

研究报告  
Research Report

## 大肠癌转移相关基因表达调控的生物信息学分析

齐鲁<sup>1</sup> 丁彦青<sup>1,2\*</sup>

1 南方医科大学基础医学院病理学系, 广州, 510515; 2 南方医科大学南方医院病理科, 广州, 510515

\* 通讯作者, dyq@fimmu.com

**摘要** 大肠癌是常见的消化道恶性肿瘤,在中国呈逐年上升的趋势。对大肠癌发生发展转移的研究能够指导临床治疗,对研发新药也有着重要意义。本文通过对表达谱数据进行分析,通过表达谱差异数据进行功能富集,研究了大肠癌转移前的早期原发肿瘤的转录调控特点以及远隔器官转移后的大肠癌的转录调控特点,筛选出了部分能够受到多重调控并在转移后肿瘤组织中高表达的关键基因,通过对这些基因相互作用关系研究,构建出转移后关键基因相互作用调控网络,为大肠癌的治疗提供更多潜在靶点。

**关键词** 大肠癌, 调控网络, miRNA, 转录因子

## Bioinformatics Analysis of Gene Expression Regulation Associated with Colorectal Cancer Metastasis

Qi Lu<sup>1</sup> Ding Yanqing<sup>1,2\*</sup>

1 Department of Pathology Southern Medical University, Guangzhou, 510515; 2 Department of Pathology Southern hospital, Guangzhou, 510515

\* Corresponding author, dyq@fimmu.com

DOI: 10.3969/gab.032.000083

**Abstract** Colorectal cancer is a common gastrointestinal malignancy cancer that shows an increasing trend in China these years. Study of colorectal cancer occurrence and metastasis can guide clinical treatment, which is also of great significance in research and development of new drugs. In this article, collecting gene expression data, as well as functional enrichment of differential data of gene expression profile were analyzed to figure out the transcriptional regulation characteristics of the early primary tumors before the colorectal cancer metastasis, a part of key genes have been screened that can be multiple regulated and of high expression in the transferred tumor tissue. From researches of interactions between these genes, we can build an interactional regulation network among key genes and find more potential targets for therapy of colorectal cancer.

**Keywords** Colorectal cancer, Regulatory network, miRNA, Transcription factor

蛋白质是基因功能的执行者,基因的转录与翻译对蛋白质的形成有着至关重要的作用,而转录与翻译又受到多种因素的调控和影响。其中转录因子与 DNA 相结合能够调控基因的表达情况,而 miRNA 能够通过结合 RNA 的 3' UTR 区域降解或者抑制 RNA 阻止 RNA 翻译为蛋白质。因此转录水平的调控和翻译水平的调控均能够影响蛋白质最终的形成及产量的多少。

大肠癌是大肠粘膜上皮来源的恶性肿瘤,是消化道恶性肿瘤中的常见类型。因其恶性程度高,易经淋巴道、血道转移,肿瘤晚期常发生远隔器官转移,危及患者生命。转移是大肠癌患者的主要死因。因此研究

与大肠癌转移相关重要基因的调控机制,能够为我们提供更多新的药物治疗靶点,以期更好的研究大肠癌的治疗方法。

现有的相关研究主要集中在与大肠癌转移相关的单个基因或 miRNA 的研究上面,或只从差异基因的上下调程度筛选大肠癌转移相关基因,因此忽略了大肠癌转移中各个差异基因的受调控情况。基因受到的转录因子及 miRNA 作用越多,其所受调控越精细,越能说明此基因的重要性。而在细胞信号的传递过程中,蛋白需与其它蛋白相互作用才能发挥其效能,因此多个重要蛋白构成的相互作用网络,成为细胞信号转导的关键部分。

本文主要通过对大肠癌转移相关差异表达基因的调控情况进行分析和筛选,进而从差异基因中筛选出受调控最多的基因并构建由这些差异基因所构成的调控网络。这些基因在大肠癌转移相关的细胞信号传递中起着关键作用,因此能够成为抑制或者治疗大肠癌转移的重要治疗靶点。

## 1 结果与分析

### 1.1 miRNA 对大肠癌转移相关差异表达基因的调控

大肠癌 T1、T2 期的基因表达情况与 M1 期有明显的差异,为了更好地挖掘出差异表达基因所包含的信息,通过 GSEA (Damian and Gorfine, 2004) 工具对差异表达基因进行不同功能集的基因富集。用 GSEA 通过 miRNA 基因集对此表达谱数据中具有相同 miRNA 靶点的高表达基因进行富集,基因集版本为 V3.0。数据库中的 221 个 miRNA 基因集中有 202 个可用,并且有 192 个基因集在 M1 表型的表达谱数据中表达上调,只有 10 个基因集在 T1T2 表型的表达谱数据中上调,且假阳性率较高,因此,本文只研究在 M1 表型中上调的基因集。M1 表型的基因集结果通过标准化富集分数进行排序,并且权衡了 P 值及错误发现率(FDR) 筛选得到几个在转移后 M1 期的大肠癌高表达相关基因中具有的共同 miRNA 作用靶序列。这些靶序列分别为 MIR-369-3P、MIR-105、MIR-484、MIR-518B、MIR-518C、MIR-518D、MIR-18A、MIR-18B、MIR-200B、MIR-200C、MIR-429、MIR-520G、MIR-520H、MIR-126、MIR-143、MIR-29A、MIR-29B、MIR-29C、MIR-19A、MIR-19B 能够结合的 3' UTR 的靶序列。每种 miRNA 的靶序列都对应着许多与 M1 期转移相关高表达基因,每种高表达基因的 3' UTR 区域都含有相应上述 miRNA 作用靶序列。虽然每个 miRNA 作用位点集内包含着许多基因,但只有部分在 M1 中的高表达的基因对此功能集起富集作用(图 1)。由于同样的调控序列可能对应着不同的 miRNA,因此上述筛选出的 miRNA 作用位点有 11 类。将上述 11 类功能集当中起富集作用的基因挑选出来,计算出现频率,研究共调控机制。结果 *KLF12* 基因的 3' UTR 区域存在上述 11 类中的 6 类 miRNA 作用位点,说明这六类 miRNA 能够同时调控高表达的 *KLF12* 基因。同理,具有五类 miRNA 作用位点的基因为 *PCDHA9*、*PCDHA3*、*PCDHA10*、*PCDHA5*、*QKI*、*PRICKLE2*、*ATXN1*。具有四类的为 *RFX4*、*OGT*、*CREB5*、*IGF1*、*TRIM33*。对上述 13 种蛋白运用 Visant (Hu

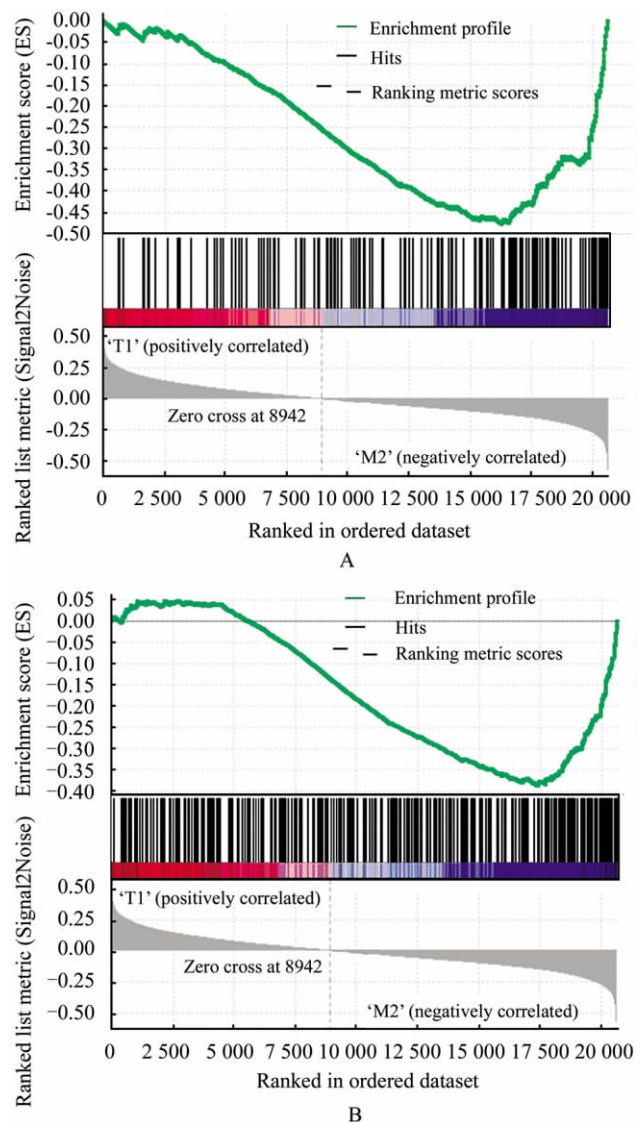


图 1 GSEA 富集图

注: A: 富集图: GTATTAT, MIR-369-3P; B: 富集图: GTAT-TAT, MIR-200B, MIR-200C, MIR-429

Figure 1 Enrichment plot of GSEA

Note: A: Enrichment plot: GTATTAT, MIR-369-3P; B: Enrichment plot: GTATTAT, MIR-200B, MIR-200C, MIR-429

et al., 2008) 工具构建蛋白质相互作用网络图(图 2) 可以看出 AXIN1 可以将其中的 7 个蛋白连接起来 AXIN1 为 Wnt 信号通路中的轴蛋白(Biechele et al., 2012) 因此 AXIN1 在大肠癌转移相关基因的调控上起着重要作用。

为了更好的分析筛选出的 13 个关键蛋白在大肠癌转移后分子网络中的关系,将上述 13 种蛋白的相互作用蛋白每个节点进一步展开,得到一个更复杂的蛋白相互作用网络图,然后计算能把上述蛋白都关联起来的中间节点的蛋白数量共 364 个。上述 13 种蛋白可能在大肠癌转移起着重要作用,但每种蛋白并不

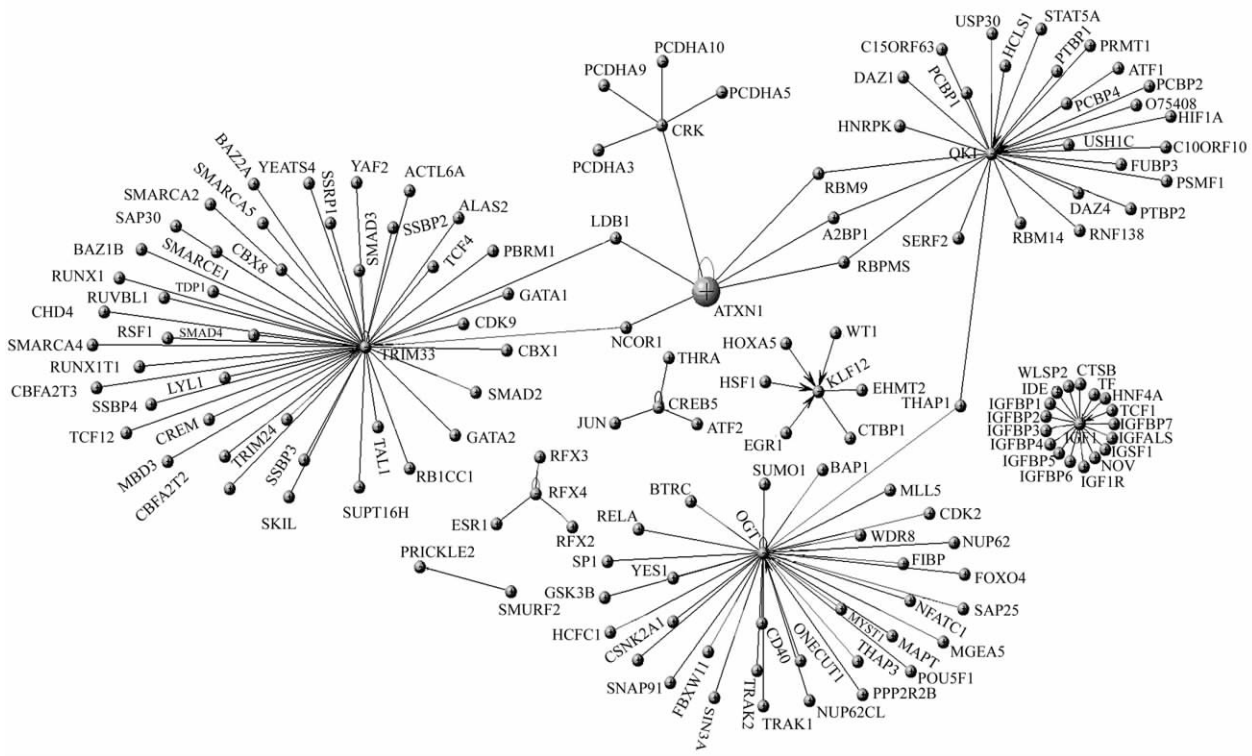


图 2 13 个蛋白与蛋白间相互作用网络  
Figure 2 13 protein-protein interaction network

是单独能够完成相应的功能,而是通过与它相互作用的蛋白构成多级网络,将生物信号进行级联传递的。所以 364 个蛋白作为桥梁,把这 13 种蛋白联系起来,构成一个复杂的网络,而这网络,可能为大肠癌转移发生发展相关的信号传递网络的一部分。因此对这些蛋白进行 Gene Set Analysis Toolkit V2 工具中的 Wikipathways(Kelder et al., 2012)通路分析(表 1)可以发现,雄激素受体信号通路、TGF- $\beta$  信号通路、EGFR1 信号通路等均有许多中间节点蛋白参与,因此,雄激素受体信号等通路与大肠癌的转移可能有着密切联系,因此也可以认为激素水平的变化对大肠癌的转移可能有着重要影响。对这 364 个蛋白共同转录因子通过 Gene Set Analysis Toolkit V2 工具进行富集分析,结果 364 个基因当中有 102 个基因上游调控区域有 SP1 转录因子结合位点,有 86 个基因有 MAZ 结合位点,有 61 个基因有 PAX4 结合位点,有 69 个基因有 LEF1 结合位点,有 72 个基因有 E12 结合位点,有 65 个基因有 FOXO4 结合位点,有 57 个基因有 AP4 结合位点,有 57 个基因有 NFAT 结合位点,有 39 个基因有 MYOD 结合位点,这些转录因子结合位点调控着这 364 个基因构成的网络,可能与大肠癌转移密切相关。

### 1.2 转录因子对大肠癌转移相关差异表达基因的调控

用 GSEA 工具通过转录因子结合位点基因集对前面已归为 T1T2、M1 两类的大肠癌表达谱数据进行共同转录因子结合位点富集。在 615 个转录因子基因集中,有 576 个可用,其中有 486 个基因集在 M1 表型中表达上调,只有 90 个基因集在 T1T2 表型中表达上调,可见转移后的肿瘤高表达基因需要较多的转录因子进行精密调控。在 M1 表型表达上调的 486 个转录因子结合位点基因集中,通过标准化富集分数,权衡 P 值及错误发现率,经筛选得到前几个转录因子结合位点分别是 EVI1、OCT1、FOXO4、SMAD3、AML、PAX2、TAL1、ALPHA E47、FREAC2、PAX4、AM-EF2、RREB1、CEBPA、TEF、SRY、GATA1、FXR 等的转录因子的结合位点,说明在转移后的大肠癌组织中,表达升高的基因均含有上述转录因子结合位点,这些对应的转录因子对大肠癌转移起着重要作用。而部分基因的转录调控区域同时含有这些转录因子结合位点,构成了组合调控,调控机制越复杂,转录因子越多,说明这些基因表达调控越精细,这部分基因在转移后的大肠癌组织中发挥的功能越重要。在此 16 个转录因子结合位点基因集中,提取对相应基因集起着富集作用的基因,计算得出同时含有 4 个以上此转录因子结合



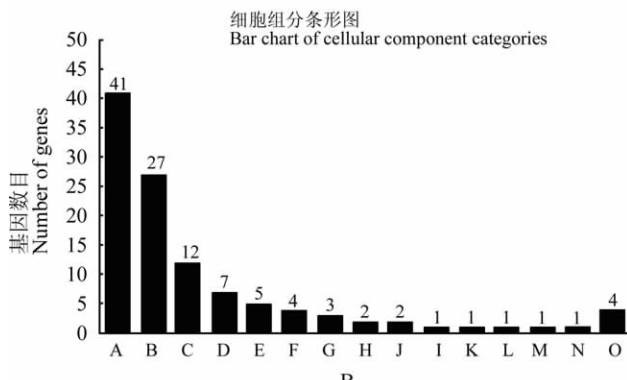
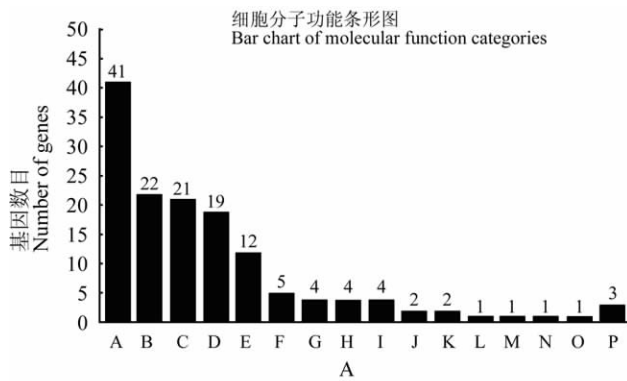
表 1 维基通路富集分析  
Table 1 Wikipedia pathway enrichment analysis

信号通路	数目	Entrez 基因身份
Signaling pathways	No.	Entrez gene IDs
Androgen receptor signaling pathway	39	2033 9612 3276 7157 25942 7329 6714 3303 1387 3725 5925 4193 2516 5901 8204 1386 5970 10499 6667 7050 2597 207 8409 2274 672 2316 1956 3304 10273 4088 8202 1025 3065 367 8648 861 2099 4089 23132
TGF- $\beta$ receptor signaling pathway	38	5599 2033 2308 4092 7157 7046 1026 7048 7341 3725 1387 5925 4609 1386 4091 1021 11171 6667 7050 4303 3312 5295 64750 4088 7421 6498 7514 4087 3205 1874 367 3065 8648 1017 2099 4089 3172 8945
EGFR1 signaling pathway	35	5599 2308 6777 25942 85021 6714 3725 7409 466 6464 4609 8503 6667 7050 4893 5359 207 6654 4690 9219 5295 9564 1956 5595 1051 4088 2885 4087 1398 1445 6772 2185 3065 8462 6776
IL-6 signaling pathway	29	5599 2033 2308 2064 7409 3725 1387 5925 6464 5521 3320 207 4303 6654 5295 958 5595 9261 1051 2885 2534 1025 6772 367 3065 2932 2185 8648 4137
B cell receptor signaling pathway	29	5599 2308 7409 3725 5925 6464 5970 1386 1021 207 4690 6654 5295 4773 4772 9564 10014 5595 9261 2885 2534 1398 11184 3059 1445 6772 2932 2185 1017
MAPK signaling pathway	29	5599 7186 7157 7046 7048 5566 3303 3725 4609 1386 4893 207 3164 3312 4914 2316 1956 5595 3304 9261 2885 1398 11184 1647 6722 356 408 5603 4137
$\delta$ -notch signaling pathway	22	2033 9612 7157 25942 3725 4856 5970 8819 207 5295 1956 5595 4088 3065 2932 9611 3516 4086 7528 1017 4089 3066
IL-3 signaling pathway	23	5599 2308 6777 5566 6714 7409 466 6464 2624 1386 2623 207 6654 5295 5595 9261 2885 2534 1398 6772 2932 7534 6776
Estrogen signalling	21	2033 6908 25942 5566 6714 3725 1387 2959 6667 207 6654 8841 672 10014 2885 9759 3065 9611 6880 2099 3066
$\alpha$ 6- $\beta$ 4 integrin signaling pathway	20	2064 1026 7531 6714 25 6464 8503 207 5295 7525 1956 4088 8660 3667 2885 2534 4087 367 7534 3691

位点的基因共有 41 个，其中有 HOXC4、PURA 基因同时含有 9 个此转录因子结合位点。因此，这些基因在转移后相关转录因子中，受到多个相关转录因子组合调控，能够对大肠癌转移的发生发展起到关键作用。对这 41 个有着大于 4 个相关转录因子结合位点的基因进行 GO 分析，得出其中有 22 个基因能够与核酸结合，其中有 19 个具有转录活性，为转录因子（图 3），说明这些转录因子同时也调控着其他基因。对这 41 个基因对应的蛋白构建相互作用网络，将每个与这 41 个蛋白相连的蛋白进一步展开，然后计算与这 41 个蛋白相连的中间节点蛋白，共有 1 100 个，也就是说有 1 100 个蛋白把这 41 个蛋白连接起来构成复杂的相互作用网络，对这 1 100 个蛋白进行 Gene Set Analysis Toolkit V2 工具中的 Wikipathways 通路分析（表 2）。可以发现，TGF- $\beta$  通路、雄激素受体信号通路等与这个网络相关，且与之前 miRNA 调控基因所构成的网络一致，因此进一步说明了大肠癌的转移与这几个信号通路相关。雄激素为固醇类激素，可直接通过细胞膜与细胞核内转录因子结合，调控转录。因此，雄激素调控转录因子，转录因子又通过相互作用网络将信号进一步传递，调控基因的表达。对这 1 100 个蛋白通过 Gene Set Analysis Toolkit V2 进行

共同转录因子进行分析，结果有 277 个基因有 SP1 结合位点，有 240 个基因有 MAZ 结合位点，有 185 个基因有 LEF1 结合位点，有 175 个基因有 FOXO4 结合位点，有 167 个基因有 NFAT 结合位点，有 186 个基因有 E12 结合位点，有 138 个基因有 PAX4 结合位点，有 119 个基因有 NFY 结合位点，有 101 个基因有 FRAC2 结合位点。可以发现，大部分转录因子与前面 miRNA 所调控基因构成网络共同转录因子重叠。说明这些转录因子在大肠癌转移相关信号调控网络中起着重要作用。

综合前面转移后大肠癌中 miRNA 所调控的基因以及转录因子所调控的基因，可以发现基因 KLF12 以及 CREB5 不仅在转录调控区域有很多相关转录因子结合位点，而且在转录后的 3' UTR 区域有着很多相关 miRNA 结合位点，因此认为这两个基因能够同时被这些转录因子以及 miRNA 所调控（图 4）。因为 KLF12 与 CREB5 本身为转录因子，因此也可以调控其他基因转录，形成级联反应，构建 KLF12 与 CREB5 相互作用调控网络图并将与他们相互作用蛋白进行展开，计算能够将 KLF12 与 CREB5 关联起来的若干个蛋白，并构成新的网络如（图 5）。可以看出在 KLF12 与 CREB5 之间有 28 个与这两个转录因子直接或间接相互作用的



注: 图 A: A: 所有; B: 核酸束缚; C: 蛋白质束缚; D: 转录调控活动; E: 离子束缚; F: 分子交换活动; G: 核苷酸束缚; H: 核染色质束缚; I: 转移酶活动; J: 酶调解活动; K: 结构分子活动; L: 电子载体活动; M: 碳水化合物束缚; N: 水解酶活动; O: 运输活动; P: 未分类; 图 B: A: 所有; B: 细胞核; C: 膜; D: 大分子络合物; E: 膜包围的内腔; F: 染色体; G: 细胞骨架; H: 细胞蛋白; I: 囊泡; J: 细胞外基质; K: 细胞外空隙; L: 细胞液; M: 线粒体; N: 包膜; O: 未分类

Figure 3 Bar chart about molecular function and cellular component  
 Note: Figure A: A: All; B: Nucleic acid binding; C: Protein binding; D: Transcription regulator activity; E: Ion binding; F: Transferase activity; G: Nucleotide binding; H: Chromatin binding; I: Molecular transducer activity; J: Enzyme regulator activity; K: Structural molecule activity; L: Electron carrier activity; M: Carbohydrate binding; N: Hydrolase activity; O: Transporter activity; P: Unclassified; Figure B: A: All; B: Nucleus; C: Membrane; D: Macromolecular complex; E: Membrane-enclosed lumen; F: Chromosome; G: Cytoskeleton; H: Cell projection; I: Vesicle; J: Extracellular matrix; K: Extracellular space; L: Cytosol; M: Mitochondrion; N: Envelope; O: Unclassified

蛋白, 这些蛋白将其联系在一起。在此对这 28 个蛋白进行 GO 分析, 28 个蛋白里面具有转录调控活性的蛋白有 26 个, 因此可以说明, 这两个蛋白是通过这 26 个转录因子连接起来的, 而转录因子相互结合成复合体可调控其他基因表达, 因此这些基因与大肠癌转移密切相关。

在前面的 GSEA 分析中, 有 90 个基因集在 T1T2

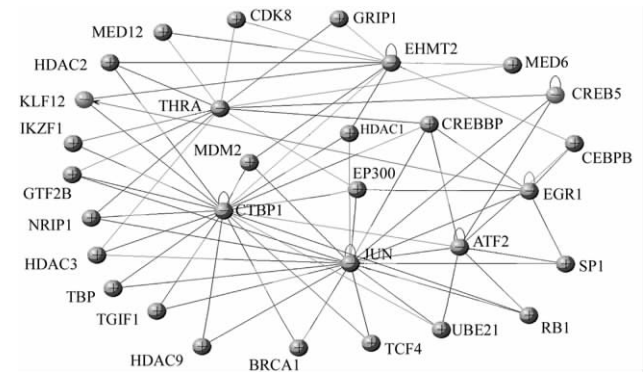


图 5 KLF12 和 CREB5 蛋白与蛋白相互作用网络

Figure 5 Interaction network constructed with the KLF12 and CREB5

表型中表达上调。说明部分基因在转移前的原发肿瘤早期起着重要作用, 通过标准化富集分数, 权衡 P 值及错误发现率, 筛选得到最相关的几个转录因子结合位点基因集为 E2F\_Q6、E2F1\_Q3、E2F1\_Q6\_01、E2F\_Q4、E2F1\_Q6、E2F\_Q3、E2F1DP2\_01、E2F\_Q2、E2F1DP1\_01、E2F1DP2\_01、E2F4DP2\_01、E2F4DP1\_01。可以发现, 全部转录因子结合位点均为 E2F 家族, 或者 E2F 与 DP 家族的二聚体, 其中主要为 E2F1。可见在转移前的原发肿瘤早期, E2F 家族转录因子对肿瘤的发生发展起着极为关键的作用, 提取对相应基因集起着富集作用的基因, 经计算有 18 个基因均含有上述 12 种转录因子结合位点, 有 62 个基因同时含有 8 个以上的上述转录因子结合位点, 说明这些基因在原发肿瘤早期起着关键作用。因此将这 62 个基因进行 Gene Set Analysis Toolkit V2 工具中的 Wikipathways 通路分析(表 3) 结果可以看出, 在转移前原发肿瘤早期, E2F 家族主要转录调控的高表达基因主要参与了细胞代谢例如 DNA 复制、核苷酸合成、染色体重组等生物进程, 因此, 说明在原发肿瘤早期主要以细胞增殖、代谢增强为主要表现。

## 2 讨论

大肠癌的发生发展转移是由多因素多步骤共同作用的结果, 发生远隔器官转移的大肠癌与未发生转移的原位大肠癌基因表达有着明显差异, 对这些差异表达基因的研究, 能够使我们更好的了解大肠癌转移的分子机制。本研究不仅限于对大肠癌 T1T2 期与 M1 期差异度高的基因进行分析, 而是通过对差异度高的基因的调控区域的 miRNA 作用位点和转录因子结合位点进行功能富集, 筛选出差异基因共有的 miRNA 作用位点和转录因子结合位点, 并将含有这

表 2 维基通路富集分析  
Table 2 Wikipedia pathway enrichment analysis

信号通路 Signaling pathways	数目 No.	Entrez 基因身份 Entrez gene IDs
TGF- $\beta$ receptor signaling pathway	89	2033 7157 8454 5520 10755 1026 1432 9372 6497 3727 2022 7161 1387 595 8554 7049 1499 5296 163 6935 7321 1386 1021 11171 9978 1019 3312 55914 857 983 4088 7421 6498 4087 51176 5933 5579 1616 3205 1874 8312 10987 3065 8648 409 3172 4089 9839 7323 51433 860 4092 10664 7046 4205 7042 5934 2113 7048 7341 3725 22938 5925 8668 4609 6678 3726 5580 6885 4091 3837 6667 7040 993 7050 5295 64750 2353 9146 10131 94 7027 9063 4802 7322 367 1017 9133 2099
Androgen receptor signaling pathway	71	2033 836 5894 7157 25898 25942 7329 3303 1387 595 8554 4193 1499 5901 841 1386 5970 29843 2597 207 902 11143 5451 672 2316 1022 857 983 10401 4088 3065 8648 5594 4089 573 7041 51588 9612 3146 860 3276 6714 2962 3725 5925 7337 5728 6047 2965 8204 834 6774 10499 6667 7050 2908 23598 2274 2100 10048 1956 10273 8202 840 1025 9063 367 3148 9604 2099 10114
TNF- $\alpha$ /NF- $\kappa$ B signaling pathway	66	7189 836 943 1457 8454 6597 1460 1147 10013 7329 7531 8290 1387 3551 4790 835 5971 2969 841 5970 6605 4215 207 3326 10971 57805 2316 857 7532 10401 1654 5515 8517 6599 10293 6772 3065 2932 3838 6198 2289 3066 7323 7186 8518 4176 6714 7341 6205 4174 10010 5781 3320 2091 1459 6128 5747 10399 6601 7133 840 4792 5371 7322 7534 6124
Cell cycle	47	2033 7157 1026 10013 891 5111 4193 890 999 1027 1021 5347 902 1019 991 1871 983 10014 7532 4088 9759 5933 1874 8900 8243 3065 2932 11200 4089 4172 3066 1111 4176 25 5925 4174 896 894 7040 993 5591 8841 6502 7027 1869 9133 1017
EGFR1 signaling pathway	60	5894 6195 5879 1050 5578 25942 1432 8290 3727 6196 6464 998 5296 8503 4215 207 4690 83737 842 857 867 6622 23624 4088 5335 4087 1398 5579 8440 6772 3065 2185 5594 3856 7410 1915 6714 1999 3725 4609 5781 1385 6774 6667 5058 7050 5931 6654 9219 2060 5295 5339 2002 2353 1956 1051 2885 160 6633 6197
IL-6 signaling pathway	47	2033 836 660 5894 5879 5516 2064 5520 4067 1432 5519 1020 1387 6196 6464 5518 5296 4790 207 842 958 5515 5335 2534 1616 6772 3065 2932 2185 6198 8648 5594 3725 5925 5580 5781 3320 6885 6774 6654 5295 5747 2353 1051 2885 1025 367
B cell receptor signaling pathway	54	5894 6195 1147 4067 1432 890 1499 6464 5911 5296 3551 6194 2969 5970 1386 1021 207 2931 4690 1019 842 1022 867 10014 5335 8517 2534 1398 5579 6772 2932 2185 6198 5594 7410 3725 5925 896 5580 5781 894 1385 6885 6774 5170 387 6654 5295 5747 2002 2885 840 4792 1017
MAPK signaling pathway	54	7189 836 4149 4763 5894 5879 7157 1432 3303 3727 998 3551 4790 835 841 1386 207 2249 842 3309 3312 2316 8517 1398 5579 1616 409 5594 839 7186 23162 7046 7042 7048 3725 4609 5601 5062 5580 6885 834 5058 7040 3164 2002 2353 1956 355 2885 840 356 408 4217 6197
IL-3 signaling pathway	43	660 5894 8835 6256 6513 5879 5578 4067 1432 6464 998 5296 4790 1386 6688 2623 5914 207 2931 10971 867 5515 2534 1398 5579 6772 2932 5594 4318 1111 6714 5601 5781 1385 6774 5058 6654 5295 10399 5747 7133 2885 7534
Estrogen signalling	39	2033 2071 1147 10013 25942 1432 25759 595 1387 3551 4790 902 207 672 1022 10014 8517 9759 3065 5594 3066 6908 6714 3725 5601 2965 2068 2959 1385 6667

些相关的作用位点较多的差异基因进行筛选,因此筛选出的基因受到众多相关的 miRNA 和转录因子的调控,功能越重要的基因受到的调控越精细,因此这些基因对大肠癌的转移有着重要作用。对这些基因构建相互作用网络,再将网络中的节点蛋白共有的转录因子结合位点进行富集,可以筛选出对此蛋白网络起着重要作用的转录因子。因此,本研究筛选出 2 个含有这些作用位点最多的差异表达基因 *KLF12* 和 *CREB5*。这两个基因本身为转录因子,连接他们的相互作用网络中大部分也为转录因子,因此这些转录因子群

相互连接,组合调控其他基因的表达,显示出其对大肠癌转移相关基因表达调控的影响力。*KLF12* 基因已被证实在胃癌中表达升高,并且高表达 *KLF12* 能够增强胃癌的侵袭和转移,证明其在低分化胃癌中能够促进肿瘤的发生发展(Nakamura et al., 2009)。*CREB5* 为 CRE 依赖的转录因子,它是唯一一个能够连接 cAMP 信号通路和 TPA 信号通路的关键蛋白(Zu et al., 1993),其与大肠癌发生发展转移的关系有待于进一步研究。本文构建了蛋白质相互作用网络,并且发现调控网络中蛋白质所对应基因的上游调控区域转录



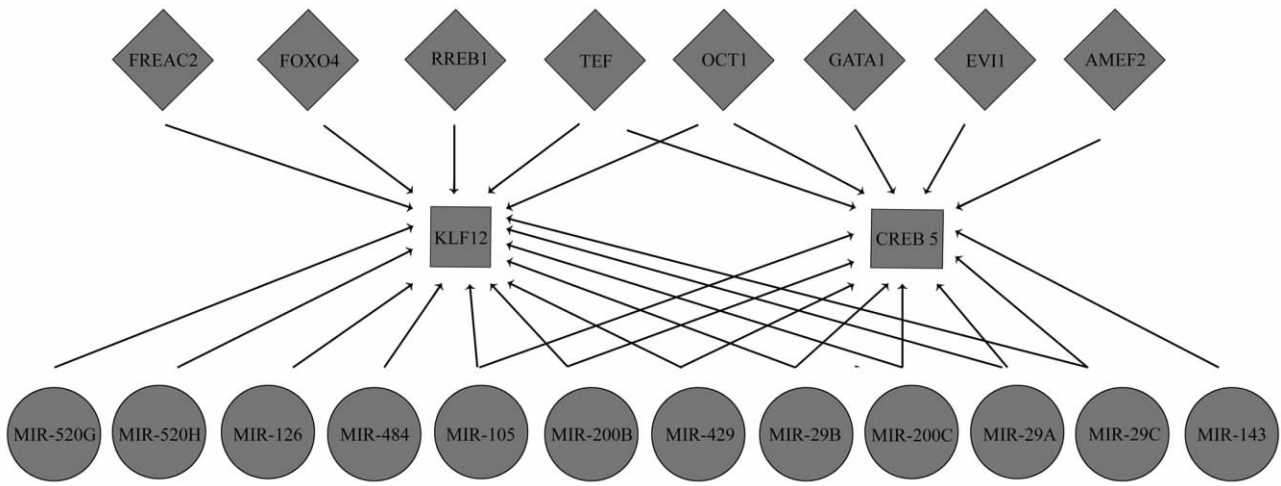


图 4 KLF12 和 CREB5 的 miRNA 和转录因子相互作用网络

Figure 4 Interaction network constructed with the miRNA and transcription factor of KLF12 and CREB5

表 3 维基通路富集分析

Table 3 Wikipedia pathway enrichment analysis

信号通路 Signaling pathways	数目 No.	Entrez 基因身份 Entrez gene IDs
DNA replication	10	51053 4173 4175 5427 990 4171 4176 5424 10714 4172
G1 to S cell cycle control	9	4173 993 4175 5427 5933 4609 4171 4176 4172
Cell cycle	9	4173 993 4175 5933 990 9088 4171 4176 4172
Nucleotide metabolism	3	6241 5424 5631
mRNA processing	4	6632 84844 6432 6426
DNA damage response	3	993 4609 2177
Homologous recombination	2	5424 10714
p38 MAPK signaling pathway (BioCarta)	2	4609 9252
TGF- $\beta$ receptor signaling pathway	3	993 5933 4609
TNF- $\alpha$ /NF- $\kappa$ B signaling pathway	3	2177 9252 4176

因子结合位点较为一致,因此筛选出对此蛋白网络起着重要作用的转录因子包括 SP1、MAZ、LEF1、FOXO4、NFAT、E12、PAX4 等。其中 SP1 能够调控多种肿瘤相关基因的表达,参与了肿瘤的增值、分化和凋亡,并能够调控雄激素受体以及 TGF- $\beta$  (Sankpal et al., 2011) 这与前面通路富集的结果一致。对关键基因相互作用网络的中间蛋白进行通路富集分析,可以看出 TGF- $\beta$  通路,雄激素受体信号通路等与网络中的蛋白密切相关,说明与大肠癌转移相关的蛋白相互作用网络中这些信号通路参与了重要作用,TGF- $\beta$  通路,雄激素受体信号通路与体内激素水平密切相关。*E2F1* 转录因子结合位点在 T1T2 期的大肠癌高表达基因转录调控区域中都存在并富集,说明 *E2F1* 基因在早期大肠癌对肿瘤的生长起着重要作用。*E2F1* 在细胞周期中参与了 DNA 的合成、修复,细胞的 G<sub>1</sub> 期到 S 期的调定点,以及有丝分裂和

凋亡等功能(Zacharatos et al., 2004) 这与前面通路富集结果一致。高表达 *E2F1* 能够促进肿瘤的生长 (Zacharatos et al., 2004), 因此,在大肠癌转移前,*E2F1* 能够对肿瘤的增殖起重要作用。大肠癌的发生发展转移是一个由多组基因,多条信号通路共同作用,相互调控的结果,其涉及的基因众多,调控关系复杂,在此通过生物信息学的方法,能够很好的找到表达调控的关键节点,对指导实验研究有着重要意义。本文寻找到了多个基因表达调控的关键节点,对这些关键节点上进行干预,或许能够减缓甚至抑制大肠癌的发生发展转移,为临床的有效治疗提供更好的治疗靶点。

### 3 材料与方法

#### 3.1 数据来源

本文选取的是来自 NCBI 网站的 GEO 数据库中

芯片数据 GSE2109 系列的大肠癌数据进行分析, 此系列本身包含 2 158 个各类肿瘤的数据并包含各种临床信息。由于本研究主要集中于大肠癌 因此在这 2 158 例表达谱数据中挑选大肠癌数据并同时包含 TNM 分期数据的表达谱数据进行分析, 可挑选出 343 例大肠癌数据 为了更好的分析未发生转移的肿瘤与已发生远隔器官转移肿瘤的基因表达差异情况 从 343 例大肠癌表达谱数据中挑选出归类于 T1、T2 和 M1 分期的大肠癌表达谱数据共 121 例进行分析 将 TNM 分期中属于 T1、T2 的归为一类 M1 的归为一类。

### 3.2 分析方法

本文主要的实验方法是通过 GSEA 筛选出与大肠癌转移相关的 miRNA 和转录因子功能集, 由于每个功能集包含很多差异表达基因 因此对筛选出的不同功能集中的多个差异表达基因取交集 取交集后的差异表达基因通过 Gene Set Analysis Toolkit V2 工具进行 GO 分析, 归类, 分析其功能类。然后对这些差异表达基因所对应的蛋白通过 visant 工具构建蛋白质相互作用网络, 提取出除了差异基因对应蛋白以外的相互作用蛋白, 再对这些作用蛋白通过 Gene Set Analysis Toolkit V2 工具进行 Wikipathways 通路富集和转录因子富集。

### 作者贡献

第一作者齐鲁对本论文进行设计、数据搜集、分析以及论文撰写 通讯作者丁彦青教授对本论文进行指导及审阅。

### 致谢

感谢评审专家对论文的评审和修改建议。

### 参考文献

- Biechele T.L., Kulikauskas R.M., Toroni R.A., Lucero O.M., Swift R. D., James R.G., Robin N.C., Dawson D.W., Moon R.T., and Chien A.J., 2012, Wnt/beta-catenin signaling and AXIN1 regulate apoptosis triggered by inhibition of the mutant kinase BRAFV600E in human melanoma, *Sci. Signal.*, 5(206): ra3
- Damian D., and Gorfine M., 2004, Statistical concerns about the GSEA procedure, *Nat. Genet.*, 36(7): 663
- Hu Z., Snitkin E.S., and DeLisi C., 2008, VisANT: An integrative framework for networks in systems biology, *Brief Bioinform.*, 9(4): 317-325
- Kelder T., van Iersel M.P., Hanspers K., Kutmon M., Conklin B.R., Chris T., Evelo C.T., and Pico A.R., 2012, WikiPathways: Building research communities on biological pathways, *Nucleic Acids Res.*, 40(D1): 1301-1307
- Nakamura Y., Migita T., Hosoda F., Okada N., Gotoh M., Arai Y., Fukushima M., Ohki M., Miyata S., Takeuchi K., Imoto I., Katai H., Yamaguchi T., Inazawa J., Hirohashi S., Ishikawa Y., and Shibata T., 2009, Krüppel-like factor 12 plays a significant role in poorly differentiated gastric cancer progression, *Int. J. Cancer.*, 125(8): 1859-1867
- Sankpal U.T., Goodison S., Abdelrahim M., and Basha R., 2011, Targeting sp1 transcription factors in prostate cancer therapy, *Med. Chem.*, 7(5): 518-525
- Zacharatos P., Kotsinas A., Evangelou K., Karakaidos P., Vassiliou L.V., Rezaei N., Kyrroudi A., Kittas C., Patsouris E., Papavassiliou A.G., and Gorgoulis V.G., 2004, Distinct expression patterns of the transcription factor E2F-1 in relation to tumour growth parameters in common human carcinomas, *J. Pathol.*, 203(3): 744-753
- Zu Y.L., Maekawa T., Nomura N., Nakata T., and Ishii S., 1993, Regulation of trans-activating capacity of CRE-BP by phorbol ester tumor promoter TPA, *Oncogene*, 8(10): 2749-2758