白质相互作用配偶到特定基因位点的一个有趣机制是形成DNA-RNA三联体，已经发现了好几种与PRC2形成复合物的几种lncRNA，通过这种方式lncRNA可能直接将染色质或转录因子引导至特定的基因组位点，这可以一定程度上解释染色质修饰复合物的位点的特异性作用，这些修饰物本身并不是以序列特意性方式与DNA结合。如位于1号染色体上的lncRNA(FAL1)，FAL1可以与PRC1的一个重要亚基BMI1相互作用，调节蛋白质的稳定性，卵巢癌细胞系中FAL1的敲低导致基因变化，降低细胞周期进程并诱导衰老，效果和BMI敲低类似(Pasmant et al., 2007)。如WDR5与小鼠胚胎干细胞中超过200个lncRNA相互作用，这些相互作用对于WDR5与染色质结合十分重要，因为WDR5与RNA结合位点的点突变导致WDR5未能与染色质稳定的结合(Trievel and Shilatifard, 2009)。总而言之，参与组蛋白修饰是lncRNA发挥功能的重要方式。

LncRNA参与DNA甲基化的调控：DNA甲基化是哺乳动物有机体中关键的表观遗传标记，在染色质组织和基因表达中起关键作用，通常基因启动子区域的甲基化通常与阻遏有关，但有研究表明，基因组中基因间区的DNA甲基化可能导致不同的基因调控模式。DNA甲基化是一个动态且可逆的过程，可控制发育和疾病期间的基因表达，研究表明lncRNA(TARID)通过诱导TCF21启动子的去甲基化激活其表达(Arab et al., 2014)。

LncRNA参与染色质重塑：除了参与共价修饰染色质相关酶的复合物等相互作用外，lncRNA还被证明能调控可以染色质重塑复合物。如在人类前列腺癌中，基因表达分析表明，SWI/SNF和SCHLAP1(lncRNA)具有相反的功能，SCHAP1可以和染色质重塑复合物SWI/SNF的亚基SNF5相互作用，并在大范围内抑制SWI/SNF和DNA的结合，最后导致基因的活性的降低(Prensner et al., 2013)。

LncRNA与转录因子的相互作用：横纹肌肉瘤相关的lncRNA，RMST(rhabdomyosarcoma associated transcript，RMST)是神经元分化所必须的，其通过与转录因子SOX2大概一般的结合位点结合接到其调控(Ng et al., 2013)。再例如lnc-DC(lnc dendritic cells)可以和转录因子STAT3结合，组织其对SHP1的磷酸化，从而激活STAT3参与树突状细胞的分化(Wang et al., 2014)。

CeRNA(competing endogenous RNA)：竞争内源RNA假说早期的miRNA研究集中在目标转录本的单向调节，miRNA序列和启动子在人和小鼠之间高度保守，但随着对miRNA靶向机制的研究，相互调节的概念逐渐发展。由于每个miRNA可以靶向数百或数千个基因，相似的一个转录本上也可能纯在多个miRNA响应位点，包含同一个响应位点的转录本彼此也可以相互结合(Salmena et al., 2011)。竞争性miRNA结合最先是合成的miRNA在体内体外观察到的，miRNA可以有效的抑制其各自靶标转录本，此后，在植物中描述了第一个ceRNA(Franco-Zorrilla et al., 2007)。很多证据表明lncRNA和miRNA是骨肉瘤(OS)中关键的调节剂，Rho相关的卷曲螺旋蛋白激酶(ROCK1)，ROCK1是一个与转移相关的被miR-335-5p负调控癌的基因，DANCR(lncRNA DANCR)调控ROCK1通过与miR-335-5p和miR-1972相互作用介导细胞增殖和转移。LncRNA GAS5通过竞争性抑制miR-222-3p对PTEN的抑制作用来抑制CRC中细胞迁移和侵袭并促进自噬(Liu et al., 2019)。

### lncRNA与结直肠癌

lncRNA可以作为肿瘤发生中的癌症生物标志物，其优势主要有：

(1)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)。

## 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。

### AOM/DSS小鼠模型组织病理学

### AOM/DSS小鼠模型分子特征

研究表明，在分子水平上AOM/DSS小鼠模型中的腺瘤和人类结直肠癌非常相似。Wnt信号通路通常在胚胎发生过程中抑制细胞分化，并参与小肠隐窝细胞的胚胎发育后调控，在异常激活时参与肿瘤形成。β-catenin(β-连环蛋白)是此信号通路中重要的分子，也是钙粘蛋白介导细胞黏附系统的关键组成部分，β-catenin参与结直肠癌隐窝细胞的生长，发育和分化相关基因的调控通过与细胞核中的T细胞因子(Tcf-4)结合，而β-catenin的突变是小鼠结直肠癌模型的早期事件，这一点与人类结直肠癌一样，在大多数结直肠癌患者都观察到β-catenin的突变。

# 材料与方法

## 实验材料

## 主要生物信息软件

1. RNAseq reads比对软件 HISAT2
2. sam文件处理软件 Samtools
3. bed文件处理软件 bedtools
4. RNAseq定量软件 featurecount
5. lncRNA分类软件 FEELnc
6. lncRNA分析软件 Annolnc2
7. R语言包：tidyverse

## 实验方法

### AOM/DSS小鼠模型构建

### RNA-Seq数据分析

本研究获得原始测序数据后都在本地工作站进行分析，按照常规的转录组分析流程，本次研究的具体分析流程如下：

1. 测序数据质控：使用FastQC v0.11.9对获得的文件格式为fastq的测序数据进行质量检测，检测的内容包括：测序数据的基本信息，reads上碱基和reads序列的质量值， reads碱基组成及N含量，reads长度，K-mer等等信息。通过MultiQC v1.9对FastQC的结果形成综合性报告，如果测序数据质量不高可使用fastp v0.21.0对测序数据进行过滤和质控，需对fastp处理后的fastq文件重新使用FastQC质量检测。
2. 测序数据比对到参考基因组：使用Hisat2 v2.0.5将测序数据比对到参考基因组，本研究使用的小鼠参考基因组为Ensembl v99版。设置为双端测序，其他均为默认参数。比对完成后使用samtools v1.7将比对后的sam文件进行排序并压缩成二进制的bam文件，并创建bam文件的索引，可用于IGV可视化。转换完成后使用samtools flagstat统计比对率。
3. 测序数据reads计数：基因组的注释文件使用的是Ensembl v99，lncRNA的注释文件使用的是NONCODE v5，保留Ensembl中所有的蛋白编码基因，去除NONCODE中和Ensembl中蛋白编码基因在同一条链上存在重叠的基因，再从Ensembl中非编码部分中去除和NONCODE中存在重叠的部分，最后合并处理后的Ensembl注释文件和NONCDOE注释文件，作为本研究的基因注释文件，此操作涉及到的工具有bedtools v2.27.1，python脚本，shell脚本。定量工具使用R语言版本的featurecount进行计数，设置两个参数isPairedEnd=TRUE，requireBothEndsMapped = TRUE，其他的为默认参数。定量后使用abundance\_estimates\_to\_matrix.pl脚本对每个样本的定量结果进行合并生成表达矩阵同时计算FPKM及TMM标准化。

### LncRNA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。

### 主成分分析和差异表达分析

主成分分析(Principal components analysis, PCA)分析是一种用于统计分析简化数据集的方法，经常用于减少数据集的维数的同时保留数据集中对方差贡献最大的特征。本研究使用的PCA分析方法使用的是PCAtools v2.0.0，基因的表达矩阵标准化使用的是TMM标准化。差异表达分析首先使用DESeq2 v1.28.1分析，因为DESeq2有其特有标准化算法，因此使用原始定量矩阵。差异表达的筛选条件为满足p.adjust<0.05同时Log2FoldChange的绝对值大于2的基因为差异表达。

### 基因功能富集分析

本研究中用到的功能富集分析方法主要是GO和KEGG功能。GO(基因本体论，The Gene Ontology)，clusterProfiler v3.16.1 。。。。。。

### 时间序列分析

针对某一个条件下不同样本之间的差异表达分析无法很好的研究生物学动态的过程，而时间序列分析可以有效的关注基因的整体表达趋势，尽管时间序列分析算法没有差异表达分析那样成熟，但是其可以从更好的维度去阐释发育，炎症的动态的变化的生命过程。本研究使用的时间序列分析的方法为maSigPro v1.60.0，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>0.7的基因被归类为直肠癌发生过程中动态表达的基因。maSigPro根据设置的阈值首先鉴定随时间具有明显变化的动态表达的基因，接着对动态表达的基因按照其表达模式分为9个cluster，每一个cluster中的基因具有相似的表达模式，但这种相似不具有统计学意义。

### 构建mRNA-lncRNA共表达网络

为了找到和结直肠癌发生相关的动态lncRNA，首先对maSigPro得到的具有相似表达模式的9个cluster中的动态表达的蛋白编码基因分别进行KEGG富集分析，寻找可以富集到结直肠癌通路的cluster，富集到结直肠癌通路上的基因被认为是关键的动态基因，与关键的动态表达基因在同一个cluster的lnc基因被认为其表达模式与关键的动态基因相同，因为基于maSigPro的这种相似性判断没有统计学意义，因此使用Himsc v4.4.1计算所有基因之间相关性，由于此过程计算量较大本地电脑无法满足，需要在生科院集群中计算，基因之间R2绝对值大于等于0.8，pvalue小于0.05的基因对被认为是存在共表达，最终得到的mRNA-lncRNA对存在于同一个maSigPro所得到的cluster中，则这些lncRNA被认为是和结直肠癌相关的动态编码基因共表达的。通过Cytoscape v3.8.0构建动态mRNA-lncRNA共表达网络。

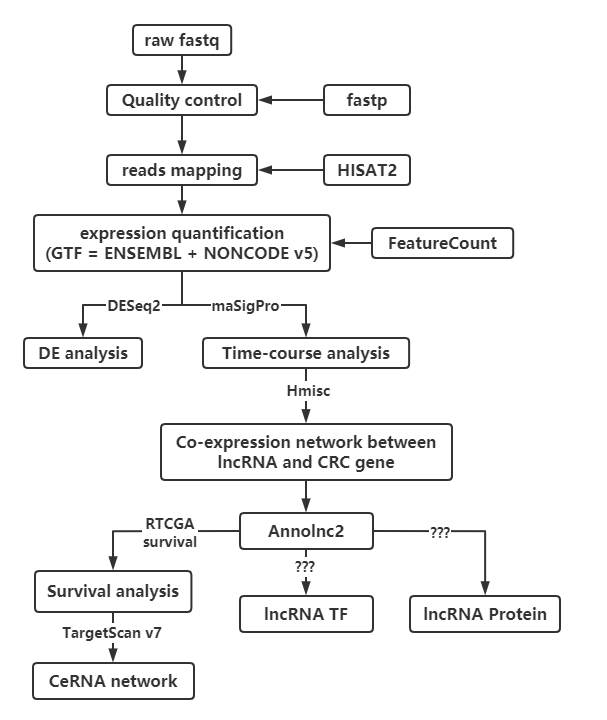
### LncRNA的系统性分析

取与结直肠癌相关的动态表达的关键基因共表达lnc基因的最长转录本进行分析，分析方法使用Annolnc2，用到的工具有python脚本，gffread v0.11.7，在本地服务器搭建好Annolnc2的分析环境，安装基础分析模块，从官网上下载本次研究需要用到的分析模块：mouse\_transcription\_regulation.tar.gz，mouse\_evolution.tar.gz，mouse\_miRNA\_interaction.tar.gz。解压后放入相应的路径供Annolnc2调用，Annolnc2提供一站式分析lncRNA服务，分析包括结构，表达，调控，遗传关联和进化。值得注意的是，本研究在分析过程中发现在线网页版本Annolnc2的分析结果和本地版的结果有些不一致，询问了作者之后得到的回复是在线网页版本由于计算量过大服务器负荷因此会有错误发生，并且有些分析由于计算机CPU的性能差异预测的结果也会不完全一样，作者表示他们正在解决这个问题。运行Annolnc2使用默认参数，得到的结果在R语言里整理分析。

### 构建ceRNA网络

根据miRNA结合lncRNA的靶标预测位点进行分析，认为结合在lnc环上且保守得分在60分以上的miRNA更大可能发挥功能，因为RNA中环的可接近性更大，本研究对所有的候选lnc基因的最长转录本进行分析，通过TargetScan v7.1数据库提取这些miRNA的靶基因，对miRNA靶基因进行KEGG功能富集，筛选到可以靶标基因可以富集到结直肠癌通路上的miRNA，认为这些lncRNA与miRNA互作网络可能参与结直肠癌的发生过程。为了进一步分析那些miRNA与结直肠癌患者生存状态存在相关性，由于miRNA的同源保守性，本研究对筛选得到的miRNA通过miRbase v22.1将小鼠的miRNA同源注释到人类的miRNA，通过TCGA数据库提供的结直肠癌患者的miRNA表达数据及临床数据进行生存分析。生存分析在R中用到的工具有，RTCGA v1.18，survival v3.1.12，得到和结直肠癌患者生存显著相关的miRNA，pvalue小于0.05。得到最终的关键miRNA反推可能结合的lncRNA，通过Gephi v0.9.2绘制CeRNA网络图。

## 技术流程图

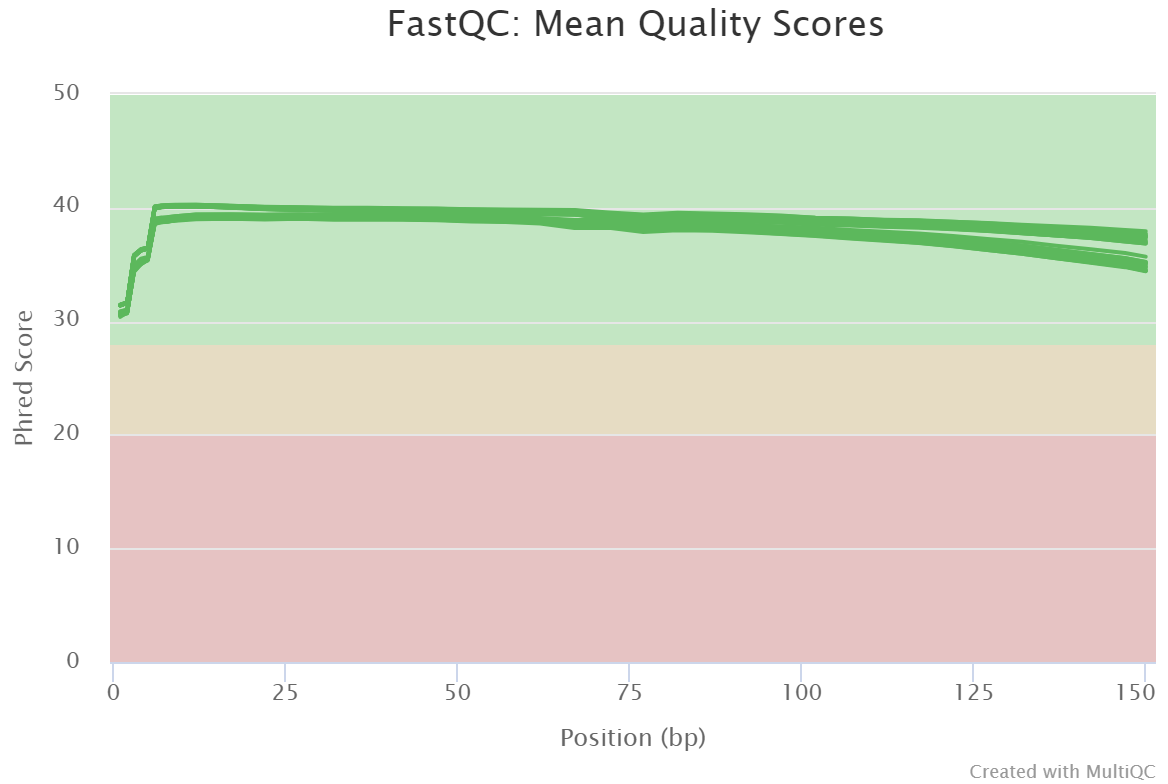


# 结果与分析

## 测序数据质量及比对率

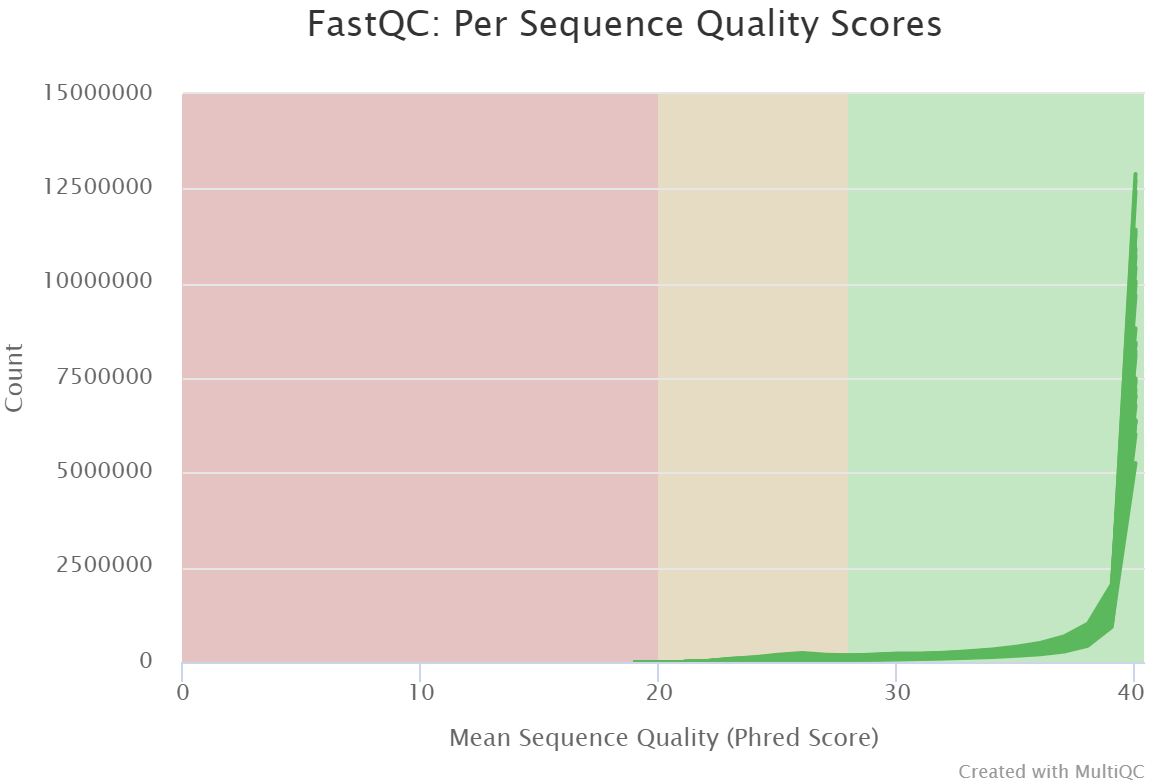
得到的原始测序数据首先用FastQC进行质量检测后发现需要进一步过滤，使用fastp对测序数据过滤后再次进行质量检测，对所有样本的检测结果使用MultiQC进行可视化，如图展示的是具有代表性的部分结果，本研究的测序数据每个读长上的每个碱基的质量分数都在30以上，绝大部分读长的质量分数在40表明此次测序数据准确度很高，每个读长的GC含量基本符合正态分布，说明建库测序过程中污染较小，且PCR偏差较小，这所有结果表明本研究的数据质量可靠，符合更进一步分析的要求。

使用samtools得到每个样本的比对率，如图所示每个样本的统计结果，可以看到所有样本的比对率均大于80%，表明每个样本中超过80%的reads是可以很好的比对到参考基因组，进一步说明本研究的测序数据质量很高，符合实验预测。



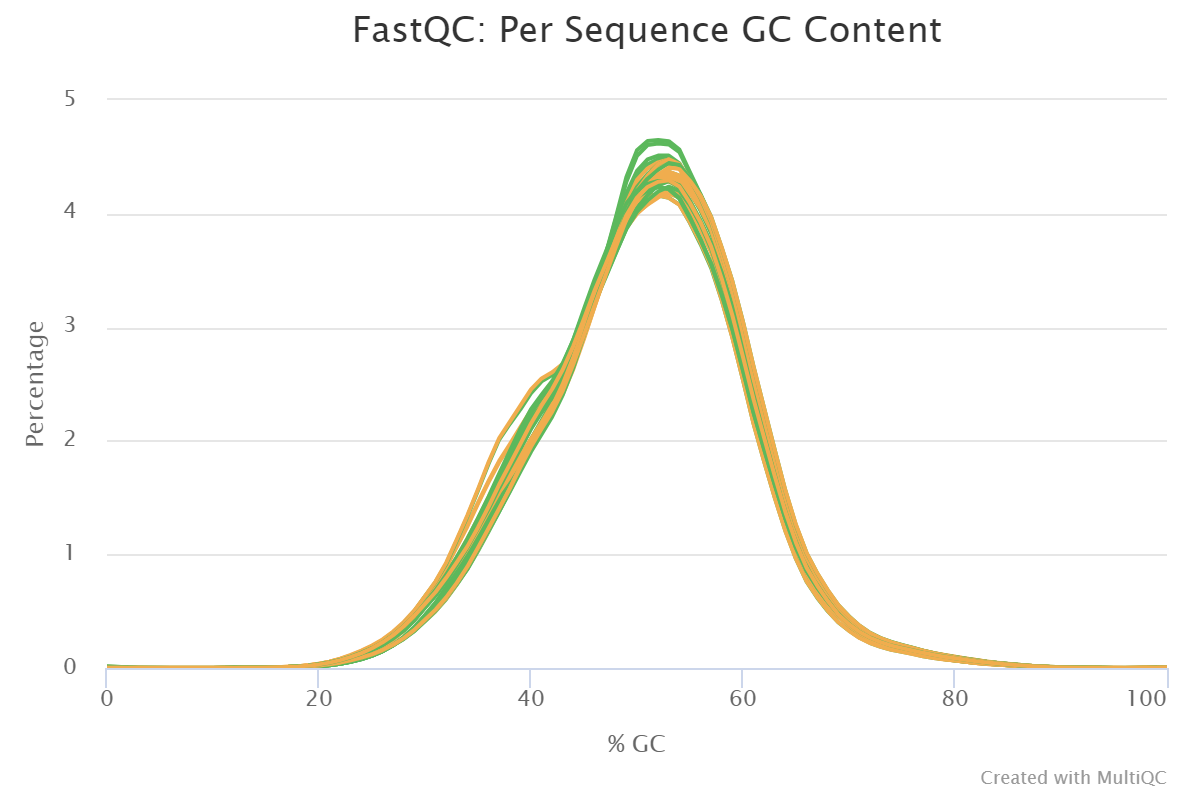
**图 3‑1 每个reads上所有碱基的平均质量分数**

**Fig. 3‑1 The mean quality value across each base position in read**



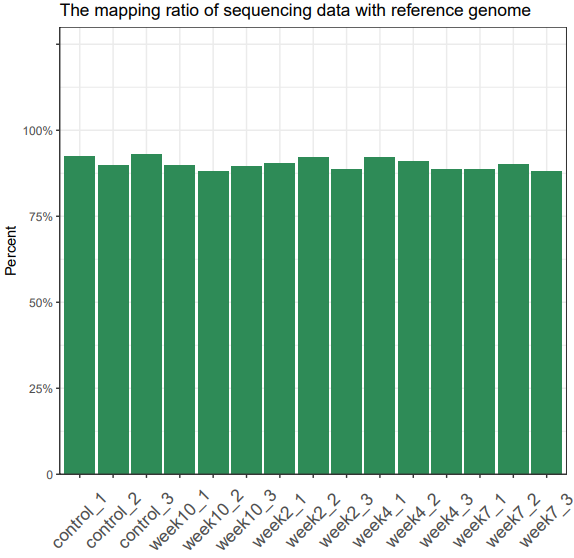
**图 3‑2 每个序列的质量得分**

**Fig. 3‑2 Per Sequence Quality Scores**



**图 3‑3 每个序列的GC含量**

**Fig. 3‑3 GC content of each sequence**

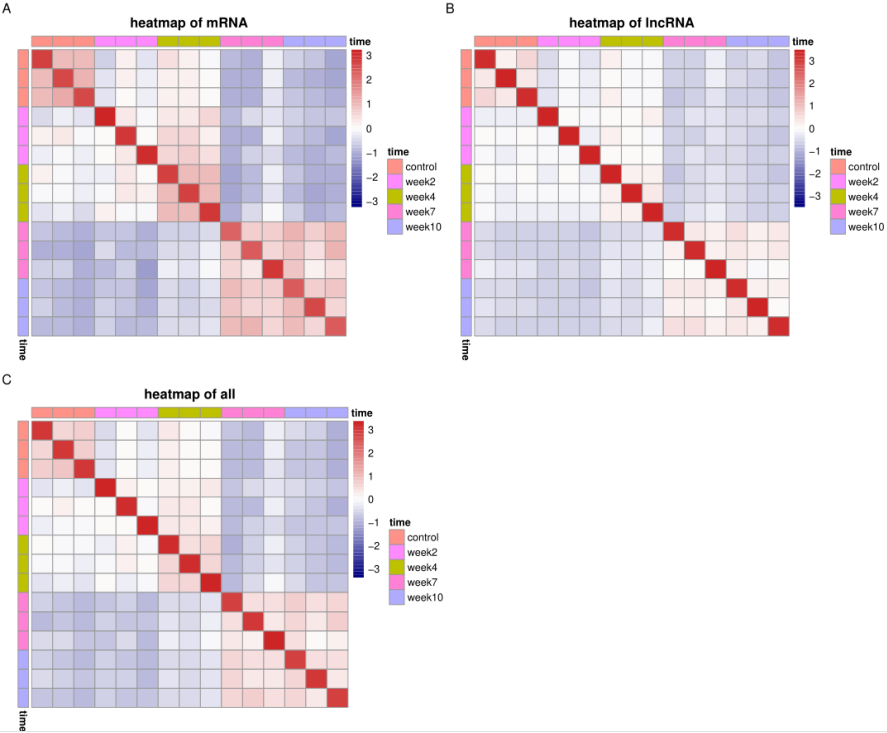


**图 3‑4 每个样本比对到参考基因组上的比对率**

**Fig. 3‑4 The mapping ratio of each sample to the reference genome**

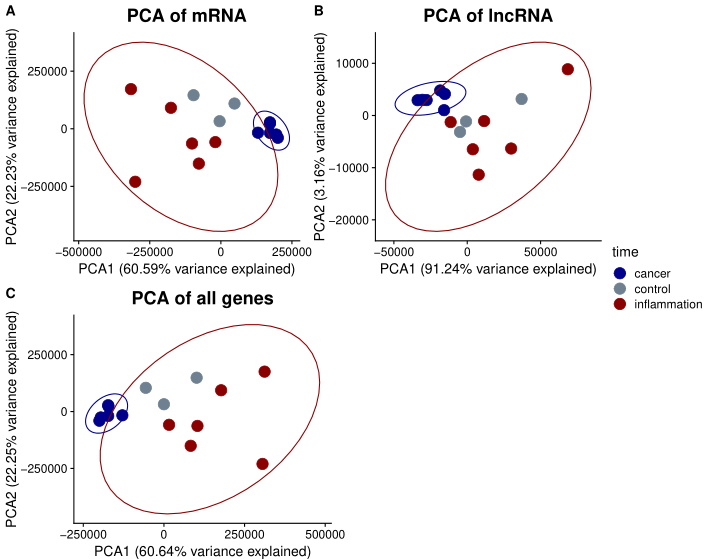
## 样本间相关性分析

为了验证样本之间是否具有良好的一致性，根据基因在每个样本的表达量计算样本间kendall相关系数并通过ComplexHeatmap可视化，如图3-5所示：不论是在所有基因水平，mRNA水平还是lncRNA水平，同一时期的样本之间的相关性大于不同时期之间的相关性，第七周和第十周样本之间的相关很高且样本间相关性在第四周和第七周区分开，正好符合AOM/DSS结直肠癌小鼠模型中这段时期由炎症转变成癌症的阶段。接着通过主成分分析（PCA），PCA分析通过对数据进行降维，将复杂多维数据描述为两个维度的差异，根据样本聚类的结果将第二周和第四周归为炎症期，将第七周和第十周定义为癌症期。如图3-6结果所示，不论是所有基因水平，mRNA水平，还是lncRNA水平癌症期样本聚在一起且与对照和炎症期分开。总之，通过计算样本间相关性及主成分分析说明本研究中每个时期的样本之间统计学上的相关性符合实验设计和生物学意义，并且第四周到第七周是由炎症转变为癌症的时期，符合之前的报道。



**图 3‑5 样本间聚类分析**

**Fig. 3‑5 The results of the sample clustering**



**图 3‑6 PCA分析**

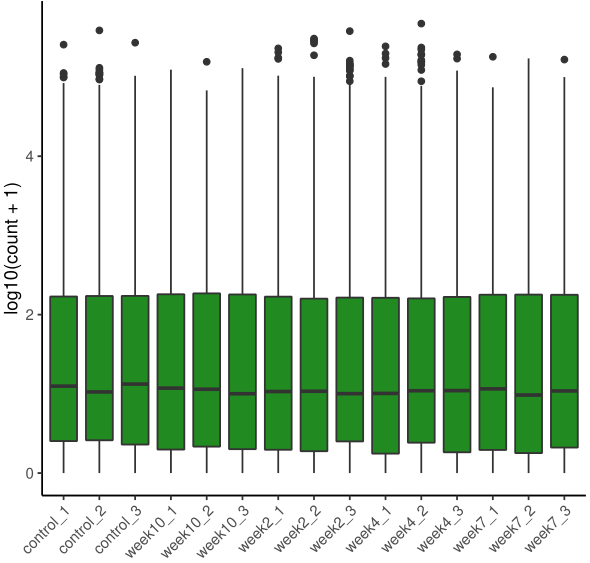
**Fig. 3‑6 PCA analysis**

## 差异表达分析

使用DESeq2分别对第二周，第四周，第七周，第十周相较于对照组分别做基因差异表达分析，p.adjust小于0.05且Log2FoldChange绝对值大于2判定为差异表达的基因，Log2FoldChange大于2的基因判定为相较于对照组上调表达，Log2FoldChange小于-2的基因则判定为相较于对照组下调表达，都不符合上述条件的基因判定为没有差异表达。

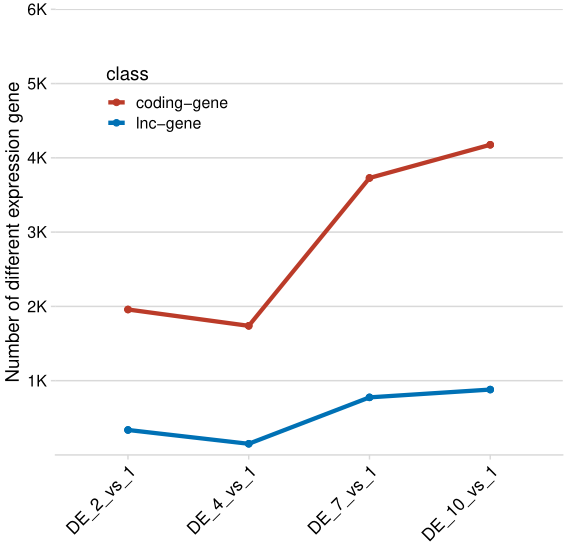
如图所示，差异表达的蛋白编码基因个数从第二周到第十周lncRNA基因均大于lncRNA基因，这点是符合预期的。且两者差异表达基因的个数变化趋势比较相似。如图所展示的具体某个时期相较于对照组差异表达基因上调下调的分布图，为了展示方便仅展示第四周和第七周差异基因的分布，图中的每一个点代表一个基因，红色代表上调，蓝色代表下调，灰色代表差异不显著或差异倍数小于2。可以观察到不论差异表达的蛋白编码基因还是差异表达的lncRNA基因上调的数目大于下调的。

使用R包clusterProfiler对差异表达的蛋白编码基因进行GO和KEGG功能富集分析如图所示，红色部分GO-cc为GO中细胞组分富集分析，蓝色部分为GO中生物过程的富集分析，黄色部分为KEGG通路分析，颜色深浅代表显著性，横轴表示基因个数(log转换后的)。。。。。。



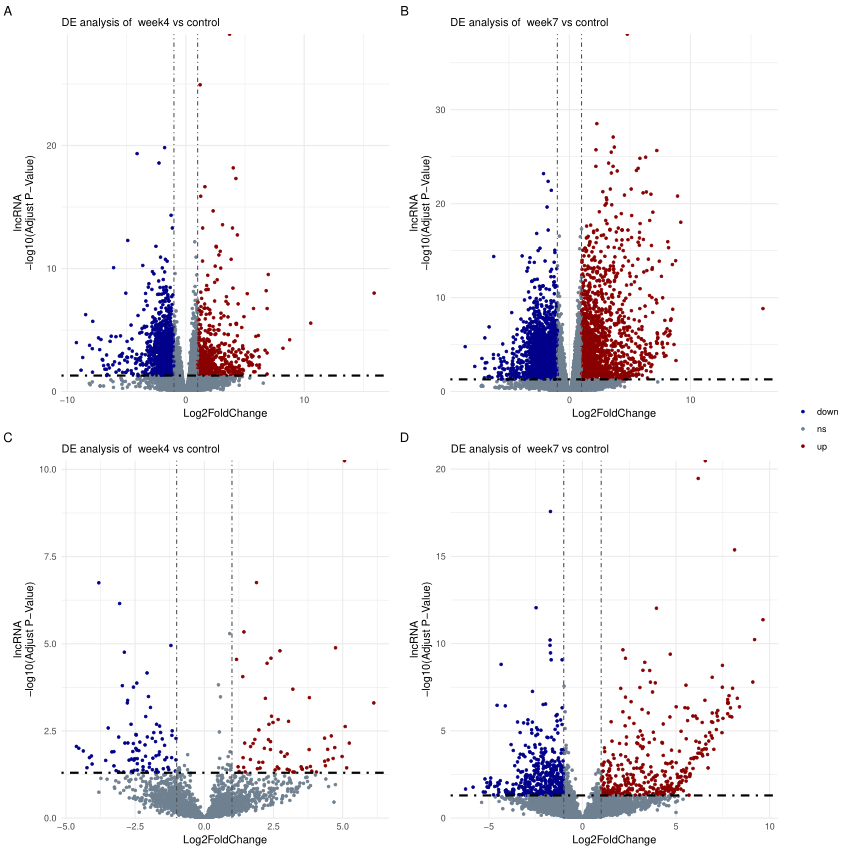
**图 3‑7 样本的表达量**

**Fig. 3‑7 Expression level of each samples**



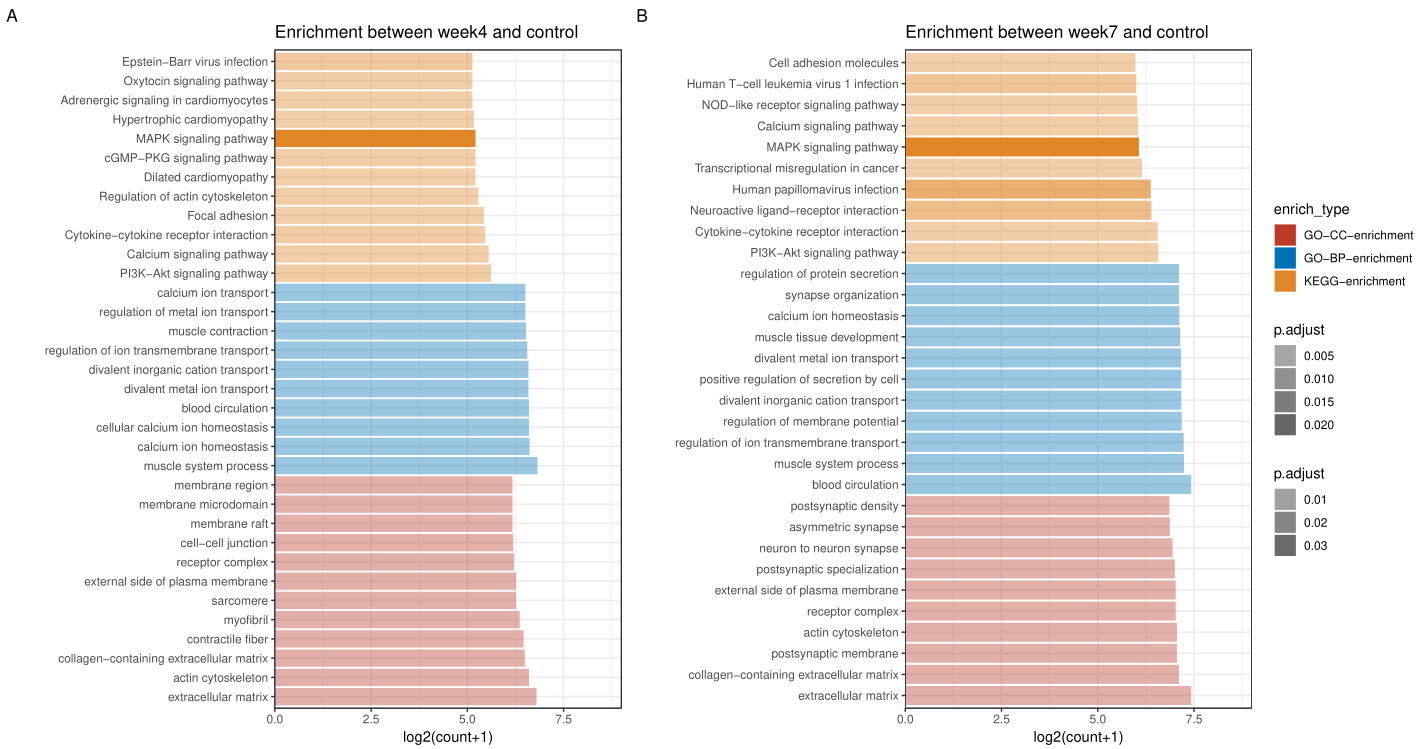
**图 3‑8 每个时期差异表达的基因个数**

**Fig. 3‑8 The num of different expressed genes in each period**



**图 3‑9 第四周和第七周差异表达基因上下调分布**

**Fig. 3‑9 Different expressed genes in week4 and week7**



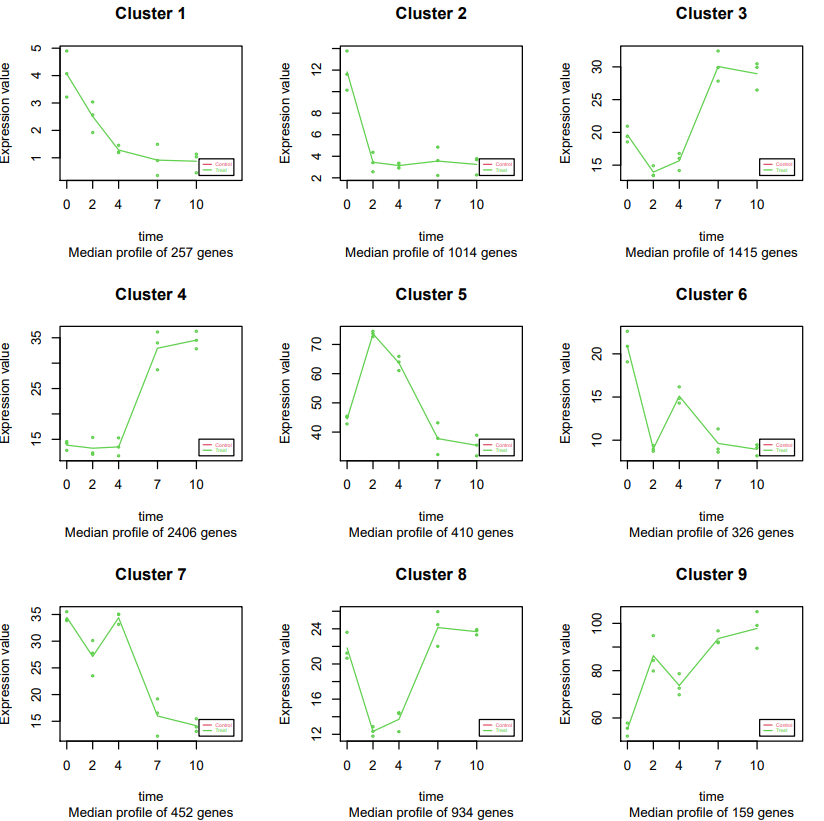
**图 3‑10 第四周和第七周差异表达基因的GO和KEGG功能富集分析**

**Fig. 3‑10 GO and KEGG enrichment analysis of DE genes in week4 and week7**

## maSigPro时间序列分析

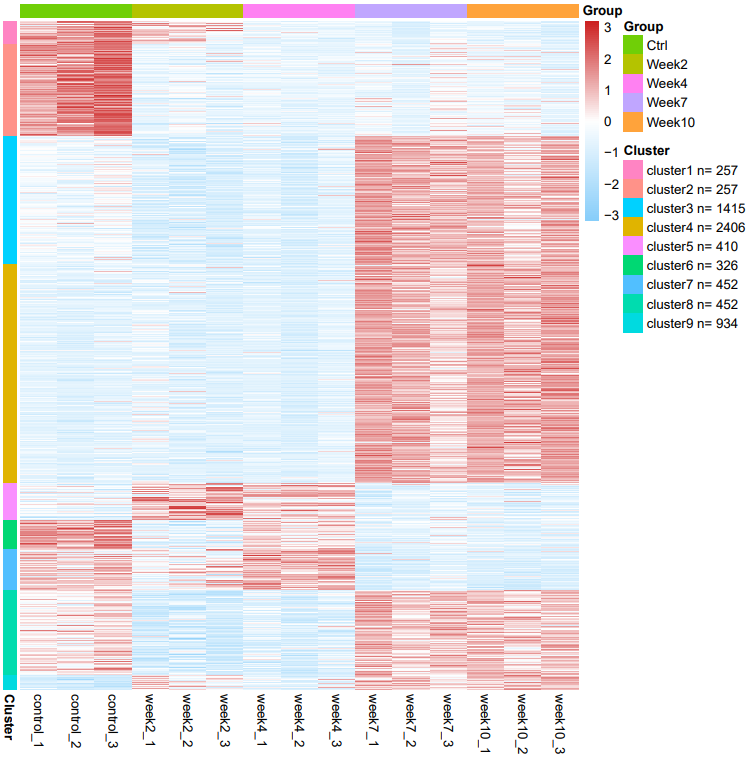
### 鉴定动态表达基因

将蛋白编码基因和lncRNA基因的原始矩阵通过edger进行CPM标准化后用于maSigPro时序分析，最终鉴定到976个动态表达的lncRNA基因，6364个动态表的蛋白编码基因。并根据这些动态表达基因的表达模式分类成9个cluster，如图所示，换句话说每个cluster中的动态编码基因和动态lncRNA基因具有相同的表达模式，每个cluster的横轴代表不同的时期，纵轴代表基因的表达水平，可以直观整体的看到每个cluster中基因表的趋势，如cluster1中的基因到第四周都表达水平下降，第四周以后表达水平平缓等等。如图所示，展示的是cluster中的基因在和每个时期的表达水平，颜色约接近红色代表基因的表达量越高，反之越低，可以看出同一个cluster中的基因在不同时期表达水平接近，表明不同的cluster之间存在区分度。如图所示展示的每个cluster中蛋白编码基因和lncRNA基因所占的百分比，可以观察到动态表达的lncRNA的个数比蛋白编码基因少。



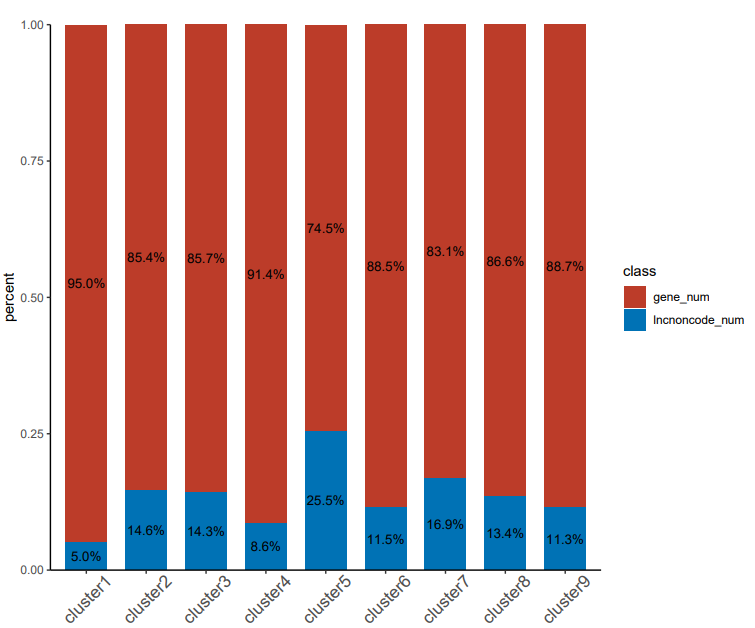
**图 3‑11 基于maSigPro的时间序列分析**

**Fig. 3‑11 Time-course analysis based on maSigPro**



**图 3‑12 不同时期不同cluster基因的表达水平**

**Fig. 3‑12 Expression level of genes in different period and cluster**



**图 3‑13每个cluster中基因的组成**

**Fig. 3‑13The composition of clusters**

### 动态表达的lncRNA基因特征

如图所示比较了动态表达的lncRNA，非动态表达的lncRNA，蛋白质编码基因在表达水平转录本长度，外显子个数的差异，可以观察到不论从哪个角度来看，动态表达的lncRNA都显著大于非动态的lncRNA，同时也显著小于蛋白质编码基因。这表明动态表达的lncRNA具有更多的选择剪接位点和拥有更多的容纳RNA的能力，从而更易于与蛋白质核酸相互作用参与生命过程，因此本研究将研究结直肠癌中的lncRNA聚焦在这些动态表达的lncRNA。



**图 3‑14动态表达lncRNA基因的特征**

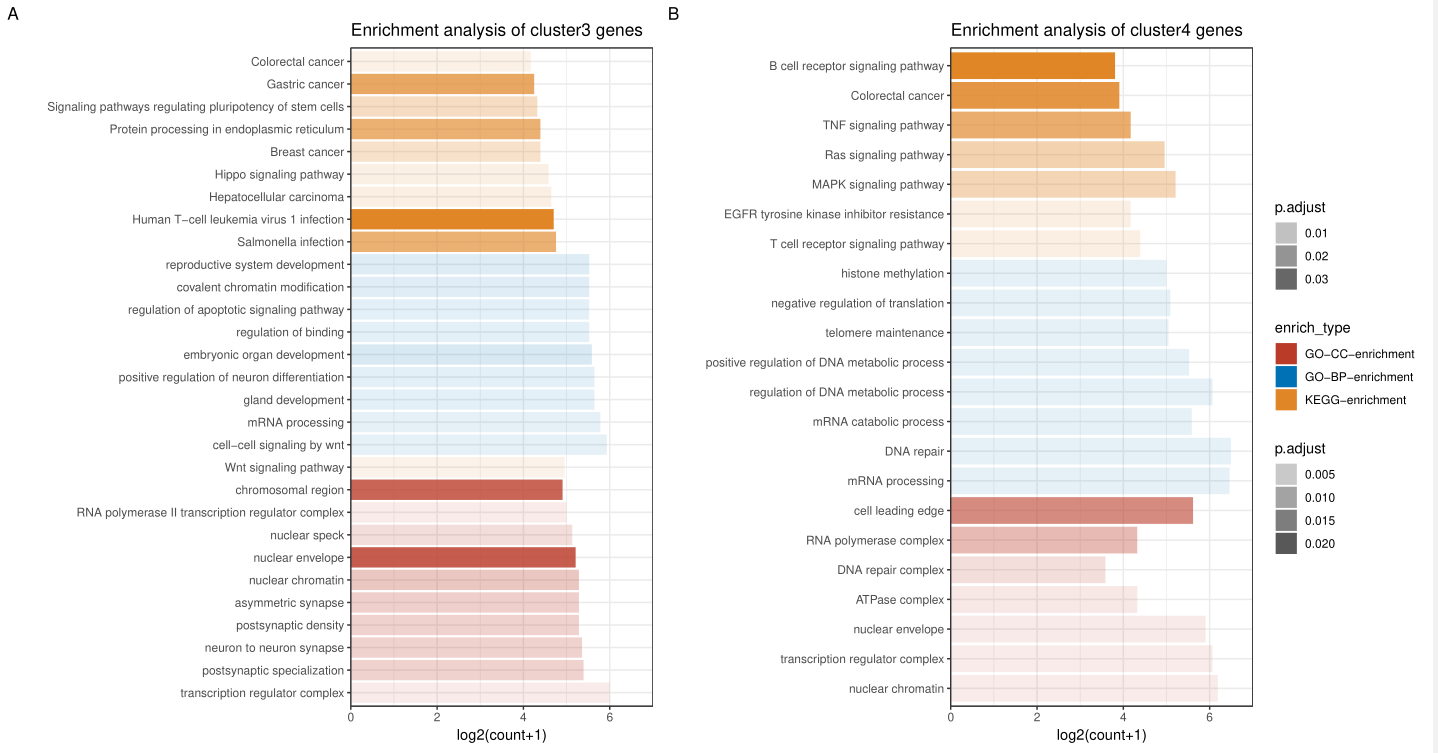
**A.动态表达的lncRNA与非动态表达的lncRNA基因及蛋白质编码基因的表达水平；B.动态表达的lncRNA与非动态表达的lncRNA基因及蛋白质编码基因的转录本长度；C.动态表达的lncRNA与非动态表达的lncRNA基因及蛋白质编码基因的外显子个数**

**Fig. 3‑14 The features of dynamic lncRNAs**

**The comparison of expression level(A), transcript length(B), exon number(C) of dynamic lncRNAs, Non-dynamic lncRNA and coding gene**

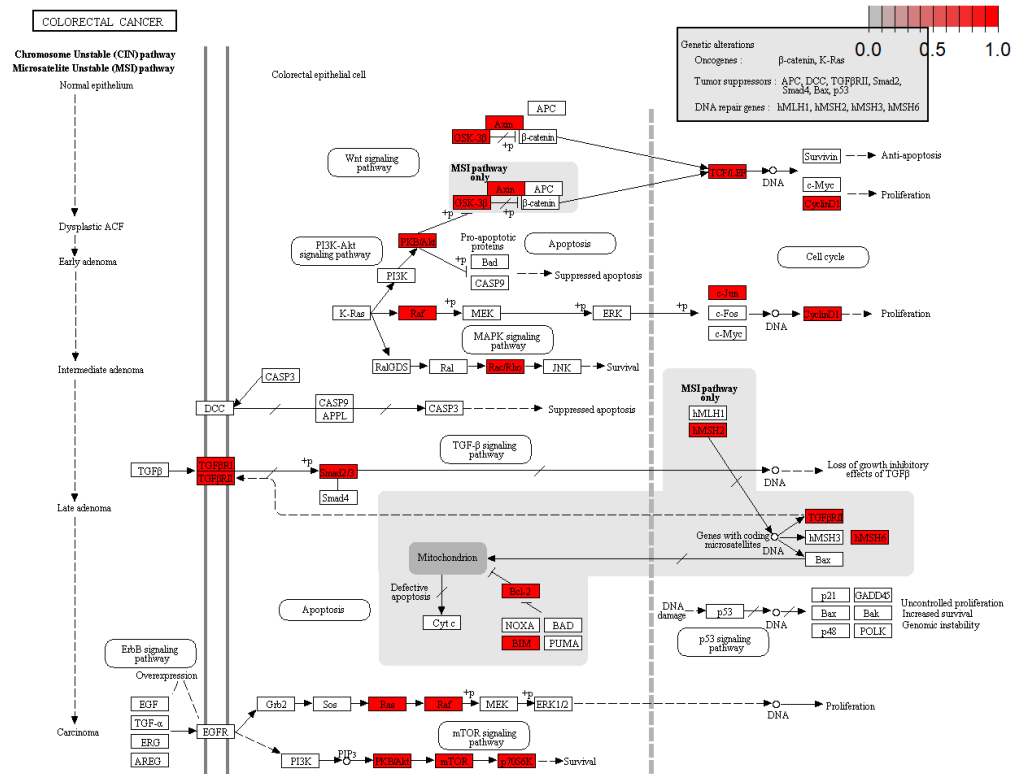
### 寻找与结直肠癌发展相关的动态基因

首先为了确定maSigPro最终得到的9个cluster中那些cluster中的动态表达基因与结直肠癌相关，本研究分别对这9个cluster通过GO和KEGG功能富集分析，有趣的是发现cluster3和cluster4中的动态表达的基因直接富集到结直肠癌通路，如图所示，此外也富集到了很多在文献中报道的与结直肠癌发展相关的通路。本研究的关注点聚焦这些和结直肠癌相关的动态表达的基因，通过KEGG富集分析的结果提取后使用Pathwayview在通路上标注这些基因的位置如图所示，可以观察到动态表达的基因广泛的分布在结直肠癌的通路各个途径，为进一步探究这些动态表达基因之间的关系，使用GeneMANIA工具对基因之间的关系进行分析，如图所示展示了这些动态表达基因的共表达关系，互作关系，及通路关系，进一步说明它们的功能是密切相关的，本研究后续的研究集中在寻找与这些基因共表达的动态表达的lncRANs。



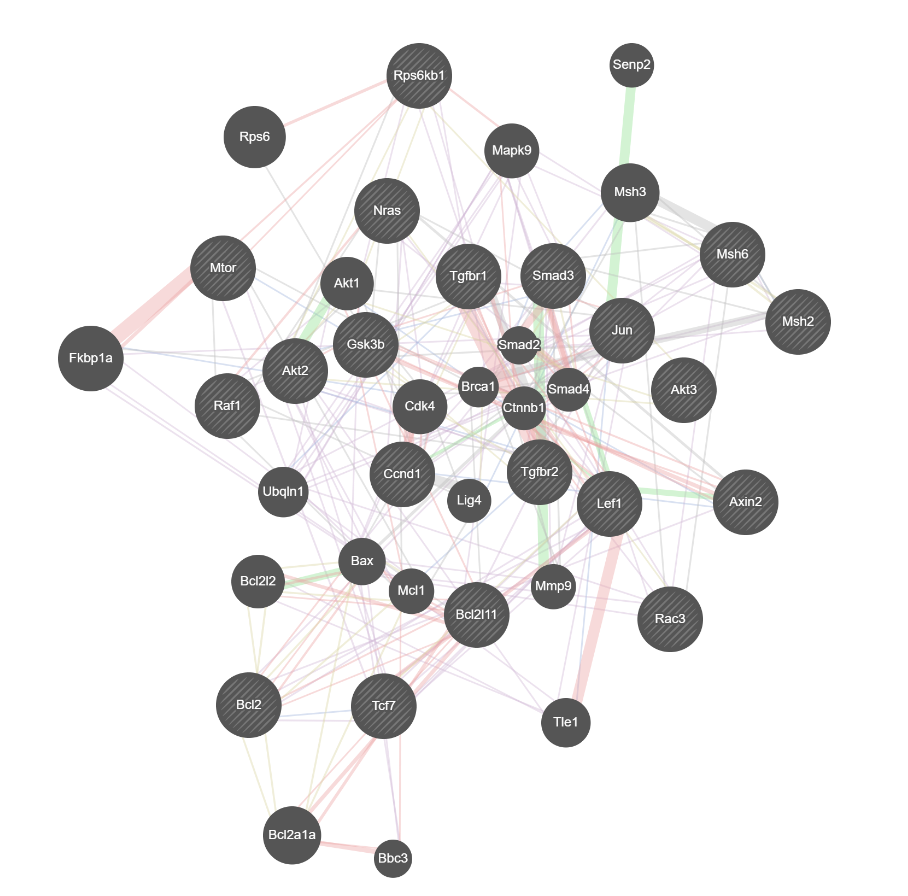
**图 3‑15 cluster3和cluster4中动态表达基因的GO和KEGG功能富集分析**

**Fig. 3‑15 GO and KEGG enrichment analysis of dynamic expression genes in cluster3 and cluster4**



**图 3‑16结直肠癌通路中动态表达的基因**

**Fig. 3‑16 Dynamic expression genes in Colorectal Cancer pathway**

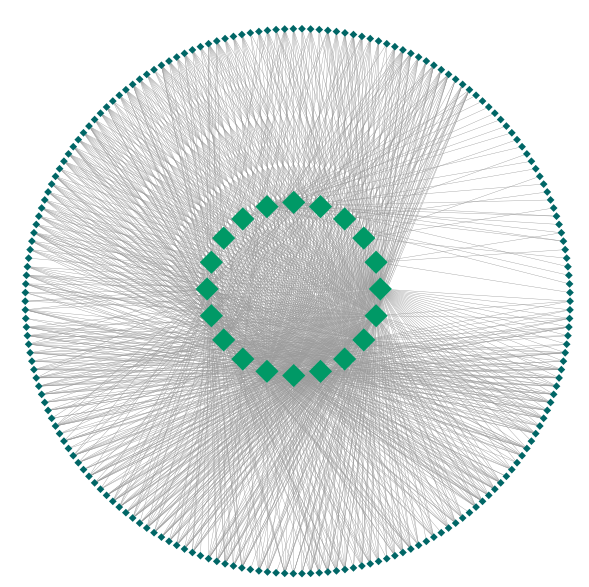


**图 3‑17 结直肠癌通路中动态表达基因的GeneMANIA分析**

**Fig. 3‑17 GeneMANIA analysis of dynamic genes in Colorectal Cancer pathway**

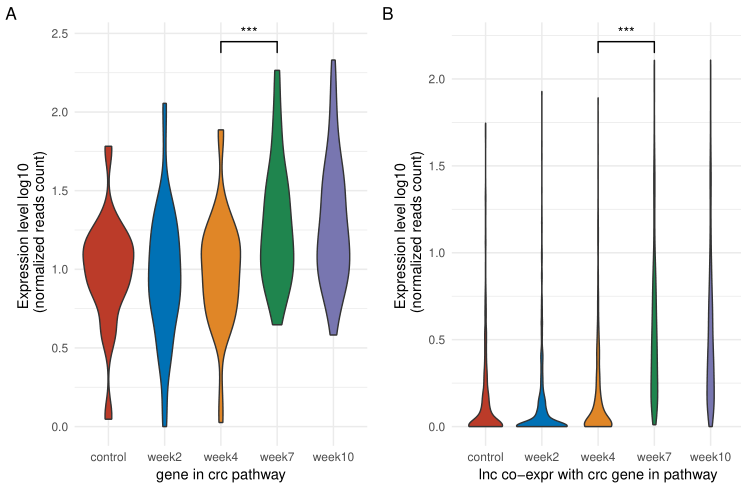
## 构建mRNA-lncRNA共表达网络

使用R包Hmisc基于pearson算法计算表达矩阵中基因之间的相关性，R2的绝对值大于等于0.8且pvalue小于0.05的被定义为基因之间存在相关性，从中筛选出同时出现在cluster3或者是cluster4中且与结直肠癌相关的动态表达的基因存在相关性的lncRNA，因为这些lncRNA和mRNA处于同一个cluster表明它们的表的模式相似且通过pearson筛选，最终认为这些mRNA和lncRNA是共表达的。最终找到178个动态表达的lncRNA基因和20结直肠相关的动态表达的编码基因共表达，认为这些lncRNA基因潜在的参与了结直肠癌的发展，基于它们之间的R2在cytoscape中绘制mRNA-lncRNA共表达网络如图所示，中间方块表示结直肠癌相关的动态表达编码基因，四周表示与其共表达的lncRNA基因。如图所示，统计了共表达网络基因的表达水平，lncRNA基因表达水平很低，到第七周显著上升并保持稳定，而蛋白编码基因的表达水平保持在较高水平在第七周显著提升同样维持稳定，这个趋势与maSigPro分类一致，说明这些动态表达的基因很有可能在这个第四周到第七周这个时期内发挥作用，而AOM/DSS小鼠模型中第四周到第七周正好对应到炎症向癌症转化的过程，因此推测这些与蛋白编码基因共表达的lncRNA基因很有可能与结直肠癌炎症及炎症转为癌症相关，这也为本研究后续分析和功能解释提供线索。



**图 3‑18 mRNA-lncRNA共表达网络**

**Fig. 3‑18 Co-expression network of mRNA-lncRNA**



**图 3‑19 共表达基因的表达水平**

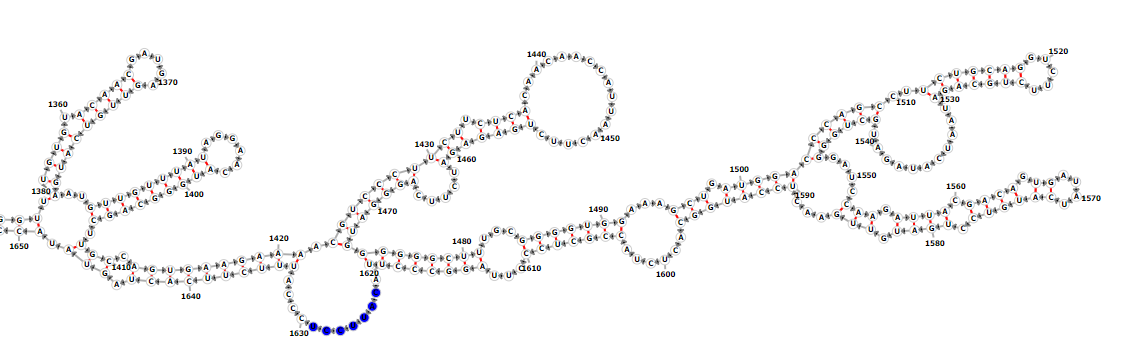
**Fig. 3‑19 Expression levels of co-expression genes**

## ceRNA调控网络构建

取共表达网络中lncRNA基因的最长转录本作为lncRNA基因对象基于Annolnc2做后续分析，为了探索与结直肠癌相关的lncRNA基因，本研究首先考虑的ceRNA调控网络，及lncRNA竞争性结合miRNA从而影响miRNA和mRNA的互作。

### lncRNA与miRNA结合靶位点预测

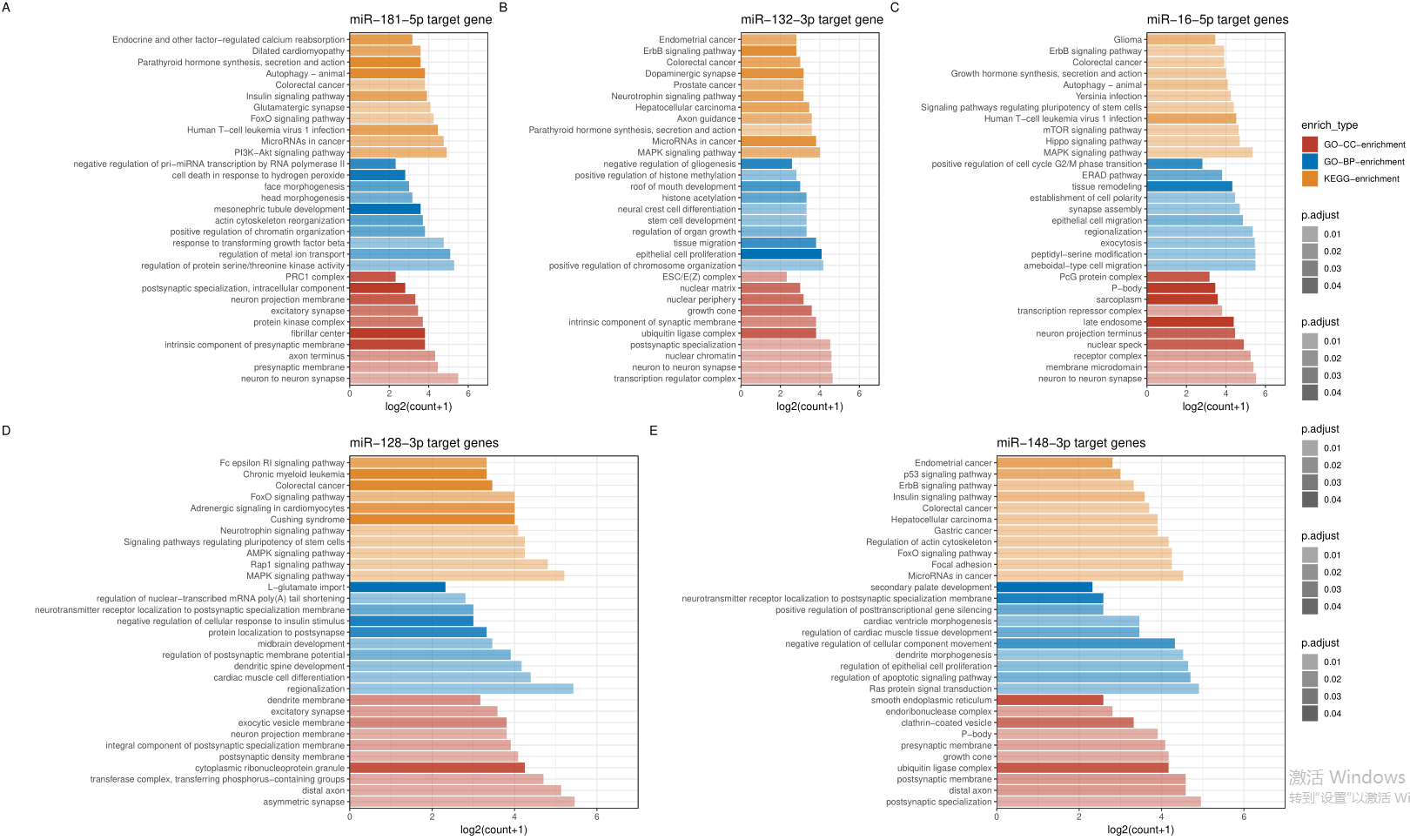
通过Annolnc2预测能与lncRNA互作的潜在miRNA及互作位点，筛选结合的保守分数大于60且互作位点位于lncRNA环上(互作位点位于换上且大于4个碱基)的miRNA，作为潜在ceRNA调控网络的miRNA对象，如图所示展示的是预测的部分可以与miRNA互作的lncRNA基因和互作的位点。本研究通过本地服务器搭建好Annolnc2对178个参与mRNA-lncRNA共表达的lncRNA基因进行批量分析，最终一共预测到了107个miRNA和lncRNA存在潜在的结合位点，为了将miRNA的范围缩小到于结直肠癌相关的miRNA



**图 3‑20 NONMMUT002330.2与miR-1-3p结合位点的预测**

**Fig. 3‑20 Prediction binding site between NONMMUT002330.2 and miR-1-3p by Annolnc2**

### miRNA靶基因功能预测及功能分析



**图 3‑21 miRNA靶基因的功能富集分析**

**Fig. 3‑21 Enrichment analysis of miRNA target genes**

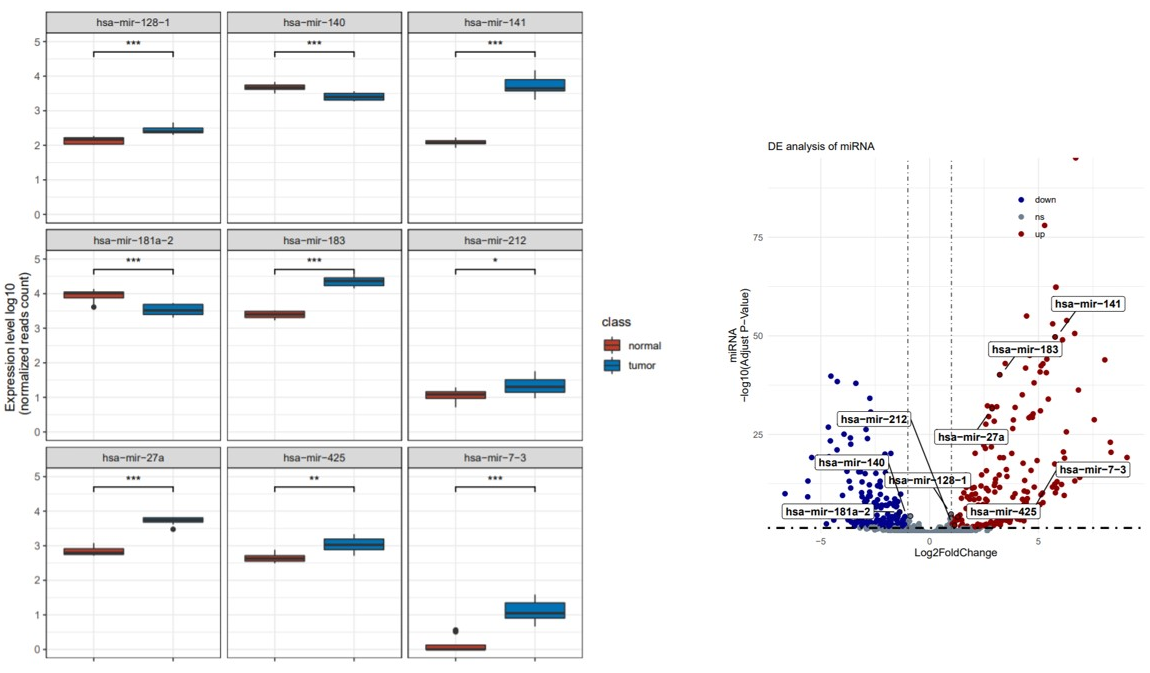
### miRNA生存分析及差异表达分析

为了探究预测到的和结直肠癌发展相关的miRNA是否在结肠癌患者中真实作用，本研究从TCGA数据库下载了8个结肠癌患者和8个正常人作为对照，探索正常组织和癌变组织差异表达的miRNA，并且从TCGA数据库中下载了487患者的miRNA表达数据及临床数据，使用survminer和survival整合表达数据和临床数据做生存分析，探索表达水平和结肠癌患者生存显著相关且差异表达的的miRNA，并通过miRbase将这些miRNA同源注释到小鼠，最终保留与动态表达lncRNA存在潜在至少3个结合位点的miRNA及动态表达lncRNA。最终确定了9个在结直肠癌患者和对照差异表达且与结直肠癌患者生存相关的miRNA。如图3-22展示同源注释的结果，3-23展示miRNA在结肠癌患者和对照之间表达水平及具体上调下调水平如果3-23右图火山图所示，如图3-24展示关键的miRNA的预后情况，可以看到不论是从差异表达还是预后，这些都是结肠癌中关键的miRNA，将这些miRNA作为ceRNA中的miRNA。



**图 3‑22 小鼠到人的miRNA同源注释**

**Fig. 3‑22 Homologous miRNA annotation from mouse to human**



**图 3‑23 TCGA中miRNA的差异表达**

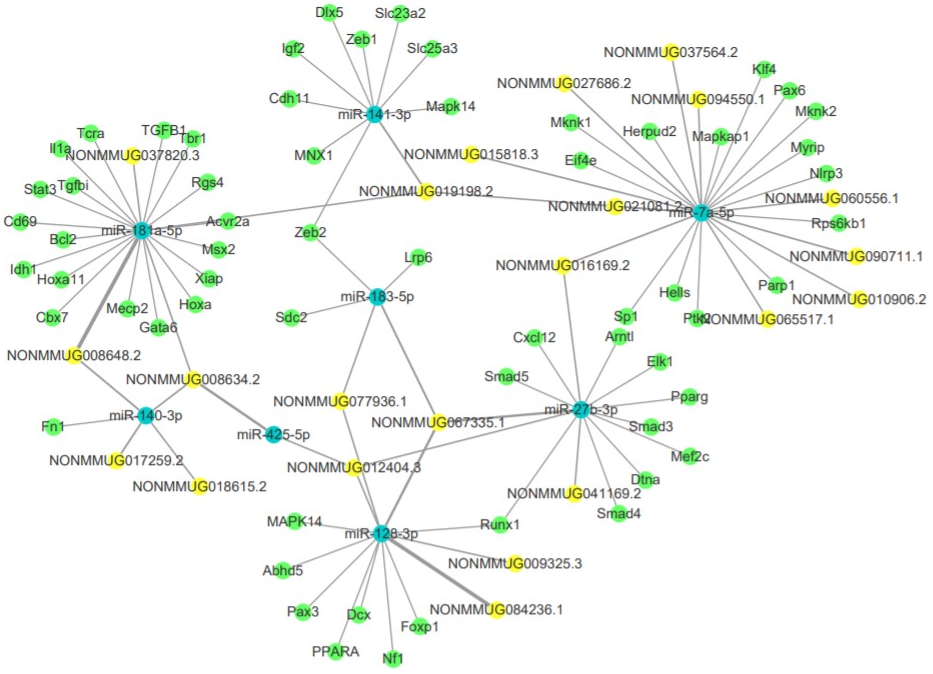
**Fig. 3‑23 Different expression analysis of miRNA in TCGA**



**图 3‑24 miRNA生存分析**

**Fig. 3‑24 Survival analysis of miRNA in human**

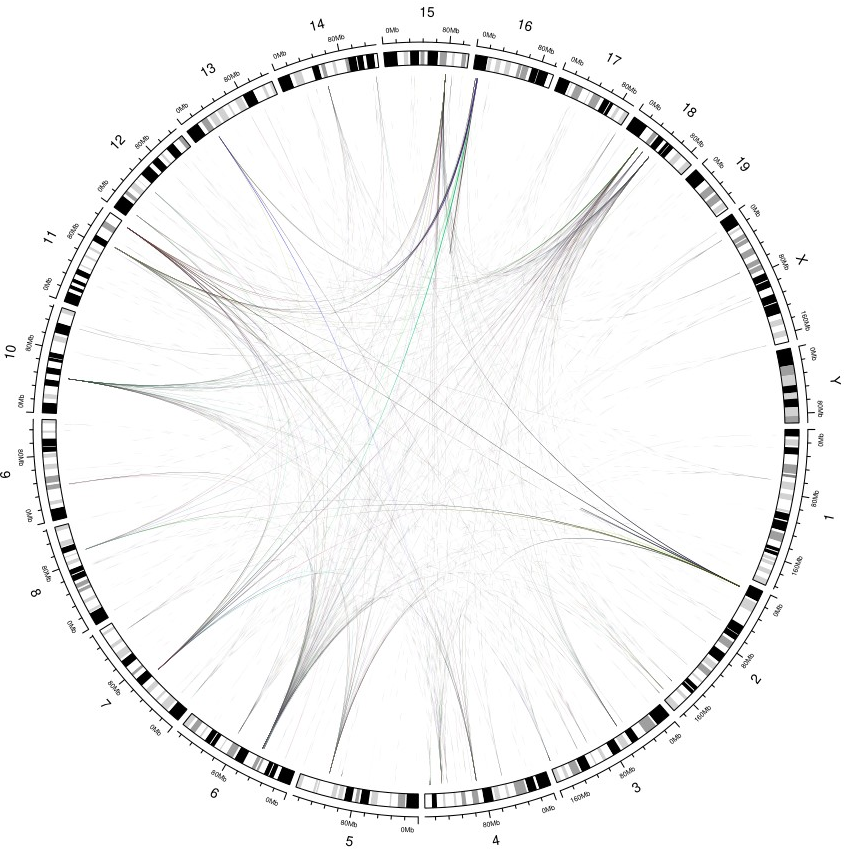
### CeRNA的构建



**图 3‑25 CeRNA网络的可视化**

**Fig. 3‑25 CeRNA network**

## 动态表达lncRNA与蛋白质互作分析



**图 3‑26动态表达lncRNA与蛋白质互作**

**Fig. 3‑26 Interactions between dy-lncRNA**

## 动态表达lncRNA与转录因子互作分析

1. ARAB K, PARK Y J, LINDROTH A M, et al. 2014. Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. Mol Cell [J], 55: 604-614.
2. 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.
3. 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.
4. ARNOLD M, SIERRA M S, LAVERSANNE M, et al. 2017. Global patterns and trends in colorectal cancer incidence and mortality. Gut [J], 66: 683-691.
5. BARON J A, BARRY E L, MOTT L A, et al. 2015. A Trial of Calcium and Vitamin D for the Prevention of Colorectal Adenomas. N Engl J Med [J], 373: 1519-1530.
6. 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.
7. CARETHERS J M, JUNG B H 2015. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. Gastroenterology [J], 149: 1177-1190 e1173.
8. CONSORTIUM E P 2012. An integrated encyclopedia of DNA elements in the human genome. Nature [J], 489: 57-74.
9. DERRIEN T, JOHNSON R, BUSSOTTI G, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res [J], 22: 1775-1789.
10. DJEBALI S, DAVIS C A, MERKEL A, et al. 2012. Landscape of transcription in human cells. Nature [J], 489: 101-108.
11. ESTELLER M, CORN P G, BAYLIN S B, et al. 2001. A gene hypermethylation profile of human cancer. Cancer Res [J], 61: 3225-3229.
12. FATIMA R, AKHADE V S, PAL D, et al. 2015. Long noncoding RNAs in development and cancer: potential biomarkers and therapeutic targets. Mol Cell Ther [J], 3: 5.
13. FEARON E R 2011. Molecular genetics of colorectal cancer. Annu Rev Pathol [J], 6: 479-507.
14. FIJTEN G H, STARMANS R, MURIS J W, et al. 1995. Predictive value of signs and symptoms for colorectal cancer in patients with rectal bleeding in general practice. Fam Pract [J], 12: 279-286.
15. FLEMING M, RAVULA S, TATISHCHEV S F, et al. 2012. Colorectal carcinoma: Pathologic aspects. J Gastrointest Oncol [J], 3: 153-173.
16. FRANCO-ZORRILLA J M, VALLI A, TODESCO M, et al. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet [J], 39: 1033-1037.
17. GRAFF R E, MOLLER S, PASSARELLI M N, et al. 2017. Familial Risk and Heritability of Colorectal Cancer in the Nordic Twin Study of Cancer. Clin Gastroenterol Hepatol [J], 15: 1256-1264.
18. GUINNEY J, DIENSTMANN R, WANG X, et al. 2015. The consensus molecular subtypes of colorectal cancer. Nat Med [J], 21: 1350-1356.
19. 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.
20. GUTTMAN M, RINN J L 2012. Modular regulatory principles of large non-coding RNAs. Nature [J], 482: 339-346.
21. HEWITSON P, GLASZIOU P, WATSON E, et al. 2008. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. Am J Gastroenterol [J], 103: 1541-1549.
22. 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.
23. IMPERIALE T F, RANSOHOFF D F, ITZKOWITZ S H, et al. 2014. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med [J], 370: 1287-1297.
24. IYER M K, NIKNAFS Y S, MALIK R, et al. 2015. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet [J], 47: 199-208.
25. JESS T, RUNGOE C, PEYRIN-BIROULET L 2012. Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies. Clin Gastroenterol Hepatol [J], 10: 639-645.
26. JIAO S, PETERS U, BERNDT S, et al. 2014. Estimating the heritability of colorectal cancer. Hum Mol Genet [J], 23: 3898-3905.
27. KEDRIN D, GALA M K 2015. Genetics of the serrated pathway to colorectal cancer. Clin Transl Gastroenterol [J], 6: e84.
28. KLATTENHOFF C A, SCHEUERMANN J C, SURFACE L E, et al. 2013. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell [J], 152: 570-583.
29. 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.
30. LI F Y, LAI M D 2009. Colorectal cancer, one entity or three. J Zhejiang Univ Sci B [J], 10: 219-229.
31. LIU L, WANG H J, MENG T, et al. 2019. lncRNA GAS5 Inhibits Cell Migration and Invasion and Promotes Autophagy by Targeting miR-222-3p via the GAS5/PTEN-Signaling Pathway in CRC. Mol Ther Nucleic Acids [J], 17: 644-656.
32. 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.
33. MERCER T R, DINGER M E, SUNKIN S M, et al. 2008. Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci U S A [J], 105: 716-721.
34. MIGLIORE L, MIGHELI F, SPISNI R, et al. 2011. Genetics, cytogenetics, and epigenetics of colorectal cancer. J Biomed Biotechnol [J], 2011: 792362.
35. MORRIS K V, MATTICK J S 2014. The rise of regulatory RNA. Nat Rev Genet [J], 15: 423-437.
36. MURPHY N, STRICKLER H D, STANCZYK F Z, et al. 2015. A Prospective Evaluation of Endogenous Sex Hormone Levels and Colorectal Cancer Risk in Postmenopausal Women. J Natl Cancer Inst [J], 107.
37. NG S Y, BOGU G K, SOH B S, et al. 2013. The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. Mol Cell [J], 51: 349-359.
38. PANG K C, FRITH M C, MATTICK J S 2006. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. Trends in Genetics [J], 22: 1-5.
39. PASMANT E, LAURENDEAU I, HERON D, et al. 2007. Characterization of a germ-line deletion, including the entire INK4/ARF locus, in a melanoma-neural system tumor family: identification of ANRIL, an antisense noncoding RNA whose expression coclusters with ARF. Cancer Res [J], 67: 3963-3969.
40. 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.
41. PINO M S, CHUNG D C 2010. The chromosomal instability pathway in colon cancer. Gastroenterology [J], 138: 2059-2072.
42. PITOT H C 1993. The molecular biology of carcinogenesis. Cancer [J], 72: 962-970.
43. PLATH K, FANG J, MLYNARCZYK-EVANS S K, et al. 2003. Role of histone H3 lysine 27 methylation in X inactivation. Science [J], 300: 131-135.
44. PRENSNER J R, IYER M K, SAHU A, et al. 2013. The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nat Genet [J], 45: 1392-1398.
45. QI P, DU X 2013. The long non-coding RNAs, a new cancer diagnostic and therapeutic gold mine. Mod Pathol [J], 26: 155-165.
46. 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.
47. SAHU A, SINGHAL U, CHINNAIYAN A M 2015. Long noncoding RNAs in cancer: from function to translation. Trends Cancer [J], 1: 93-109.
48. SALMENA L, POLISENO L, TAY Y, et al. 2011. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell [J], 146: 353-358.
49. SHIN A, KIM K Z, JUNG K W, et al. 2012. Increasing trend of colorectal cancer incidence in Korea, 1999-2009. Cancer Res Treat [J], 44: 219-226.
50. SINGH M P, RAI S, PANDEY A, et al. 2019. Molecular subtypes of colorectal cancer: An emerging therapeutic opportunity for personalized medicine. Genes & Diseases [J].
51. 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.
52. TOYOTA M, AHUJA N, OHE-TOYOTA M, et al. 1999. CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci U S A [J], 96: 8681-8686.
53. TRIEVEL R C, SHILATIFARD A 2009. WDR5, a complexed protein. Nat Struct Mol Biol [J], 16: 678-680.
54. VENTURA L, MANTELLINI P, GRAZZINI G, et al. 2014. The impact of immunochemical faecal occult blood testing on colorectal cancer incidence. Dig Liver Dis [J], 46: 82-86.
55. VISONE R, BACALINI M G, DI FRANCO S, et al. 2019. DNA methylation of shelf, shore and open sea CpG positions distinguish high microsatellite instability from low or stable microsatellite status colon cancer stem cells. Epigenomics [J], 11: 587-604.
56. VITIELLO M, TUCCOLI A, POLISENO L 2015. Long non-coding RNAs in cancer: implications for personalized therapy. Cellular oncology [J], 38: 17-28.
57. WALSH J M, TERDIMAN J P 2003. Colorectal cancer screening: scientific review. JAMA [J], 289: 1288-1296.
58. 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.
59. WANG P, XUE Y, HAN Y, et al. 2014. The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. Science [J], 344: 310-313.
60. YANG L, XIONG Z, HE W, et al. 2018. Proximal shift of colorectal cancer with increasing age in different ethnicities. Cancer Manag Res [J], 10: 2663-2673.
61. 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.
62. YUE F, CHENG Y, BRESCHI A, et al. 2014. A comparative encyclopedia of DNA elements in the mouse genome. Nature [J], 515: 355-364.