

DOI:10.3724/SP.J.1008.2015.00619

基于系统生物学整合技术挖掘结直肠癌形成中的核心通路和驱动基因

苗 华^{1△}, 曹付傲^{2△}, 赵权权², 缪宗原¹, 叶 淳¹, 徐小雯², 王汉涛^{2*}

1. 浙江省平湖市第一人民医院外科, 平湖 314200
2. 第二军医大学长海医院肛肠外科, 上海 200433

[摘要] **目的** 探索结直肠癌(colorectal carcinoma, CRC)形成中的核心通路和关键基因。**方法** 利用 meta 分析技术从以往 5 项 CRC 发生相关转录组学研究中筛选癌及癌旁差异表达基因。采用 ComBat 方法合并 5 项研究的癌组织基因表达谱数据, 针对上述差异表达基因用 PCIT 软件进行共表达网络构建。利用 CFinder 软件揭示该共表达网络中存在的核心亚网络, 并用 Gather 软件确定主要核心亚网络所富集的生物学功能。以重要核心亚网络(或通路)为重点, 联合节点基因在 CRC 中的表达变化方向和相应染色体区域扩增或缺失的信息, 发现亚网络功能形成中的候选驱动基因。**结果** Meta 分析转录组学研究共发现差异表达基因 2 073 个, 其中在 5 项研究的癌组织中一致上调 1 174 个, 一致下调 899 个, 这些基因在 CRC 样本中形成的共表达网络包括 798 个基因节点和 1 462 条边, 存在 22 个核心亚网络。最大的核心亚网络由 77 个基因和 436 条边组成, 功能涉及细胞周期和增殖信号调控, *UBE2C*、*MYBL2*、*FAM83D*、*AURKA*、*TPX2* 等 11 个基因被预测为该信号功能的驱动基因。**结论** 细胞周期和增殖信号通路是结直肠癌发生中的核心通路, *UBE2C* 和 *AURKA* 等 11 个基因是该核心通路的驱动基因。

[关键词] 结直肠肿瘤; 核心通路; 驱动基因; 计算生物学

[中图分类号] R 735.3 **[文献标志码]** A **[文章编号]** 0258-879X(2015)06-0619-08

Integrated genomic analysis-based identification of core pathways and driver genes associated with colorectal carcinoma

MIAO Hua^{1△}, CAO Fu-ao^{2△}, ZHAO Quan-quan², MIAO Zong-yuan¹, YE Chun¹, XU Xiao-wen², WANG Han-tao^{2*}

1. Department of Surgery, The First People's Hospital of Pinghu, Pinghu 314200, Zhejiang, China
2. Department of Colorectal Surgery, Changhai Hospital, Second Military Medical University, Shanghai 200433, China

[Abstract] **Objective** To investigate the core pathways and driver genes associated with development of colorectal carcinoma (CRC). **Methods** Meta-analysis was employed to screen differently expressed genes between CRC and the adjacent normal mucosa in 5 studies. ComBat was used to combine the gene data of the 5 studies and then the differently expressed genes were used to construct a stable co-expression network in CRC using PCIT software. CFinder software was used to extract the core sub-networks of the constructed coexpression network, and Gather software was used for functional enrichment analysis of the core subnetworks. Finally, those genes with up-regulation and DNA gain in CRC or down-regulation and DNA loss were identified as driver genes. **Results** Our meta-analysis found 2 073 differently expressed genes between CRC and the adjacent normal mucosa, including 1 174 up-regulated and 899 down-regulated in 5 the CRC studies. A coexpression network was constructed with those differently expressed genes; it had 798 nodes, 1 462 edges, and 22 core subnetworks. The largest sub-network consisted 77 genes and 436 edges, and the function mainly involved cell cycle and proliferation regulation, with the driver genes including *UBE2C*, *MYBL2*, *FAM83D*, *AURKA* and *TPX2*. **Conclusion** Cell cycle and proliferation pathways are the core ones in the development of CRC, and 11 genes including *UBE2C* and *AURKA* are the driver genes of the pathways.

[Key words] colorectal neoplasms; core pathways; driver genes; computational biology

[Acad J Sec Mil Med Univ, 2015, 36(6): 619-626]

结直肠癌(colorectal carcinoma, CRC)是全球最常见的恶性肿瘤之一, 全球每年新发病例 100 多

[收稿日期] 2014-12-13 **[接受日期]** 2015-04-28

[作者简介] 苗 华, 副主任医师. E-mail: 13567330917@139.com; 曹付傲, 硕士, 主治医师. E-mail: caofuao@gmail.com

△共同第一作者(Co-first authors).

*通信作者(Corresponding author). Tel: 021-31161607, E-mail: hantaowang@126.com

万人,且有上升趋势^[1]。CRC的发病与饮食、环境、遗传等多种因素相关,其分子机制复杂,目前尚不明确。近年来,随着微阵列和高通量测序等组学技术的发展,组学检测逐渐进入广大生物医学研究者的工作平台。但是,一次组学检测常发现数百甚至上千CRC相关基因^[2-4],使研究者难以确定哪些是重要的,哪些是次要的。同时,组学检测费用昂贵,致使许多CRC研究的样本量较小,再加上CRC个体之间存在较大异质性^[5-7],使得不同组学研究所得到的差异基因表达谱有很大不同^[8]。因此,开发新的数据分析和整合策略对组学数据挖掘、发现重要的甚至新的CRC发生相关关键分子十分重要。

系统生物学观点认为:在大批疾病相关基因和通路中,并非所有都有重要意义,其中许多都是被动发生变化(“乘客”基因),只有部分基因的变化是驱动性的(“司机”基因),研究者应该发现那些发挥“司机”作用的基因和通路^[9-10]。在本研究中,我们首先利用meta分析技术从既往5项研究中寻找CRC发生相关差异表达基因,然后构建差异表达基因在癌组织中的共表达网络,进而发现共表达网络中的核心亚网络并分析其生物功能和基因节点的剂量-效应关系^[10],最后确定CRC发生中的核心通路和驱动分子,从而为今后CRC发生机制的研究提供候选分子。

1 资料和方法

1.1 基因表达谱芯片数据下载及预处理

从GEO数据库中搜索并下载5项与CRC发生相关的表达谱芯片研究(GSE8671^[11]、GSE22598^[12]、GSE23878^[13]、GSE9348^[14]和GSE37364^[15])的原始数据。所有研究的检测平台均为Affymetrix Plus 2.0,研究所包含的癌旁和癌组织样本均超过10例。具体情况为:GSE8671包含配对的癌和癌旁样本各32例;GSE22598包含配对的癌和癌旁样本各17例;GSE23878包含非配对的癌旁样本24例和癌样本35例;GSE9348包含癌旁样本12例和癌样本70例;GSE37364包含癌旁样本38例和癌样本27例。在R 3.0.1环境下,原始数据(.cel格式)首先用fRMA^[16]提取表达谱数据,基于barcode文件^[16]剔除那些在5项研究中均不表达的探针,再进一步基于Jetset文件^[17]选择那些标记为“Perfect”的探针,

最终留存10 717个探针进行深入分析。

1.2 差异表达基因分析

5项CRC相关研究的表达谱数据在格式调整后,用MetaDE软件^[18]继续在R统计平台下进行单独及合并的差异表达基因分析。具体过程为:首先根据MetaDE软件建议过滤掉30%低表达的基因和30%样本间变异小的基因,然后计算单独研究中癌和癌旁组织的差异表达基因(其中GSE8671和GSE22598为配对设计),最后用maxP法^[18]合并各单独研究的结果并选择假阳性率(false discovery rate, FDR)≤0.01的差异表达基因。本研究剔除5项研究中差异表达方向不一致的差异表达基因进行后续分析。

1.3 共表达网络的构建及可视化

为了构建差异表达基因在CRC中的共表达网络,首先利用sva软件包^[19]中的“combat”命令将5项CRC相关研究的表达谱数据合并,手工提取差异表达基因所对应的表达谱数据,之后用PCIT软件^[20]对癌组织样本进行共表达网络构建(PCIT可利用偏回归系数法去掉那些间接相关的共表达基因对)。为了获得稳定的共表达网络,我们将181个癌组织样本随机分成A组($n=90$)和B组($n=91$),计算A、B组以及(A+B)组的共表达网络,选择各组共表达分析有统计学意义且相关系数绝对值>0.6(且 $P<0.05$)的基因对,并在Cytoscape 3.1.0软件^[21]平台下获得3组研究一致存在的共表达网络。

1.4 核心网络的搜索及功能分析

对于形成的共表达网络,我们利用CFinder软件^[22]进行核心亚网络的搜索,其中参数K为4。对于CFinder所确定的核心亚网络,进一步用Gather^[23](<http://gather.genome.duke.edu/>)对其所富集的GO(Gene Ontology)和KEGG相关生物学功能进行分析,并选择Bayes因子≥6的功能为有统计学意义。

1.5 核心亚网络相关驱动基因的分析

那些mRNA水平受到DNA拷贝数影响的基因在癌症形成中更可能发挥“驱动”作用^[9-10,24]。为了确定核心亚网络中的驱动基因,我们从美国明尼苏达癌症中心cBioPortal网站(<http://www.cbioportal.org/>)下载所有核心亚网络基因在CRC样本^[25]中的基因拷贝数变异信息(即下载GISTIC文件);其中,“-2和-1”代表基因拷贝数缺失,“2和1”代表基因拷贝

数扩增,“0”代表基因拷贝数正常。接下来,首先计算核心亚网络基因在CRC样本中的扩增频率和缺失频率,然后联合这些基因在CRC中的表达变化方向,确定那些在癌组织中扩增频率高且表达升高和那些在癌组织中缺失频率高且表达降低的基因为亚网络功能形成中的候选驱动基因。

2 结果

2.1 CRC及其癌旁组织之间的差异表达基因 当 $FDR \leq 0.01$ 时,MetaDE 分析发现 GSE8671、

GSE22598、GSE23878、GSE9348 和 GSE37364 等 5 项研究分别获得 3 173、2 130、2 726、1 770 和 2 961 个差异表达基因。利用 maxP 法合并 5 项研究的 P 值后共得到 2 443 个差异表达基因,其中在 4 项及以上研究的癌组织中一致上调的有 1 304 个基因,一致下调的有 1 018 个基因;在 5 项研究的癌组织中一致上调的有 1 174 个基因,一致下调的有 899 个基因。图 1 为 2 443 个基因在 5 项研究中表达的整体情况。

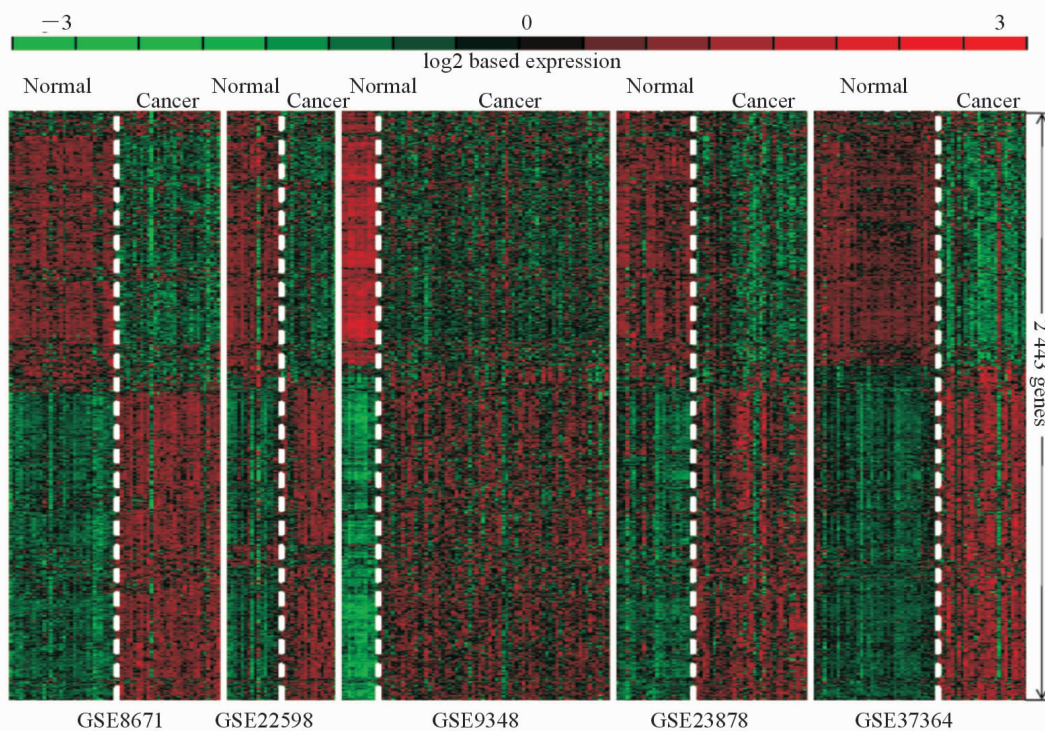


图 1 CRC 及其癌旁组织差异表达基因在 5 项研究中的表达模式

Fig 1 Expression pattern of differentially expressed genes identified from 5 studies between CRC and adjacent normal mucosa
CRC;Colorectal carcinoma

2.2 CRC 相关基因在癌组织中的共表达网络 在 5 项研究中一致上调或一致下调的 2 073 个基因被用于共表达网络构建。2 073 个差异基因在 A 组、B 组和(A+B)组中的两两相关系数的频数分布如图 2,发现有统计学意义的相关系数依次为 272 986、232 649 和 253 506 个(其中一个系数对应一个共表达基因对);进一步将 3 组中相关系数绝对值 >0.6 的挑出,依次获得 3 897、3 721 和 3 926 个基因对,分别对应 996、987 和 1 021 个差异表达基因。利用 Cytoscape 3.1.0 对 3 个共表达网络进行合并,发现

在 3 者中一致存在的网络是由 798 个基因和 1 462 条边组成。

2.3 核心亚网络及其生物功能 CFinder 从 CRC 共表达网络中发现 22 个核心亚网络,其中最大的核心亚网络有 77 个节点基因,其余核心亚网络从大到小依次为 29 个节点基因(1 个)、10 个节点基因(1 个)、9 个节点基因(1 个)、8 个节点基因(1 个)、7 个节点基因(3 个)、6 个节点基因(4 个)、5 个节点基因(2 个)和 4 个节点基因(8 个)。其中,最大的 5 个核心亚网络(SubNet 1~5)如图 3 所示。

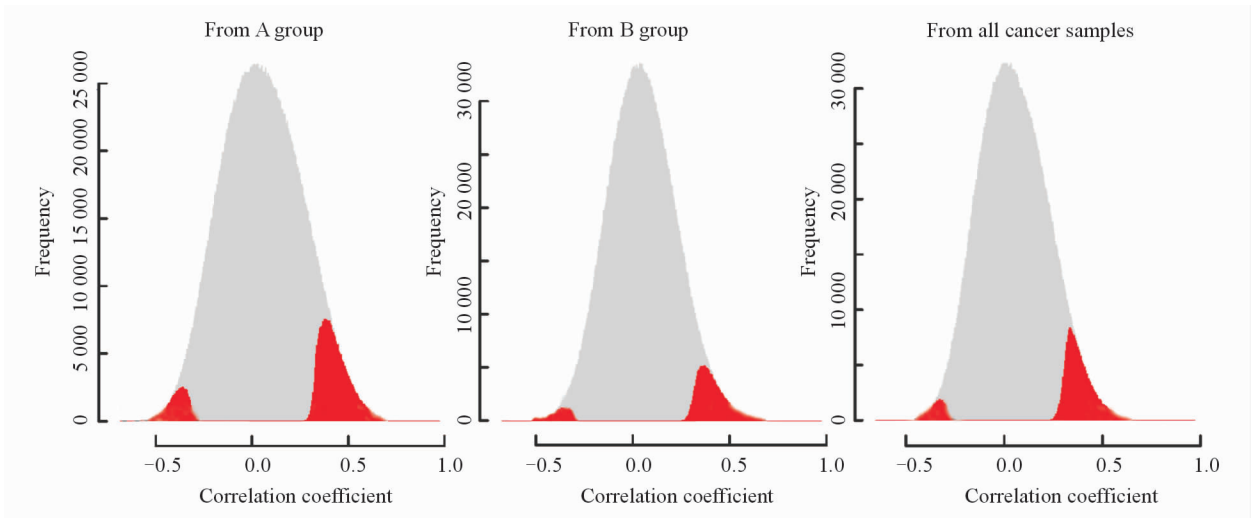


图2 A组、B组和(A+B)组样本中有意义共表达相关系数的分布

Fig 2 Distribution of meaningful correlation coefficients from samples of A, B, and (A+B) groups

Meaningful coefficients marked red. $n=90$ for A group, $n=91$ for B group, and $n=181$ for all

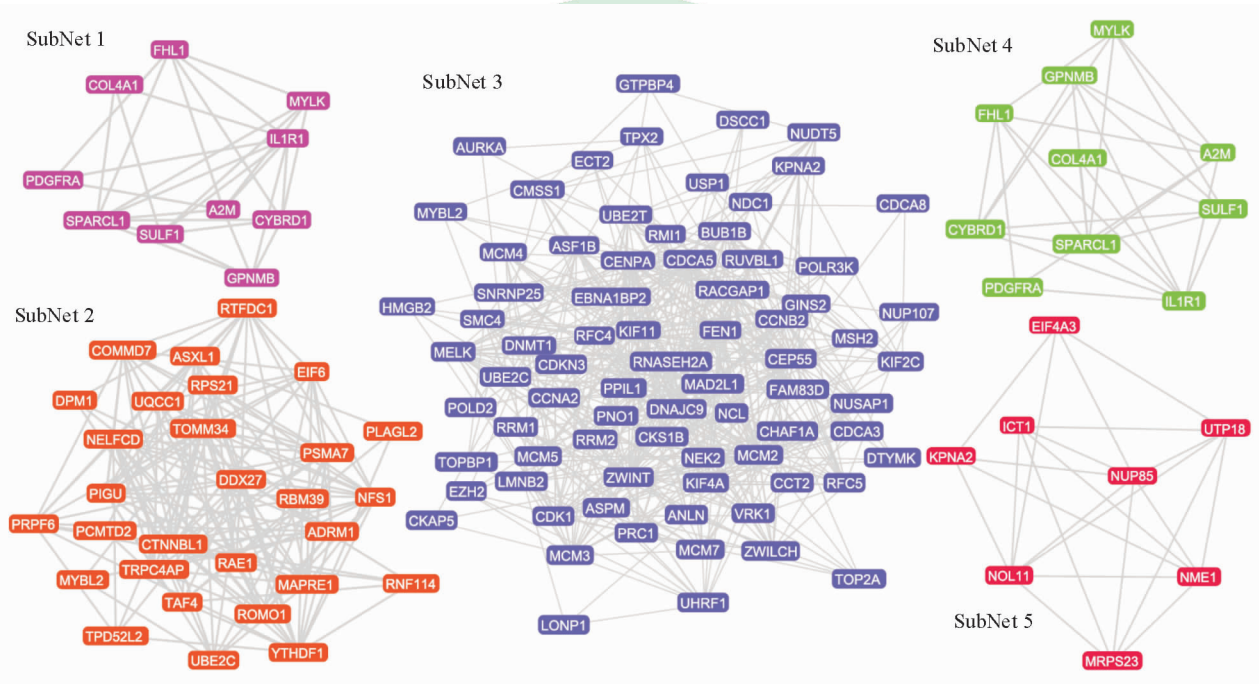


图3 CRC发生相关差异基因所形成的5个核心亚网络(SubNet 1~5)

Fig 3 Five core sub-networks (SubNet 1-5) constructed based on differentially expressed genes in CRC

亚网络功能富集度分析发现(表1):SubNet 1的功能主要涉及蛋白质等大分子的生物合成;SubNet 3的功能主要涉及细胞周期调控和增殖,该通路已被广泛报道与多种癌症发生密切相关^[9,26];SubNet 2、SubNet 4和SubNet 5未发现富集某种生物学功能,但其亚网络节点基因分别富集于不同的染色体(表1);尤其SubNet 2,其29个节点基因中分别有16个和11个位于染色体20q13和20q11,提示CRC组织DNA拷贝数变异影响SubNet 2中基因的表达。20q13和20q11染色体区域拷贝数增

加,以及20q13染色体区域多态性位点已被报道和CRC的发生显著相关^[24,27]。

2.4 核心亚网络关键基因的筛选 对于功能富集明确的SubNet 3,进一步考察其所包含基因在CRC中上调、下调方向和对对应染色体区扩增或缺失之间的对应关系。为此,我们首先计算cBioPortal网站中下载的257例CRC样本中20876个基因的扩增频率、缺失频率及扩增频率减去缺失频率的差值,并对所有的差值绘制概率密度图(图4A)。根据概率密度图界定:(扩增频率-缺失频率) > 0.23 为扩

增, (扩增频率 - 缺失频率) < -0.32 为缺失, 其余为正常。随后 SubNet 3 中 77 个基因在 CRC 中的拷贝数变异相应分为扩增、缺失和正常, 由于这些基因在 CRC 中均呈高表达 (图 1), 故 *UBE2C*、*MYBL2*、*FAM83D*、*AURKA* 和 *TPX2* 等 11 个基

因的高表达与染色体扩增正相关 (图 4B), 可能为细胞周期和增殖调控的驱动基因; 虽然 *BUB1B*、*CCNB2* 和 *NUSAP1* 等基因在 CRC 中高表达, 但其染色体区域却高频率缺失, 提示这 3 个基因的表达变化可能由其他机制 (如 miRNA 等) 调控。

表 1 SubNet 1~5 的功能富集度分析情况

Tab 1 Functional enrichment analysis of gene nodes in SubNet 1-5

	Enriched characteristics	Genes	Hit genes	Adjusted <i>P</i> value	Bayes factor
SubNet 1					
1	GO:0006412 [6]: protein biosynthesis	10	4	< 0.0001	9
2	GO:0009059 [5]: macromolecule biosynthesis	10	4	< 0.0001	9
3	GO:0044249 [5]: cellular biosynthesis	10	4	< 0.0001	7
4	GO:0006413 [8]: translational initiation	10	2	< 0.0001	7
5	GO:0009058 [4]: biosynthesis	10	4	< 0.0001	7
SubNet 2					
1	20q13	29	16	< 0.0001	56
2	20q11	29	11	< 0.0001	37
SubNet 3					
1	path:hsa04110: cell cycle	77	10	< 0.0001	20
2	GO:0007049 [5]: cell cycle	77	38	< 0.0001	63
3	GO:0008283 [4]: cell proliferation	77	41	< 0.0001	57
4	GO:0000067 [6]: DNA replication and chromosome cycle	77	18	< 0.0001	36
5	GO:0006260 [6]: DNA replication	77	16	< 0.0001	34
6	GO:0007067 [8]: mitosis	77	15	< 0.0001	32
7	GO:0000278 [6]: mitotic cell cycle	77	17	< 0.0001	32
8	GO:0006259 [5]: DNA metabolism	77	24	< 0.0001	32
9	GO:0000087 [7]: M phase of mitotic cell cycle	77	15	< 0.0001	32
10	GO:0000279 [6]: M phase	77	16	< 0.0001	31
11	GO:0000280 [7]: nuclear division	77	15	< 0.0001	28
12	GO:0000910 [5]: cytokinesis	77	10	< 0.0001	17
13	GO:0043283 [4]: biopolymer metabolism	77	29	< 0.0001	17
14	GO:0006261 [7]: DNA-dependent DNA replication	77	8	< 0.0001	15
15	GO:0006270 [8]: DNA replication initiation	77	5	< 0.0001	13
16	GO:0000074 [6]: regulation of cell cycle	77	13	< 0.0001	11
17	GO:0007154 [3]: cell communication	77	3	< 0.0001	10
18	GO:0007088 [7]: regulation of mitosis	77	5	< 0.0001	9
19	GO:0050875 [3]: cellular physiological process	77	63	< 0.0001	7
SubNet 4					
1	7p22	9	4	< 0.0001	15
SubNet 5					
1	17q25	8	3	< 0.0001	9

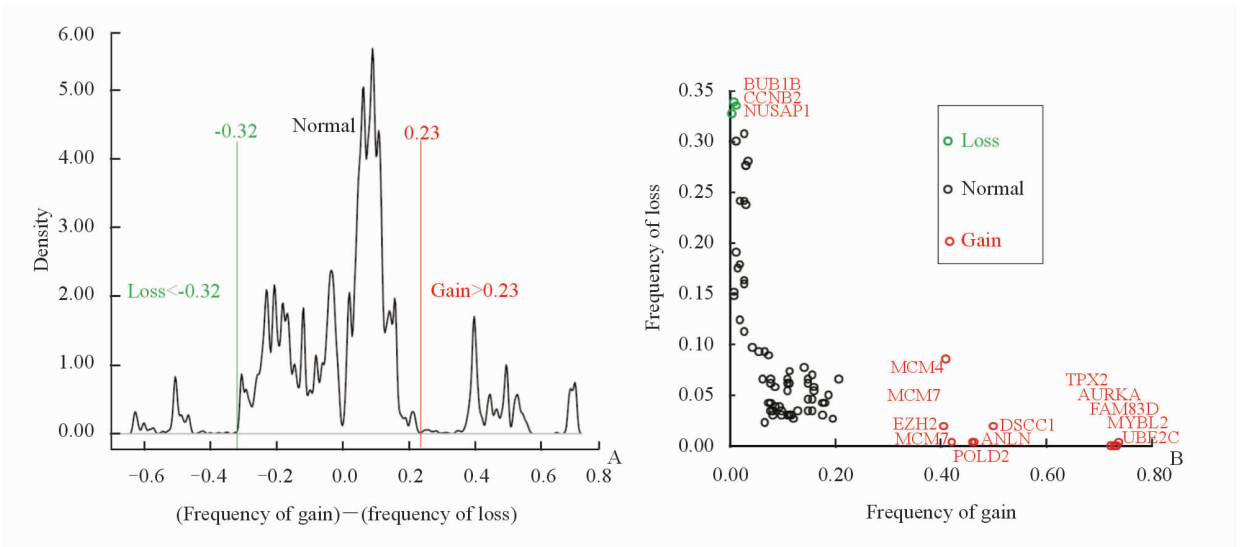


图4 核心亚网络 SubNet 3 中关键驱动基因的确定

Fig 4 Identification of the driver genes of SubNet 3 in CRC

A: Development of cut-off values for chromosomal aberrations in CRC; B: The relationship of mRNA expression and corresponding copy number variation of genes in SubNet 3. CRC: Colorectal carcinoma

3 讨论

CRC 的发生具有很大异质性^[5-7], 导致患者的临床表现和预后具有很大不同, 要研究 CRC 发生相关的基因表达谱必须具有较大的样本量。然而, 由于基因表达谱芯片价格昂贵, 很少有单个实验室能够使用大量的基因芯片进行 CRC 及其癌旁差异表达基因的筛选。在本研究中, 我们基于既往 5 项有关 CRC 发生的转录组学研究, 用 MetaDE 软件包中的 maxP 方法^[18] 整合发现 CRC 及其癌旁组织中存在 2 443 个差异表达基因, 其中一致性上调的有 1 174 个基因, 一致性下调的有 899 个基因, 5 项研究不一致的有 370 个基因(图 1)。这样的分析不仅扩增了研究的样本量, 也找到了不同实验室之间可以重复的研究结果, 显著提高了 CRC 发生相关基因的筛选效率。

表达谱芯片技术发现的数百甚至数千疾病相关基因往往组成生物学通路信号执行相关功能^[28], 故对于 CRC 差异表达基因的理解不能仅停留在基因倍数的差异, 而应该对其形成的信号网络进行探索。共表达网络是一种探索疾病相关基因间正向或负向相关变化的方法, 许多有共表达关系的基因更倾向于形成生物学通路信号^[20]。为了得到稳定的共表达网络, 我们首先将 5 项 CRC 研究中的癌组织表达谱数据合并, 随机分为 2 组, 并与所有样本组进行 3

次共表达网络构建, 取其共有且相关系数绝对值 > 0.6 的共表达基因对构成 1 个由 798 个基因和 1 462 条共表达关系组成的稳健的网络, 该网络清楚地展示了 CRC 相关差异表达基因在癌症表型维持中的组织模式。根据共表达网络的拓扑结构, CFinder 发现了 5 个较大的核心亚网络 SubNet 1~5(图 3), 提示这些亚网络有可能通过不同的功能促进 CRC 的发生和发展。尽管 SubNet 2、4、5 所富集的功能并不能用目前 KEGG 和 Gene Ontology 数据库所搜集的通路信号功能解释, 但是, 我们发现 SubNet 3 的功能主要涉及细胞周期和增殖调控, SubNet 1 的功能主要涉及蛋白质等大分子合成(表 1)。事实上, 细胞周期和增殖调控等生物学功能均报道与癌症的发展密切相关^[9, 26], 提示我们的共表达网络分析在揭示 CRC 发生机制上是有效的。

在癌症相关亚网络中的众多基因中, 许多基因的表达变化是被动的或伴随的, 只有少数基因才发挥“始动”作用(亦即“司机”基因), 发现这些具有始动作用的基因才是研究的关键^[9-10, 24]。许多致癌因素首先通过破坏基因组 DNA 稳定性, 然后导致 DNA 拷贝数变异和基因组突变, 并通过这些异常的结构变化来促进癌相关基因在 mRNA 和(或)蛋白层面量和质的变化^[28-29]。一些学者认为: 癌症相关差异表达基因中, 同时具有相应基因组结构变化(突变、扩增、缺失等)者可能是癌症形成的“司机”基

因^[9-10,24]。围绕该假设,我们进一步获得 SubNet 3 中所有基因在 CRC 组织中对应染色体区域扩增和缺失的频率,并将其与这些基因在 CRC 中的表达变化方向相结合,以发现 SubNet 3 中的“司机”基因。我们发现 *UBE2C*、*MYBL2*、*FAM83D*、*AURKA* 和 *TPX2* 等 11 个基因的高表达与染色体扩增正相关,提示这些基因是 SubNet 3 能形成的始动因素,是本研究中的关键基因(图 4)。事实上,*UBE2C* 和 *AURKA* 已被报道在细胞周期调控中发挥关键作用^[30-31]。

归纳而言,本研究利用 meta 分析的方法从既往 5 项 CRC 发生相关基因组研究中获得了 2 443 个癌和癌旁组织差异表达基因;并进一步经共表达网络分析发现细胞周期和增殖调控等生物学功能是 CRC 发生中的核心通路;最后,联合基因拷贝数变异数据发现 *UBE2C*、*MYBL2*、*FAM83D*、*AURKA* 和 *TPX2* 等 11 个基因的高表达与染色体扩增正相关,是 CRC 细胞周期和增殖调控通路的关键驱动基因,为今后课题组深入研究 CRC 发生的分子机制提供大量研究基础。当然,本研究也存在不足,例如在探讨基因表达变化和基因组拷贝数变异关系时,所用的信息来自不同的样本群,当两个样本群存在较大变异时,研究推论将会受到影响。

[参考文献]

[1] Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013[J]. CA Cancer J Clin, 2013, 63: 11-30.

[2] Xu Y, Xu Q, Yang L, Liu F, Ye X, Wu F, et al. The effect of colonoscopy on whole blood gene expression profile: an experimental investigation for colorectal cancer biomarker discovery [J]. J Cancer Res Clin Oncol, 2015, 14: 591-599.

[3] Solé X, Crous-Bou M, Cordero D, Olivares D, Guinó E, Sanz-Pamplona R, et al. Discovery and validation of new potential biomarkers for early detection of colon cancer[J]. PLoS One, 2014, 9: e106748.

[4] Wang Y, Zheng T. Screening of hub genes and pathways in colorectal cancer with microarray technology [J]. Pathol Oncol Res, 2014, 20: 611-618.

[5] Marisa L, de Reyniès A, Duval A, Selves J, Gaub M P, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization,

validation, and prognostic value[J]. PLoS Med, 2013, 10: e1001453.

[6] De Sousa E Melo F, Wang X, Jansen M, Fessler E, Trinh A, de Rooij L P, et al. Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions[J]. Nat Med, 2013, 19: 614-618.

[7] Hinoue T, Weisenberger D J, Lange C P, Shen H, Byun H M, Van Den Berg D, et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer[J]. Genome Res, 2012, 22: 271-282.

[8] Chang W, Gao X, Han Y, Du Y, Liu Q, Wang L, et al. Gene expression profiling-derived immunohistochemistry signature with high prognostic value in colorectal carcinoma[J]. Gut, 2014, 63: 1457-1467.

[9] Mine K L, Shulzhenko N, Yambartsev A, Rochman M, Sanson G F, Lando M, et al. Gene network reconstruction reveals cell cycle and antiviral genes as major drivers of cervical cancer[J]. Nat Commun, 2013, 4: 1806.

[10] Akavia U D, Litvin O, Kim J, Sanchez-Garcia F, Kotliar D, Causton H C, et al. An integrated approach to uncover drivers of cancer [J]. Cell, 2010, 143: 1005-1017.

[11] Sabates-Bellver J, van der Flier L G, de Palo M, Cattaneo E, Maake C, Rehrauer H, et al. Transcriptome profile of human colorectal adenomas [J]. Mol Cancer Res, 2007, 5: 1263-1275.

[12] Okazaki S, Ishikawa T, Iida S, Ishiguro M, Kobayashi H, Higuchi T, et al. Clinical significance of *UNC5B* expression in colorectal cancer[J]. Int J Oncol, 2012, 40: 209-216.

[13] Uddin S, Ahmed M, Hussain A, Abubaker J, Al-Sanea N, AbdulJabbar A, et al. Genome-wide expression analysis of Middle Eastern colorectal cancer reveals *FOXM1* as a novel target for cancer therapy [J]. Am J Pathol, 2011, 178: 537-547.

[14] Hong Y, Downey T, Eu K W, Koh P K, Cheah P Y. A ‘metastasis-prone’ signature for early-stage mismatch-repair proficient sporadic colorectal cancer patients and its implications for possible therapeutics [J]. Clin Exp Metastasis, 2010, 27: 83-90.

[15] Galamb O, Sipos F, Solymosi N, Spisák, S, Krenács T, Tóth K, et al. Diagnostic mRNA expression

- patterns of inflamed, benign, and malignant colorectal biopsy specimen and their correlation with peripheral blood results[J]. *Cancer Epidemiol Biomarkers Prev*, 2008, 17: 2835-2845.
- [16] McCall M N, Bolstad B M, Irizarry R A. Frozen robust multiarray analysis (fRMA)[J]. *Biostatistics*, 2010, 11: 242-253.
- [17] Li Q, Birkbak N J, Györfy B, Szallasi Z, Eklund A C. Jetset: selecting the optimal microarray probe set to represent a gene[J]. *BMC Bioinformatics*, 2011, 12: 474.
- [18] Wang X, Kang D D, Shen K, Song C, Lu S, Chang L C, et al. An R package suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection[J]. *Bioinformatics*, 2012, 28: 2534-2536.
- [19] Leek J T, Johnson W E, Parker H S, Jaffe A E, Storey J D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments[J]. *Bioinformatics*, 2012, 28: 882-883.
- [20] Watson-Haigh N S, Kadarmideen H N, Reverter A. PCIT: an R package for weighted gene co-expression networks based on partial correlation and information theory approaches[J]. *Bioinformatics*, 2010, 26: 411-413.
- [21] Shannon P, Markiel A, Ozier O, Baliga N S, Wang J T, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks [J]. *Genome Res*, 2003, 13: 2498-2504.
- [22] Adamcsek B, Palla G, Farkas I J, Derényi I, Vicsek T. CFinder: locating cliques and overlapping modules in biological networks[J]. *Bioinformatics*, 2006, 22: 1021-1023.
- [23] Chang J T, Nevins J R. GATHER: a systems approach to interpreting genomic signatures [J]. *Bioinformatics*, 2006, 22: 2926-2933.
- [24] Mekenkamp L J, Haan J C, Koopman M, Vink-Börger M E, Israeli D, Teerenstra S, et al. Chromosome 20p11 gains are associated with liver-specific metastasis in patients with colorectal cancer[J]. *Gut*, 2013, 62: 94-101.
- [25] Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer[J]. *Nature*, 2012, 487: 330-337.
- [26] Cuzick J, Swanson G P, Fisher G, Brothman A R, Berney D M, Reid J E, et al. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study[J]. *Lancet Oncol*, 2011, 12: 245-255.
- [27] Sillars-Hardebol A H, Carvalho B, Tijssen M, Beliën J A, de Wit M, Delis-van Diemen P M, et al. TPX2 and AURKA promote 20q amplicon-driven colorectal adenoma to carcinoma progression[J]. *Gut*, 2012, 61: 1568-1575.
- [28] Campbell P J, Yachida S, Mudie L J, Stephens P J, Pleasance E D, Stebbings L A, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer[J]. *Nature*, 2010, 467: 1109-1113.
- [29] Alhopuro P, Sammalkorpi H, Niittymäki I, Biström M, Raitila A, Saharinen J, et al. Candidate driver genes in microsatellite-unstable colorectal cancer [J]. *Int J Cancer*, 2012, 130: 1558-1566.
- [30] Xie C, Powell C, Yao M, Wu J, Dong Q. Ubiquitin-conjugating enzyme E2C: A potential cancer biomarker [J]. *Int J Biochem Cell Biol*, 2014, 47: 113-117.
- [31] Mehra R, Serebriiskii I G, Burtneß B, Astsaturov I, Golemis E A. Aurora kinases in head and neck cancer [J]. *Lancet Oncol*, 2013, 14: e425-e435.

[本文编辑] 孙岩