

Identification of Lung Cancer Related Function Modules Based on Co-Expression Network*

Yana Lv^{1#}, Yuehan He^{2#}, Zhengqiang Miao², Xu Jia², Chenchen Feng², Lina Chen^{2†}

¹College of Basic Medicine, Harbin Medical University, Harbin

²College of Bioinformatics Science and Technology, Harbin Medical University, Harbin
Email: lvyana198623@163.com, heyuehan56@126.com, †chenlina@ems.hrbmu.edu.cn

Received: Mar. 21st, 2013; revised: Mar. 27th, 2013; accepted: Apr. 27th, 2013

Copyright © 2013 Yana Lv et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract: Objective: Identifying lung cancer disease-related functional modules is important to understand the mechanism of lung cancer. **Methods:** In this paper, we propose an integration method of mining disease-related functional module. Using microarray data of normal and lung cancer samples, firstly, rank-based method was applied to construct gene co-expression network. Secondly, gene co-expression modules were mined through Qcut, then disease-related functional modules were screened based on the joint measure of lung cancer differentially expressed genes and the functional consistency. **Results:** 7 significant disease-related functional modules were screened, which were closely linked with the development of lung cancer by literature confirmation. Further it found that our method could not only return the functional consistency modules, but also find two modules were associated with specific functional annotations named “virus response” that could not be identified by other methods. **Conclusions:** The method provided additional insights for finding new functional module, which will be helpful for the studies on the pathogenesis of human complex diseases.

Keywords: Co-Expression Network; Gene Expression; Module Mining; Lung Cancer

基于共表达网络挖掘肺癌相关模块*

吕亚娜^{1#}, 何月涵^{2#}, 苗正强², 贾 婧², 冯陈晨², 陈丽娜^{2†}

¹哈尔滨医科大学基础医学院, 哈尔滨

²哈尔滨医科大学生物信息科学与技术学院, 哈尔滨

Email: lvyana198623@163.com, heyuehan56@126.com, †chenlina@ems.hrbmu.edu.cn

收稿日期: 2013年3月21日; 修回日期: 2013年3月27日; 录用日期: 2013年4月27日

摘 要: 目的: 识别肺癌疾病相关功能模块, 对了解肺癌疾病的发病机制至关重要。**方法:** 本文提出一个挖掘疾病相关功能模块的整合方法。采用包括正常和肺癌样本的微阵列数据, 首先, 应用 rank-based 方法构建基因共表达网络; 其次, 通过 Qcut 挖掘基因共表达模块; 然后基于肺癌差异表达基因及基因模块功能一致性的联合测度, 最终筛选出疾病相关功能模块。**结果:** 研究发现, 我们的方法获得 7 个显著疾病相关功能模块, 经文献证实都与肺癌的发生发展有着密切的联系。进一步分析发现不仅能获得与传统方法功能一致的模块, 而且还发现了传统方法没有获得的病毒层面的两个模块(模块 351 和 352)。**结论:** 我们的方法能够有效地发现新的功能模块, 为探索癌症致病机理提供新的视角及依据。

关键词: 共表达网络; 基因表达; 模块挖掘; 肺癌

*基金项目: 国家自然科学基金(61272388); 黑龙江省自然科学基金(F201237); 黑龙江省研究生创新科研资金(YJSCX2012-209HLJ); 黑龙江省卫生厅科研课题(2012-810)。

#第一作者。

†通讯作者。

1. 引言

癌症是一种严重危害人类健康的常见病、多发病^[1,2],在所有类型的癌症中,肺癌的死亡率在5年内高达80%~85%^[3,4]。癌症是细胞失控生长的复杂疾病,而人类复杂疾病是由多个基因共同作用导致的^[5,6]。当代,人类复杂疾病研究的一个重要挑战是寻找与疾病相关的基因及基因之间的协同关系,并全面阐明其在复杂疾病致病过程中的作用机理^[7,8]。

基因之间的相互作用与复杂的生物分子网络系统息息相关,如代谢网络、蛋白质-蛋白质相互作用网络、基因调控网络和基因共表达网络等^[9-12],其中基因共表达网络可用于筛选候选基因、预测基因功能、识别疾病模块和生物学标记等,已迅速成为从分子生物学网络层面研究人类复杂疾病的热点^[13-17]。因此,通过基因共表达网络,我们可以更好了解人类复杂疾病的致病机理,从而提高疾病的诊断和治疗。但现有的构建基因共表达网络的方法,几乎都是利用基因之间相似测度,统一的阈值,两个基因之间的相似性值在一定的阈值以上,就把两个基因连接^[18,19]。有研究表明使用同一个阈值即 value-based 方法,对研究生物学网络中基因的关联性具有显著的限制^[20,21]。基因通常是通过不同机制调控的,基因间的不同功能聚类可由不同的共表达模式来揭示。因此,基因可能在一个功能聚类中有强的相互共表达关系,而在另一功能聚类中可能是弱的共表达关系。如果选择一个严格全局阈值,弱相关的基因就没有被连接,那么在共表达网络中会导致丢失一些有意义的基因。Ruan 等人^[22]提出一个构建基因共表达网络的方法,rank-based 方法,它是利用局部相似阈值,不仅考虑了强的共表达基因之间的关联,还将弱的共表达基因连接到网络中。

网络的一个重要特性是具有模块化的结构,而模块中的节点密集彼此连接,属于同一模块中的生物分子彼此相互作用进行一个特定的生物功能,而在不同的正常和疾病状态下展示相似的表达模式。现在有许多图论分割算法被开发,如分子复合物检测(MCODE)、限制邻近搜索聚类(RNSC)、K-均值聚类、马尔可夫聚类(MCL)^[23-26]等。然而,在这些算法中都存在一定的局限性,例如:聚类,在图形分割中一个主要困难是确定分区的数量。有些算法虽不需要提前

确定分区数,但是它们要考虑其他的参数,也是很难获得,例如:MCL,一个最好的图形分割算法,它需要一个通货膨胀的参数,参数设置为不同的值,可导致不同的结果^[27]。Zhang W.等人^[28]提出一种有效的启发式算法 Qcut,是一种非参数方法,它能够在网络中自动确定最适当的模块数量。因此,基于基因表达谱构建基因共表达网络、聚类算法及疾病差异表达基因联合应用可以有效的预测疾病基因模块,并进一步借助功能富集分析可以对疾病基因模块进行更有效地预测和证实。

本文提出了一个整合的模块挖掘方法,采用具有正常和肺癌样本的表达谱,应用 rank-based 方法构建的基因共表达网络,通过非参数的图形分割算法 Qcut 挖掘基因共表达模块,结合肺癌差异表达基因和基因模块的功能一致性,筛选出疾病相关功能模块。研究发现,我们的方法不但能获得传统方法所获得到的大部分模块,而且还可以发现传统方法没有获得的病毒层面的模块,通过文献证实这些模块与肺癌紧密相连,这为肺癌的研究提供新的视角及依据。

2. 材料与方法

2.1. 实验材料

在本研究中,肺癌表达数据 GSE7670 来自于 GEO (<http://www.ncbi.nlm.nih.gov/geo/>)数据库,包含 22,284 个探针,56 个样本,其中包括 28 个正常样本和 28 个肺癌样本。

应用 DAVID 数据库(<http://david.abcc.ncifcrf.gov/>)中的功能注释工具,通过超几何分布对差异基因模块进行 Gene Ontology 生物学功能注释和 KEGG 生物学通路富集分析(显著性水平为 0.05),来进一步筛选疾病相关功能模块。

2.2. 实验方法

为了预测肺癌疾病相关功能模块,我们提出了一个整合的模块挖掘方法,采用具有正常和肺癌样本的表达谱,应用 rank-based 方法构建的基因共表达网络,利用 Qcut 算法挖掘基因共表达模块,结合肺癌差异表达基因和基因模块的功能一致性,筛选出疾病相关功能模块,其具体流程如下(如图 1)。

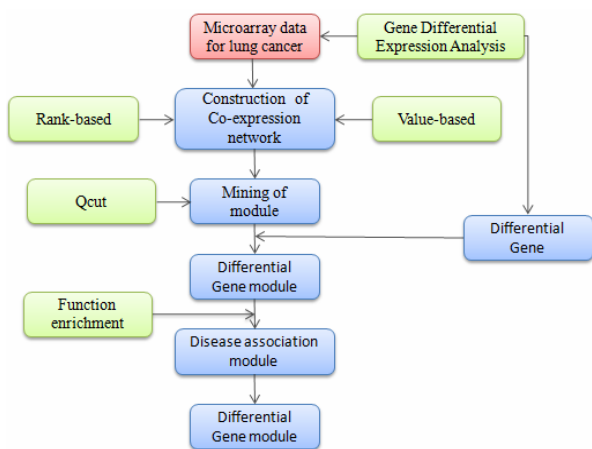


Figure 1. Work flow
图 1. 流程图

2.2.1. 构建共表达网络

用皮尔森相关系数表示基因之间相似性^[29]:

$$r = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N \sum x_i^2 - (\sum x_i)^2} \sqrt{N \sum y_i^2 - (\sum y_i)^2}} \quad (1)$$

其中 N 表示样本的数量, x_i 和 y_i 分别表示基因 x 和 y 在第 i 个样本中的表达值, r 表示两个基因的共表达系数。

在 value-based 方法中, 选择一个严格的全局阈值 R, 如果两基因之间的相似性 $r > R$, 就把这两个基因之间的关系视为共表达。在 rank-based 方法中, 首先, 计算每对基因之间的相似性 r, 对于每一个基因 x_i , 排列和这个基因 x_i 具有相似性的其他基因 y_i 。然后, 每一个基因连接 d 个和它最相似的基因。与 value-based 方法相比, rank-based 方法本质上, 针对不同的基因使用不同的局部相似性阈值。

2.2.2. 筛选差异表达基因

首先, 对于基因表达谱数据(GSE7670)进行预处理, 由于在表达谱中存在多个探针位点对应同一个基因, 也有同一个探针位点对应多个基因的情况, 因此我们需要对表达谱数据中重复出现的基因取平均, 用基因表达值得到的平均值代表该基因的表达值。其次, 应用 SAM 软件对正常样本和肺癌样本的表达谱数据进行差异表达基因筛选, 如果错误发现率 $FDR < 0.05$, 基因表达就被视为显著差异。

2.2.3. 基于整合的模块挖掘方法筛选疾病相关模块

根据基因模块的功能一致性特点, 我们提出一个

整合的模块挖掘算法, 即: 在 rank-based 方法构建的基因构建基因共表达网络中, 采用模块挖掘算法 Qcut, 结合肺癌差异表达基因和基因模块的功能一致性, 筛选出疾病相关功能模块。

模块检测: 模块分割算法, Qcut, 优化一个模块函数^[30]Q, 被定义为:

$$Q(\Gamma_k) = \sum_{i=1}^k (e_{ii} - a_i^2) \quad (2)$$

其中 Γ_k : 把一个组内的所有节点划分成 k 个组的聚类, e_{ii} : 第 i 个聚类中含有两个节点的边所占的比例, a_i : 第 i 个聚类中至少含有一个节点的边所占的比例。直观看, 就是观测值减去期望值, 所得差值越大, 说明分割就越准确。

疾病相关功能模块: 首先, 将至少包括一个差异基因的模块定义为差异基因模块。其次, 基于 DAVID 数据库, 对差异基因模块进行 Gene Ontology 生物学功能和 KEGG 生物学通路功能注释分析。最后, 将与疾病基因功能注释显著一致的差异基因模块($FDR < 0.05$)定义为疾病相关功能模块。

3. 结果

3.1. 构建基因共表达网络

利用包括正常和肺癌样本的表达谱数据 (GSE7970), 分别采用传统的 value-based 方法和 rank-based 方法构建基因共表达网络。在 value-based 方法中, 首先, 计算每对基因之间的皮尔森相关系数; 其次, 对每一条共表达关系进行 T 检验, 排除不显著的共表达关系, 选择 Pearson 相关系数的阈值在 0.7 以上, 其构建的共表达网络最为合理, 经反复试验, 我们最终确定 Pearson 相关系数的阈值为 0.77; 最后, 构建的 value-based 共表达网络由 6861 个节点, 68715 共表达关系组成。在 rank-based 方法中, 首先, 也是先计算每对基因之间的皮尔森相关系数; 其次, 分别选择 d 为 3、4 和 5 三个值可满足网络中高度可靠的共表达关系, 最终为了保证和 value-based 共表达网络中节点和边的规模相近, 我们保留每个节点 Pearson 相关系数值排秩在前 5 位($d = 5$)的共表达关系; 最后, 构建的 rank-based 共表达网络由 13,929 节点和 59,918 共表达关系组成。

3.2. 整合模块挖掘方法筛选疾病相关模块

为了获得肺癌相关功能模块，我们采用整合的模块挖掘方法，首先，利用 Qcut 算法在上述共表达网络中挖掘基因共表达模块；其次，结合肺癌差异表达基因和基因模块的功能一致性；最终，筛选出疾病相关功能模块。

3.2.1. 基因共表达模块挖掘

采用 Qcut 聚类算法针对 rank-based 和 value-based 构建的两个基因共表达网络进行模块挖掘，分别获得 1274 和 251 个基因模块(见表 1 和表 2)。

3.2.2. 差异基因模块筛选

从 GEO 数据库中下载包括正常和肺癌样本的表达谱数据 GSE29111。利用 SAM 软件进行差异分析，获得 1114 个肺癌差异表达基因。然后，基于肺癌差

Table 1. Module information of rank-based co-expression network (part)

表 1. Rank-based 共表达网络的模块信息(部分)

M_R	Size	Genes
r1	3	PRO2964, KITLG, TGFBR1
r2	4	E2F4, CDX1, DAXX, NAAA
r3	11	PTP4A3, KRT33B, LLGL1, CYP2U1, LOC100131613, SLC1A2, LOC100131062, FAM163A, SNN, AHSG, OR7A5
r4	3	ATXN7L1, CXCR5, APOC3
r5	3	KIAA1661, DNAH2, DDC
r6	4	TRAPPC9, ACSM2A, USP32, ACSM2B
r7	6	BAIAP2L2, FZR1, IFNA13, C14orf115, IFNA1, C2orf83
r8	7	S100A1, SLC5A3, C22orf31, DKFZP564O0823, AURKA, GCATP6, V0A4
r9	4	ERCC8, B3GALT4, KRT37, WFDC8
r10	8	ATP5G2, LOC732447, DAZ1, DAZ2, DAZ4, DAZ3, PGK1, PCDHB1

Note: M_R denotes ID of module in rank-based co-expression network, Size denotes size of module, Genes denote name of gene.
注: M_R 表示 rank-based 共表达网络中的模块 ID, Size 表示模块大小, Gene 表示基因的名字。

异表达基因，在 1274 个和 251 模块中分别筛选出 368 个和 47 个差异基因模块(如表 3 和表 4 所示)。

Table 2. Module information of value-based co-expression network (part)

表 2. Value-based 共表达网络的模块信息(部分)

M_V	Size	Gene
v1	4	MORC1, HADHA, LOC100132923, PMCHL1
v2	6	PNPLA3, KISS1, SYNM, PAX5, PCDHA2, LOC399904
v3	11	MED14, BCL7A, SSTR3, CYCSP52, CYCS, SIRT2, BMP7, PLCB1, KCNS1, MLANA, PKD2L1
v4	3	ZNF362, INPP4A, RAG2
v5	7	PRRX2, MASP2, NNAT, PHTF2, KANK2, SNX26, COL14A1
v6	14	CHD9, KIF3A, ATXN2L, GJB4, KIAA1751, ADH1A, MCM3AP, EPHA7, LOC728946, UBE3A, SNRNP200, KRTAP1-1, ZNF701, WBSR23
v7	4	CNGB3, GRIN2B, ELOVL5, GPR68
v8	7	KCNA6, NRAP, BAIAP2L2, KIAA1045, ZNF81, HNRPDL, FCER2,
v9	10	CCDC85B, MYT1L, HRH4, SSTR4, HTN1, HOXD12, C9, WT1, UBE2D1, SMURF1, WISP3
v10	4	ATXN7, RUFY3, TTBK2, AKAP13

Note: M_V denotes ID of module in value-based co-expression network, Size denotes size of module, Genes denote name of gene.
注: M_V 表示 value-based 共表达网络中的模块 ID, Size 表示模块大小, Gene 表示基因的名字。

Table 3. Information of differential gene module in rank-based co-expression network (part)

表 3. Rank-based 共表达网络中差异基因模块信息(部分)

D_R	Size	Genes
1	6	GMFG+, PFDN5, STX8, MRPL34, FAM32A, S1PR4+
2	3	VAT1, FBXO7+, VPS24+
3	4	RTN3+, BRPF1, FAM62A, OPTN+
4	8	FAM82A2, HK3+, SLC47A1, FABP4+, SFRP5, GPA33, CYP27A1, GPD1+
5	10	CRIP1, MARCO+, ST6GALNAC2, PTPN6, GPRC5A+, APOB48, R LTA4H+, S100A4+, TSPO, CD52+
6	3	HTT, PNPLA6+, PHF17+
7	11	FLVCR2+, VPS8, ITGAM, MRC1L1, PCOLCE2+, MAST3, SNX2, MSR1+, MRC1, VSIG4+, BACE1
9	5	HIGD1B+, PDE2A+, GIPC2+, AATK+, KIAA0040+
10	9	KAL1+, ADRB2+, FAM134B+, SPTBN1+, AGER+, CLIC5+, LOC100131610, TNNC1+, C10orf116+

Note: D_R denotes ID of differential gene module in rank-based co-expression network, Size denotes size of module, Gene denotes name of gene, "+" denotes differential gene.
注: D_R 表示 rank-based 共表达网络中差异基因模块 ID, Size 表示模块大小, Gene 基因的名字, 带有 "+" 的基因名字是差异基因。

3.2.3. 疾病相关功能模块筛选

为了筛选具有显著生物意义的差异基因模块，分

Table 4. Information of differential module in value-based co-expression network (part)
表 4. Value-based 共表达网络差异基因模块信息(部分)

D_V	Size	Gene
1	3	C20orf20+, MUC1, CCNO
2	6	PIH1D1, TNFRSF11, ABOLA1, ARHGEF16+, GYG2+, ABCF3
3	8	YIPF2+, DECR2,RPP40+, TOMM6, COX6C, NPM1+, TMED3+, PRICKLE4
4	4	ZCCHC2, PHF11+, ARGLU1+, UBA7
5	6	DCP1A, MKNK1, RGL1+, IFFO1, RUFY2+, CRT3+
6	10	COL1A1+, SULF1+, COL1A2+, COL3A1+, SERPINE2, COL5A1+, COL5A2+, MFAP2, LRRIC17, FAP
7	4	TRAV20, TRA@, TRD@+, TRAJ17
8	4	FBXO7+, CYB5R3+, ATF4, SPAG7
9	3	EIF4E2+, TIMM17A, PSMC1
10	3	MALL, CEACAM6, CD24+

别对两个网络获得的 341 和 33 个差异基因模块进行 GO 功能注释和 KEGG 通路富集分析, 最后, 筛选与疾病基因功能注释显著一致的差异基因模块(FDR < 0.05), 即疾病相关功能模块。分别筛选出与疾病基因功能注释(intrinsic to plasma membrane, mitochondrion, mitochondrial lumen 等)显著一致的 7 个疾病相关功能模块, 和与疾病基因功能注释(mitochondrion, plasma membrane part, RNA binding)显著一致的 10 个疾病相关功能模块。

3.2.4. 疾病相关功能模块分析

基于整合的模块挖掘方法, 获得疾病相关功能模块(见图 2), 我们发现在 rank-based 共表达网络和 value-based 共表达网络中, 共有 13 对模块具有相同功能, 而 rank-based 共表达网络中有 3 个模块的功能(模块 239、351 和 352)和 value-based 共表达网络中 2 个模块的功能(模块 14 和 21)存在差异。

我们分析两个网络中模块共有的功能, 发现部分模块的功能与 plasma membrane part、extracellular region、mitochondrion 等 12 个功能相关, 经报道这些

功能与肺癌是相关的。Selvendiran K.^[31]研究发现 Mg-ATPase 是细胞质膜标志酶, 其活性降低, 并将导致质膜功能的改变, 细胞膜电位下降, 质膜通透性增加, 这可能会导致癌细胞生理功能及分化能力的下降。质膜与必需经过翻译后再加工的 K-Ras 基因相关, Cho K.J.等人^[32]研究发现 K-Ras 质膜抑制剂对作为一种新型 K-ras 基因特定的抗癌疗法是具有实用性。Thomas K.J.等人^[33]研究表明线粒体裂变缺陷有助于瘤性肺上皮细胞(A549)的抗凋亡作用。

在 rank_based 共表达网络中 3 个功能特异的模块, 其富集的功能分别为: 模块 239 富集功能为 Axoneme 和 cilium, 另外发现模块 351 和模块 352 中的大部分基因富集在病毒应答功能类, 通过文本挖掘证实此功能与肺癌相关。模块 352 分别与白血病病毒-1HTLV-1、新城鸡瘟病毒 NDV、H1N1 病毒和人体免疫缺损病毒 HIV 等有着较强的生物学关联。Bian J. 和 Wang K.等人^[34]的研究发现, 新城鸡瘟病毒 NDV 具有潜在的溶瘤作用, 它可以通过影响 MAPK 通路, 进而诱导肿瘤细胞凋亡。在模块 352 中还发现存在与 NDV 密切相关的 ISG15 和 MUC4 等基因。Lu G.等^[35]发现, IRF-3 在抗病毒免疫反应过程中起到了重要作用; 而基因 ISG15 正是通过抑制 IRF-3 的降解, 进而增强个体的先天的抗病毒应答能力。Behr M.等人^[36]使用 EMSA 实验方法, 仅在病毒感染的 L929 细胞系中发现与人干扰素刺激基因 ISG15 密切相关的蛋白质复合物, 这暗示着对该基因以及基因相关的生物学过程进行研究或有可能揭示肺癌的内在机制。Nakaya T.等^[37]研究表明, 在病毒感染过程中, IFN-alpha/beta 和许多 IFN 诱导基因都可以激活宿主的抗病毒反应, 这些基因包含 IRF 绑定的瞬时元件, 在不同的生物通路中, IRF-3 或 ISGF3 体现出诱导 ISG15、ISG54 和 IP-10 基因的特性。肺癌与 HTLV-1 病毒也有很强的关联: Nomori H.和 Mori T.等人^[38]经过四年, 对 212 个肺癌患者的跟踪研究发现, HTLV-I 病毒可能是与验证或者免疫反应有关的细支气管肺泡癌的致病风险。此外, 对于模块 351 中包含的基因 SMAD5 和 PPCS, 现有研究^[39,40]发现它们是与 Epstein-Barr(EB)病毒相关的, 而 Guertler A.等人^[41]又发现 EB 病毒可以从年轻的肺癌患者身上发现放射性敏感性。而模块中的另一个基因 FUT6, Norden R.等人^[42]发现疱疹病毒可以通过

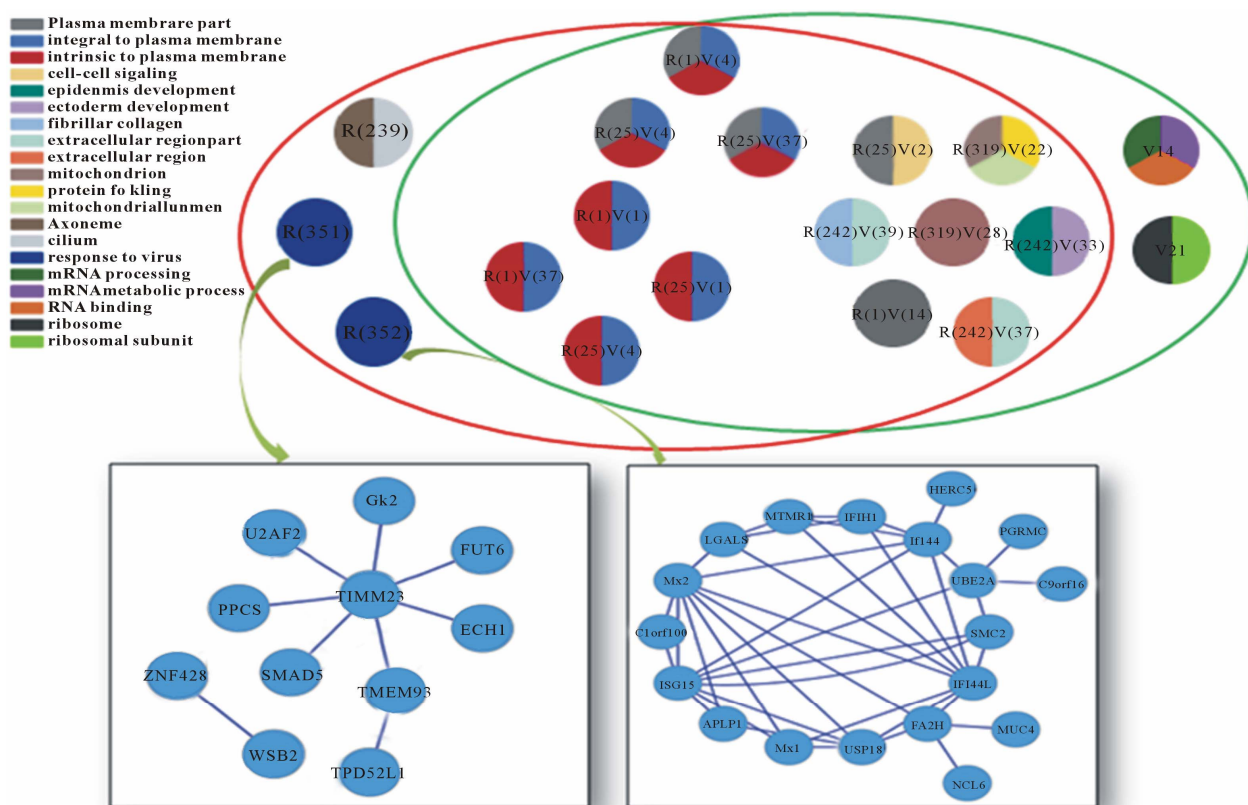


Figure 2. Disease-related functional modules in both network. Small circle within red oval and green oval intersect part, which represents the same function pair of module in the two networks, small circle within disjoint part, which represents module of function differences in two network, different colors represent different functional classes in small circle, the letter R and V in small circles, which represent rank-based expression networks and value-based co-expression network, respectively. Green arrow point to the graph, which is modules 351 and 352 visualization, the light blue represents gene, blue edges represent gene co-expression relationship.

图 2. 两个网络中疾病相关功能模块。红色椭圆和绿色椭圆相交内的小圆圈代表在两种网络中具有相同功能的模块对，不相交部分内小圆圈代表两种网络中功能存在差异的模块，小圆圈不同颜色代表不同的功能类，小圆圈上字母 R 和 V 分别代表 rank-based 共表达网络和 value-based 共表达网络。绿色箭头指向图形分别是模块 351 和模块 352 的可视化，图内浅蓝色节点代表基因，蓝色边代表基因间共表达关系。

激活 FUT3, FUT5, FUT6 基因的转录过程, 来诱导受累细胞中选择蛋白酶受体的表达, 而 Christensen C.L 等人^[43]证实靶向的疱疹病毒胞嘧啶脱氨酶尿嘧啶磷酸转移自杀基因疗法可以诱导肺癌特异性细胞毒性并延迟肿瘤的生长。在 value-based 共表达网络中的 2 个模块, 模块 14 富集功能为 mRNA processing、mRNA metabolic process 和 RNA binding, 模块 21 富集功能为 ribosome 和 ribosomal subunit, 对其功能进行的文献证实, 其富集功能也与肺癌相关。在模块 14 中还发现与 mRNA 密切相关的 SUZ12 和 KRAS 等基因。Xi S 等人^[44]实验证实 SUZ12 和 KRAS 是 miR-487b 的靶点, SUZ12 和 KRAS 的过表达是与原发性肺癌中 miR-487b 的抑制相关。

4. 讨论

本文提出一个挖掘疾病相关功能模块的整合方

法。利用包括正常和肺癌样本的微阵列数据, 应用 rank-based 方法构建基因共表达网络, 通过 Qcut 挖掘基因共表达模块, 然后联合肺癌差异表达基因和基因模块的功能一致性策略, 筛选出疾病相关功能模块。我们的方法获得了 7 个显著肺癌相关功能模块, 经文献证实与肺癌的发生发展有着密切的联系。进一步分析发现我们不仅能获得与传统方法功能一致的模块, 而且还发现了传统方法没有获得的病毒层面的两个模块。我们这种基于整合策略挖掘疾病相关功能模块的方法具有三个主要的优点:

1) 基于微阵列数据, 介绍了 rank-based 方法来有效地构建基因共表达网络。与传统构建方法相比, 我们构建的 rank-based 共表达网络, 能够更好捕捉到潜在生物系统的全局拓扑, 包括强的和弱的共表达模块。

2) 提出了一个整合的模块挖掘策略, 筛选出疾病相关功能模块。我们采用的 Qcut 算法, 有一个独立

特性, 参数是自由的。与传统图形分割算法方法相比, Qcut 算法可以挖掘的更有生物意义的模块, 联合肺癌差异表达基因及基因模块的功能一致性, 筛选出与疾病功能障碍显著相关的功能模块。

3) 可发现传统方法无法识别的新的疾病相关功能模块。将整合的模块挖掘方法, 应用在 rank-based 构建的共表达网络中不仅获得与传统的 value-based 方法构建的共表达网络功能一致的模块, 而且还发现了传统方法没有获得的病毒层面的两个模块, 这些模块功能经文献证实与肺癌的发生发展有着密切的联系。

5. 结论

为了预测肺癌疾病相关功能模块, 本文提出了一个整合的模块挖掘方法, 将其成功的应用到 rank-based 方法构建的基因共表达网络, 发现新的模块, 这为理解肺癌的发生的病理基础起到重要作用, 为研究肺癌的发生发展提供了新的研究思路, 同时也为研究其它复杂疾病提供了参考。

参考文献(References)

- [1] A. Corti, M. Giovannini, C. Belli and E. Villa. Immunomodulatory agents with antivasular activity in the treatment of non-small cell lung cancer: Focus on TLR9 agonists, IMiDs and NGR-TNF. *Journal of Oncology*, 2010: Article ID: 732680.
- [2] S. Y. Kim, W. C. Hahn. Cancer genomics: integrating form and function. *Carcinogenesis*, 2007, 28(7): 1387-92.
- [3] Y.-C. Wang, B.-S. Chen. A network-based biomarker approach for molecular investigation and diagnosis of lung cancer. *BMC Medical Genomics*, 2011, 4: 2.
- [4] S. N. Zahra, N. A. Khattak and A. Mir. Comparative modeling and docking studies of p16ink4/Cyclin D1/Rb pathway genes in lung cancer revealed functionally interactive residue of RB1 and its functional partner E2F1. *Theoretical Biology and Medical Modelling*, 2013, 10(1): 1.
- [5] M. Dupage, T. Jacks. Genetically engineered mouse models of cancer reveal new insights about the antitumor immune response. *Current Opinion in Immunology*, 2013, in press.
- [6] X. Fang, M. Netzer, C. Baumgartner, et al. Genetic network and gene set enrichment analysis to identify biomarkers related to cigarette smoking and lung cancer. *Cancer Treatment Reviews*, 2013, 39(1): 77-88.
- [7] L. Mariño-Ramírez, K. Tharakaraman, O. Bodenreider, et al. Identification of cis-regulatory elements in gene co-expression networks using A-GLAM. *Methods in Molecular Biology*, 2009, 541: 1-22.
- [8] F. M. Huang, H. C. Chen, M. A. Khan, et al. CYP2A6, CYP1A1, and CYP2D6 polymorphisms in lung cancer patients from Central South China. *Medical Oncology*, 2013, 30(2): 521.
- [9] E. Mosca, M. Barcella, R. Alfieri, et al. Systems biology of the metabolic network regulated by the Akt pathway. *Biotechnology Advances*, 2011, 30(1): 131-141.
- [10] H. Liu, J. Su, J. Li, et al. Prioritizing cancer-related genes with aberrant methylation based on a weighted protein-protein interaction network. *BMC Systems Biology*, 2011, 5: 158.
- [11] M. Cantile, R. Franco, G. Schiavo, et al. The HOX genes network in uro-genital cancers: Mechanisms and potential therapeutic implications. *Current Medicinal Chemistry*, 2011, 18(32): 4872-4884.
- [12] Y. S. Pyon, X. Li and J. Li. Cancer progression analysis based on ordinal relationship of cancer stages and co-expression network modularity. *International Journal of Data Mining and Bioinformatics*, 2011, 5(3): 233-251.
- [13] E. Prifti, J. D. Zucker, K. Clément, et al. Interactional and functional centrality in transcriptional co-expression networks. *Bioinformatics*, 2010, 26(24): 3083-3089.
- [14] K. Aoki, Y. Ogata and D. Shibata. Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant and Cell Physiology*, 2007, 48(3): 381-390.
- [15] S. Yu, L. Zheng, Y. Li, et al. Causal co-expression method with module analysis to screen drugs with specific target. *Gene*, 2013, 518(1): 145-151.
- [16] J. Zhang, Y. Xiang, L. Ding, et al. Using gene co-expression network analysis to predict biomarkers for chronic lymphocytic leukemia. *BMC Bioinformatics*, 2010, 11(Suppl 9): S5.
- [17] Z. J. hang, S. Ni, Y. Xiang, et al. Gene co-expression analysis predicts genetic aberration loci associated with colon cancer metastasis. *International Journal of Computational Biology and Drug Design*, 2013, 6(1-2): 60-71.
- [18] S. Lehtinen, F. X. Marsellach, S. Codlin, et al. Stress induces remodelling of yeast interaction and co-expression networks. *Molecular Biosystems*, 2013, in press.
- [19] L. López-Kleine, L. Leal and C. López. Biostatistical approaches for the reconstruction of gene co-expression networks based on transcriptomic data. *Briefings in Functional Genomics*, 2013, in press.
- [20] P. Tsaparas, L. Mariño-Ramírez, O. Bodenreider, et al. Global similarity and local divergence in human and mouse gene co-expression networks. *BMC Evolutionary Biology*, 2006, 6: 70.
- [21] I. K. Jordan, L. Mariño-Ramírez, Y. I. Wolf and E. V. Koonin. Conservation and coevolution in the scale-free human gene co-expression network. *Molecular Biology and Evolution*, 2004, 21(11): 2058-2070.
- [22] J. Ruan, A. K. Dean and W. Zhang. A general co-expression network-based approach to gene expression analysis: Comparison and applications. *BMC Systems Biology*, 2010, 4: 8.
- [23] P. G. Sun, L. Gao and S. Han. Prediction of human disease-related gene clusters by clustering analysis. *International Journal of Biological Sciences*, 2011, 7(1): 61-73.
- [24] M. Habibi, C. Eslahchi and L. Wong. Protein complex prediction based on k-connected subgraphs in protein interaction network. *BMC Systems Biology*, 2010, 4: 129.
- [25] A. Ebrahimzadeh, J. Addeh and Z. Rahmani. Control chart pattern recognition using K-MICA clustering and neural networks. *ISA Transactions*, 2011, 51(1): 111-119.
- [26] K. Kim, W. Kim and S. Kim. ReMark: an automatic program for clustering orthologs flexibly combining a Recursive and a Markov clustering algorithms. *Bioinformatics*, 2011, 27(12): 1731-1733.
- [27] S. Srihari, K. Ning and H. W. Leong. Refining Markov Clustering for protein complex prediction by incorporating core-attachment structure. *Genome Information*, 2009, 23(1): 159-168.
- [28] J. Ruan, W. Zhang. Identifying network communities with a high resolution. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2008, 77(1 Pt 2): Article ID: 016104.
- [29] L. Mao, Van J. L. Hemert, S. Dash, et al. Arabidopsis gene co-expression network and its functional modules. *BMC Bioinformatics*, 2009, 10: 346.
- [30] M. E. Newman, M. Girvan. Finding and evaluating community structure in networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 2004, 69(2 Pt 2): Article ID: 026113.
- [31] K. Selvendiran, D. Sakthisekaran. Chemopreventive effect of piperine on modulating lipid peroxidation and membrane bound enzymes in benzo(a)pyrene induced lung carcinogenesis. *Bio-*

- medicine & Pharmacotherapy, 2004, 58(4): 264-267.
- [32] D. van der Hoeven, K. J. Cho, X. Ma, et al. Fendiline inhibits K-Ras plasma membrane localization and blocks K-Ras signal transmission. *Molecular and Cellular Biology*, 2013, 33(2): 237-251.
- [33] K. J. Thomas, M. R. Jacobson. Defects in mitochondrial fission protein dynamin-related protein 1 are linked to apoptotic resistance and autophagy in a lung cancer model. *PLoS One*, 2012, 7(9): Article ID: e45319.
- [34] J. Bian, K. Wang, X. Kong, et al. Caspase- and p38-MAPK-dependent induction of apoptosis in A549 lung cancer cells by Newcastle disease virus. *Archives of Virology*, 2011, 2(3): 12-23.
- [35] G. Lu, J. T. Reinert, I. Pitha-Rowe, et al. ISG15 enhances the innate antiviral response by inhibition of IRF-3 degradation. *Cell and Molecular Biology*, 2006, 52(1): 29-41.
- [36] M. Behr, K. Schieferdecker, P. Bühr, et al. Interferon-stimulated response element (ISRE)-binding protein complex DRAF1 is activated in Sindbis virus (HR)-infected cells. *Journal of Interferon & Cytokine Research*, 2001, 21(11): 981-990.
- [37] T. Nakaya, M. Sato, N. Hata, et al. Gene induction pathways mediated by distinct IRFs during viral infection. *Biochemical and Biophysical Research Communications*, 2001, 283(5): 1150-1156.
- [38] H. Nomori, T. Mori, K. Iyama, et al. Risk of *Bronchioloalveolar carcinoma* in patients with human T-cell lymphotropic virus type 1 (HTLV-I): Case-control study results. *Annals of Thoracic and Cardiovascular Surgery*, 2011, 17(1): 19-23.
- [39] Q. Yin, X. Wang, C. Fewell, et al. MicroRNA miR-155 inhibits bonemorphogenetic protein (BMP) signaling and BMP-mediated Epstein-Barr virus reactivation. *Journal of Virology*, 2010, 84(13): 6318-6327.
- [40] S. Kashiwagi, T. Kumasaka, N. Bunsei, et al. Detection of Epstein-barr virus-encoded small RNA-expressed myofibroblasts and IgG4-producing plasma cells in sclerosing angiomatoid nodular transformation of the spleen. *Virchows Archiv*, 2010, 453(3): 275-282.
- [41] A. Guertler, A. Kraemer, U. Roessler, et al. The WST survival assay: An easy and reliable method to screen radiation-sensitive individuals. *Radiation Protection Dosimetry*, 2011, 143(2-4): 487-490.
- [42] R. Nordén, K. Nyström and S. Olofsson. Inhibition of protein deacetylation augments herpes simplex virus type 1-activated transcription of host fucosyltransferase genes associated with virus-induced sLex expression. *Archives of Virology*, 2010, 155(3): 305-313.
- [43] C. L. Christensen, T. Gjetting, T. T. Poulsen, et al. Targeted cytosine deaminase-uracil phosphoribosyl transferase suicide gene therapy induces small cell lung cancer-specific cytotoxicity and tumor growth delay. *Clinical Cancer Research*, 2010, 155(3): 305-313.
- [44] S. Xi, H. Xu, J. Shan, et al. Cigarette smoke mediates epigenetic repression of miR-487b during pulmonary carcinogenesis. *Journal of Clinical Investigation*, 2013, 123(3): 1241-1261.