

DOI: 10.3969/j.issn.1005-8982.2018.24.002

文章编号: 1005-8982 (2018) 24-0006-07

## 加权重基因共表达网络分析识别慢性肾脏病足细胞损伤关键节点基因 *MAGI2*\*

王万鹏<sup>1</sup>, 唐伯玉<sup>1</sup>, 沈剑箫<sup>2</sup>, 李永刚<sup>1</sup>, 蒯巧林<sup>1</sup>,  
高甄典<sup>1</sup>, 李娟<sup>1</sup>, 梁森林<sup>1</sup>, 陈皓瑜<sup>1</sup>

(1. 江苏省涟水县人民医院, 江苏 淮安 225400; 2. 上海交通大学医学院附属仁济医院, 上海 200127)

**摘要: 目的** 通过加权重基因共表达网络分析 (WGCNA) 识别慢性肾脏病 (CKD) 足细胞损伤相关的基因模块与关键基因。**方法** 从基因表达综合数据库 (GEO) 中下载包含足细胞损伤体外模型及 CKD 患者肾小球组份的表达谱 GSE66107、GSE93798、GSE30528、GSE32591 4 个数据集。利用 R 软件包做数据预处理并选取错误发现率 (FDR) < 0.05 及差异倍数  $\geq 1.5$  作为差异表达基因 (DEG); 结合数据集 GSE993395 (133 例 CKD 患者肾小球样本) 用 WGCNA 进行模块化分析并对模块内基因进行功能注释 (GO), 根据 GO 结果和含有文献报道的足细胞标准基因 (PSG) 情况确定足细胞损伤相关模块, 根据模块内基因连接度 (Kwithin) 挑选枢纽 (*hub*) 基因; 利用 Cytoscape 软件及插件 Cluego 构建足细胞损伤相关模块基因的加权重共表达网络。**结果** 4 个数据集共得到 趋势一致的 DEG 7 957 个, 其中 15 个 DEG 在 4 个数据集中均有差异 (coDEG)。GSE993395 中包含的 4 031 个 DEG 被用于 WGCNA 分析; 最终得到 12 个基因模块, 其中 green 模块包含最多的 coDEG 和 PSG, 进一步通过 Kwithin 发现其中同时作为 coDEG 和 PSG 的 *MAGI2* 基因是其 *hub* 基因之一。**结论** WGCNA 表达网络分析方法是一个高效的系统生物学方法, 应用该方法发现 CKD 时关键节点基因 *MAGI2* 在足细胞损伤中可能起到重要作用。

**关键词:** 生物信息学; 加权重基因共表达网络分析; 慢性肾脏病; 足细胞; 膜相关鸟苷酸激酶转化蛋白 2  
**中图分类号:** Q-33 **文献标识码:** A

## Identification of podocyte damage related *MAGI2* gene by weighted gene co-expression network analysis in chronic kidney disease\*

Wan-peng Wang<sup>1</sup>, Bo-yu Tang<sup>1</sup>, Jian-xiao Shen<sup>2</sup>, Yong-gang Li<sup>1</sup>, Qiao-lin Kuai<sup>1</sup>,  
Zhen-dian Gao<sup>1</sup>, Juan Li<sup>1</sup>, Sen-lin Liang<sup>1</sup>, Hao-yu Chen<sup>1</sup>

(1. Lianshui People's Hospital, Huaian, Jiangsu 225400, China; 2. Department of Nephrology, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai 200127, China)

**Abstract: Objective** To screen podocyte damage related critical genes in chronic kidney disease (CKD) patients by weighted gene co-expression network analysis (WGCNA). **Methods** Podocyte damage related database of GSE66107, GSE93798, GSE30528, GSE32591 were explored from gene expression GEO database. The R software package was used for data pre-processing. Differentially expressed genes (DEG) were identified for those with FDR < 0.05 and fold change  $\geq 1.5$ . DEG in GSE993395 data was analyzed by WGCNA. According to the functional annotation of Gene Ontology (GO) and the podocyte standard gene (PSG) which were reported in literatures, the podocyte injury-associated modules were defined, and the hub gene is selected according to the

收稿日期: 2017-12-01

\* 基金项目: 国家自然科学基金青年科学基金项目 (No: 81600549, 8170030521); 淮安市自然科学研究计划 (No: HAB201737)

average connectivity degree of the module (Kwithin). Then Cytoscape software and plugin Clugo was utilized for network visualization. **Results** Totally 7957 DEGs were picked out by comparing the expressed data of normal subject and podocyte or glomerulus under typical damage conditions, including 15 DEG that were differently expressed in 4 GSEs. About 4031 DEGs were screened by WGCNA, and 12 gene modules were identified. Green module contained most coDEGs or PSGs, which was suggested as podocyte injury-associated module. Furthermore, *MAGI2* gene was one of the hub genes that was coDEG and PSG. **Conclusions** WGCNA is an efficient system for screening of target genes; *MAGI2* may play an important role in podocyte damage in CKD.

**Keywords:** bioinformatics; weighted gene co-expression network analysis WGCNA; chronic kidney disease (CKD); podocyte; *MAGI2*

慢性肾脏病 (chronic kidney disease, CKD) 与足细胞损伤常常互为因果, 足细胞损伤机制极其复杂, 多种理化、生物因素均可影响足细胞的功能。本研究拟采用加权重基因共表达网络分析 (weighted gene co-expression network analysis, WGCNA) 挖掘从美国国立生物技术信息中心 (National Center for Biotechnology Information, NCBI) 的基因表达综合数据库 (gene expression omnibus, GEO) 下载到的足细胞损伤体外模型以及 CKD 患者肾小球表达谱数据集, 分析 CKD 状态下肾小球足细胞损伤后加权重基因共表达网络关系, 识别 CKD 状态下足细胞损伤的相关基因模块及模块内枢纽 (hub) 基因, 为进一步探索 CKD 足细胞损伤的发生、发展机制提供一种新思路。

## 1 材料与方法

### 1.1 数据来源

下载数据集 GSE66107、GSE93798, GSE30528、GSE32591 和 GSE99339, 其中 GSE66107 为嘌呤霉素氨基核苷 (purinomycin amino nucleoside, PAN) 诱导的足细胞损伤体外模型; GSE93798、GSE30528、GSE32591 分别为通过激光微切割获得的 IgA 肾病、糖尿病肾病 (diabetic kidney disease, DKD)、狼疮性肾炎 (lupus nephritis, LN) 患者肾小球 mRNA 表达数据, 以上 4 个数据集用于差异表达基因 (differentially expressed genes, DEG) 分析; 数据集 GSE99339 由多种 CKD 患者肾小球表达谱数据构成, 疾病包括: LN、IgA 肾病、膜性肾小球肾炎 (membranous glomerulonephritis, MGN)、高血压性肾病 (hypertensive nephropathy, HT)、局灶节段性肾小球硬化 (focal segmental glomerulosclerosis, FSGS)、肾肿瘤切除 (tumor nephrectomy, TN)、微小病变肾病 (minimal change disease, MCD)、DKD、薄基膜病 (thin membrane

disease, TMD), 共 133 例患者用于 WGCNA 分析。见表 1。芯片的探针注释信息来自 Affymetrix 公司下载的原始文件。

### 1.2 数据预处理

采用 R 语言 (3.4.0 版) 中的 Affy 包对原始数据进行预处理, 包括 ReadAffy 函数读取原始文件 (后缀名为 .CEL); 全局归一化矫正 (Robust Multi-array Average, RMA) 函数归一化及对数化; mas5calls 函数检测基因表达情况, 函数结果表现为“表达 (P)”或“缺失 (A)”, 本研究选择在各数据集样本中表达超过 50% 的探针用于后续研究; KNN 法补充缺失值; limma 包进行差异表达值计算, 并用贝叶斯方法进行多重检验校正, 最终选取阈值错误发现率 (false discovery rate, FDR) < 0.05 及倍数变化 (fold change, FC)  $\geq 1.5$  得到 DEG。

### 1.3 WGCNA

**1.3.1 模块构建及可视化** GSE99339 数据集包含 5 个批次 (batch), 使用 sva 包中 ComBat 函数去除批次效应后运行 WGCNA 包进行基因共表达网络的构建与模块鉴定。共表达网络模块内各基因利用 DAVID 6.7 (<https://david.ncifcrf.gov/>) 在线工具进行基因本体 (gene ontology, GO) 分析。利用 Cytoscape 绘制 WGCNA 导出的加权重共表达基因网络, 并通过插件 Cluego (v2.3.5) 与 Cluepedia (v1.3.5) 关联 String 数据库, 进行模块内基因 GO (功能) 和京都基因组百科全书 (Kyoto Encyclopedia of Genes and Genomes, KEGG) 通路富集分析与蛋白间互作网络 (protein-protein interactions, PPI) 可视化分析。Cluego 插件参数设置为  $pV \leq 0.05$ , 其余为默认参数。

**1.3.2 足细胞特异性阳性标准基因** JU 等<sup>[1]</sup> 在过去的研究中使用高通量数据以机器学习为基础发展处一种迭代算法, 通过该算法分别定义统计学意义上的

表 1 数据集纳入及分析结果

GEO 数据集	GSE66107		GSE93798		GSE30528		GSE32591			
	PAN	对照组	IgA 肾病	对照组	DKD	对照组	LN	对照组		
例数	3	3	20	22	9	13	32	14		
平台	GPL570		GPL22945		GPL571		GPL16443			
RNA 来源	足细胞		肾小球		肾小球		肾小球			
原始探针数	54 675		19 763		22 277		12 029			
剩余探针数	26 357		11 118		11 296		7 194			
DEG	6 154		3 882		1 422		840			
DEG-ALL					7 957					
coDEG					15					
GEO 数据集	GSE99339									
	LN	IgA 肾病	MGN	HT	FSGS	TN	MCD	DKD	TMD	FSGS-MCD
例数	30	26	21	15	12	11	7	7	3	1
平台	GPL19184				RNA 来源				肾小球	
原始探针数	12 134				剩余探针数				7 114	
包含 DEG					4 031					

50 种足细胞特异性标准基因 (podocyte standard gene, PSG)。PSG 包括: *ACTN4*、*NPHS1* 及 *NPHS2* 等。

## 2 结果

### 2.1 足细胞损伤体外模型及 CKD 患者肾小球差异表达基因分析

经过预处理并删除低表达探针 (表达率 <50% 数据集样本数) 后剩余探针分别为 26 357、11 118、11 296 及 7 194 个, 合并重复探针和删除无注释探针后, 最终分别得到 6 154、3 882、1 422、840 个 DEG (FDR<0.05, FC ≥ 1.5) (见表 1)。使用 R 语言 ggplot 包绘制 DEG 热图, 可见所得 DEG 可以区分出大部分 CKD 患者或者损伤的足细胞 (见图 1A~D)。本研究 4 个数据集共得到 9 217 个 DEG, 共同 DEG (coDEG) 15 个, 其中上调基因 4 个 (见图 1E), 下调基因 11 个 (见图 1F), 分别为 *CDKN1C*, *MYO5C*, *C1orf21*, *KBTBD11*, *NR3C2*, *FZD2*, *MAGI2*, *LPL*, *ZFPM2*, *USP46*, *PRKAR2B*, *TRIM38*, *DESI1*, *CTSS*, *ITGB2*。

### 2.2 WGCNA 构建基因共表达模块

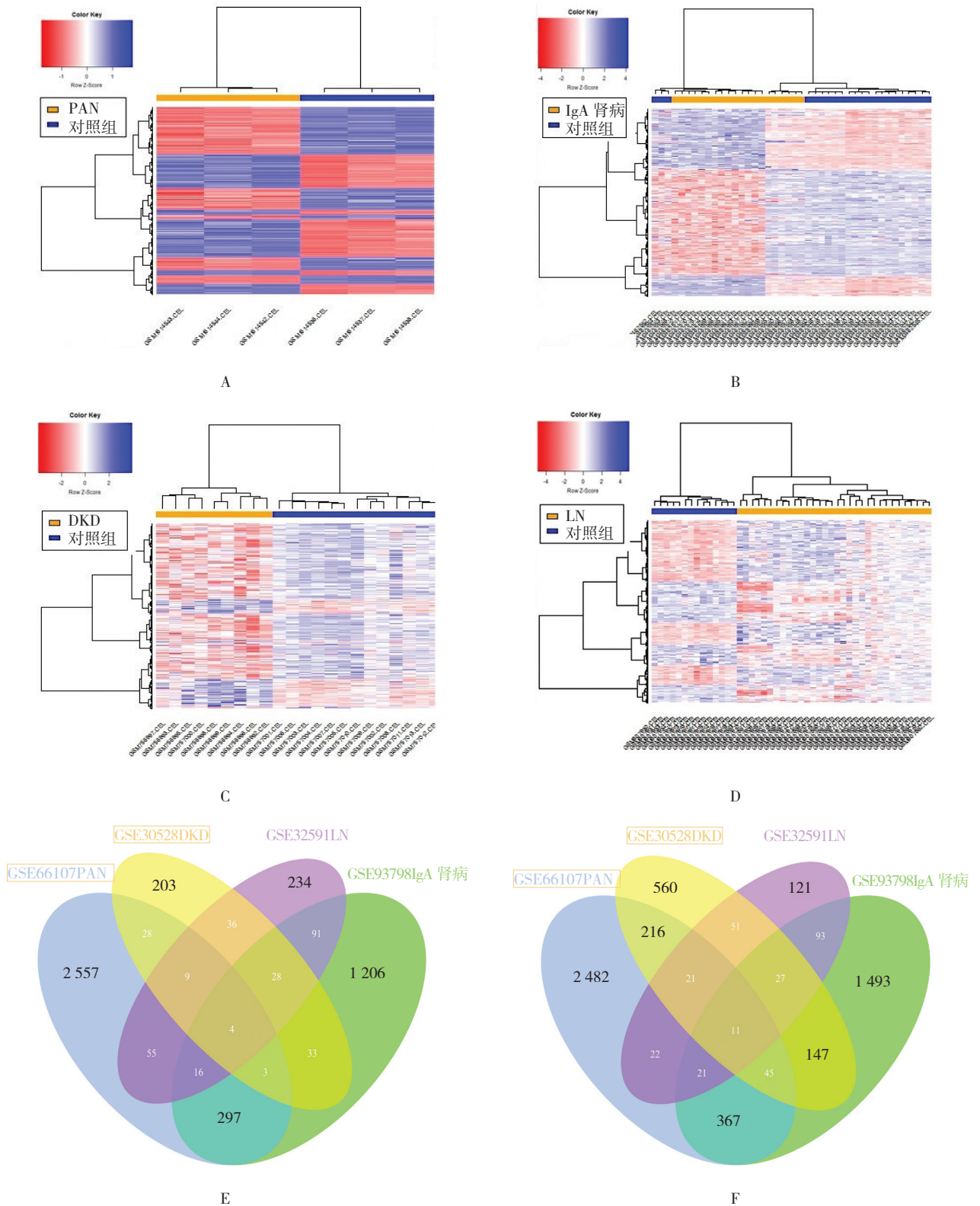
9 217 个 DEG 中有 1 260 个 DEG 表达趋势在 2 个或 2 个以上数据集中存在相反 (与对照组比较), 删除后剩余 7 957 个 DEG, 其中 4 031 个 DEG 存在于 GSE99339 预处理后的表达谱中, 被用于 WGCNA 分

析 (见表 1)。网络模块分析结果易受离群样本的影响, 因此在构建网络之前去除离群的样本数据显得尤为重要。相关研究常采用芯片间相关度 (inter. array correlation, IAC) 来评估芯片数据的分布情况。IAC 可以通过计算任意一对芯片上所有探针表达水平的相关系数获得, 芯片之间的关系可以用树形图来展示, 平均 IAC 值较低的样本, 或者在树形图上无法聚类的样本即为离群样本, 应被去除。本研究使用 WGCNA 包中的函数 `dist()` 计算样本间的 IAC, 并通过 `hclust()` 函数绘制 133 个样本 4 031 个 DEG 构成的树形图, 结果未发现明显离群值和疾病聚集 (见图 2A); 根据动态分层剪切树算法最终得到 12 个基因模块 (见图 2B), 以颜色的英文命名, 其中 grey 模块包含 518 个 DEG, 表示未纳入任何模块的基因集合。

### 2.3 足细胞损伤相关基因模块鉴定及分析

12 个基因模块与未归类的 grey 模块 (见表 2), 本研究中共包含 PSG 22 个, 其中尤以 green 和 red 模块最多, 分别为 10 和 4 个。同时通过 DAVID 进行模块内基因功能富集发现 green 和 red 模块所含基因与肾小球发育、细胞外基质解聚、细胞机械刺激等足细胞常见的病理状态密切相关, 因此笔者认为这两个模块与足细胞损伤关系较为密切。另一方面, 与两个模块比较, 发现 green 模块中包含更多的 coDEG, 因此可

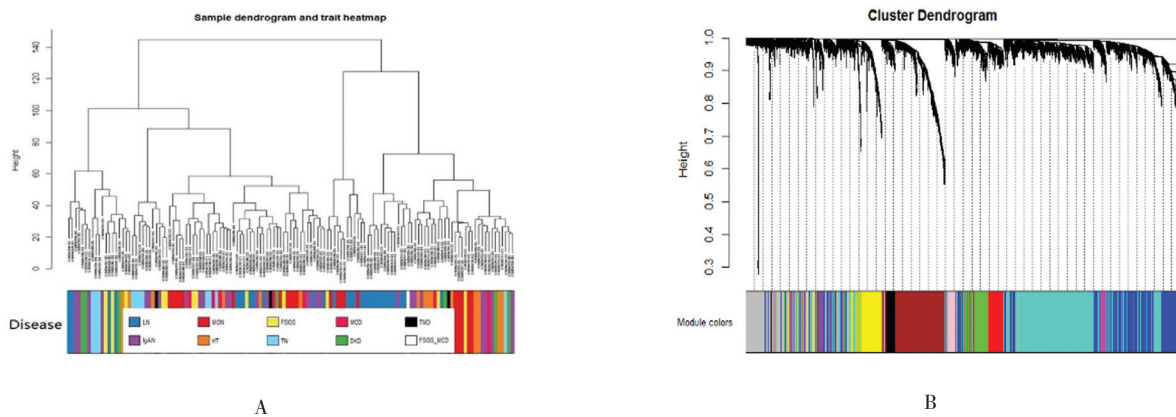
以推测 green 模块中的基因可能参与多种 CKD 发展中的足细胞损伤。



A: PAN 诱导的足细胞损伤差异基因 (GSE66107); B: IgA 肾病患者肾小球差异基因 (GSE93798); C: DKD 患者神效差异基因 (GSE30528); D: LN 患者肾小球差异基因 (GSE32591); E、F: 4 数据集差异基因韦恩图 (E: 上调; F: 下调)

图 1 各数据集差异基因表达热图及韦恩图





A: hclust ( ) 绘制 133 个样本 4 031 个 DEG 构成的树形图分析离群值; B: 动态剪切法构建基因共表达模块

图 2 WGCNA 分析构建基因共表达模块

### 2.4 足细胞损伤相关模块加权网络分析

hub 基因是指在模块中连接度最高的一系列基因, 与全局网络中的枢纽基因比较, 模块中的枢纽基因往往更有生物学意义。本文定义的 WGCNA hub 基因为模块内平均连接度 (Kwithin) 前 5% 的基因<sup>[2]</sup>。结果显示, green 模块中 MAGI2 基因可以同时作为即是 green 模块的 hub 基因也是本研究中的 coDEG 和 PSG (见表 2), 结果高度提示 MAGI2 不仅具有 CKD 足细胞损伤的分子标志物潜在价值, 也可能在早期足

细胞损伤中起到重要生物学作用。选择关联权重阈值  $\geq 0.05$  构建 mRNA 共表达网络 (见图 4A), 并使用 Cytoscape 及 ClueGO 插件并联合 String 数据库构建 green 模块内所有基因的 KEGG/GO/PPI 互作网络, 进一步挖掘模块内基因参与的生物学进程及蛋白相互作用, 结果显示与 MAGI2 表达或蛋白互作高度相关的多种基因已经被报道参与足细胞损伤, 包括 NPHS2、FGF1、BMP7 等 (见图 4B)。

表 2 构建出的模块资料及所含基因总结

Module	例数	PSG (n=22)	coDEG (n=15)	hub	coDEG-hub	PSG-hub
turquoise	1 214	IQGAP1 PALLD	-	61	-	-
blue	708	CASK AGRN	FZD2 ZFPM2 DESI1	36	-	-
brown	526	-	-	27	-	-
grey	518	-	LPL	26	-	-
yellow	235	PLAUR BASP1	MYO5C TRIM38 ITGB2 CTSS	12	ITGB2 CTSS	-
green	211	TJP1 FYN MME PLCE1 MAGI2 TCF21 NES CLIC5 NPHS2 DDN	CDKN1C KBTBD11 C1orf21 MAGI2 PRKAR2B NR3C2	11	MAGI2	MAGI2 PLCE1 CLIC5 MME
red	195	FAT1 SULF1 CD2AP UCHL1	USP46	10	-	SULF1
black	100	CDH13	-	5	-	-
pink	87	-	-	5	-	-
magenta	70	CX3CL1	-	4	-	-
purple	58	-	-	3	-	-
greenyellow	57	-	-	3	-	-
tan	52	-	-	3	-	-

注: PSG: 足细胞标准基因; coDEG: 4 数据集共同差异基因; hub: 根据模块内基因连接度 (Kwithin) 得出的差异基因; coDEG-hub: coDEG 与 hub 交叉基因; PSG-hub: PSG 与 hub 交叉基因

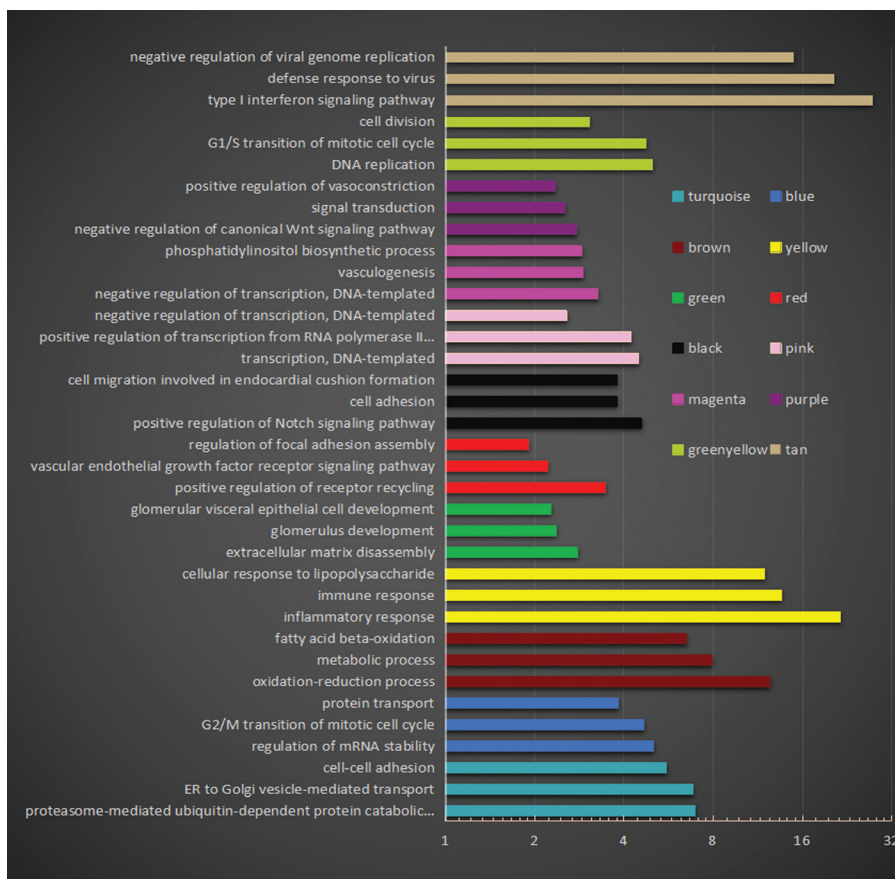
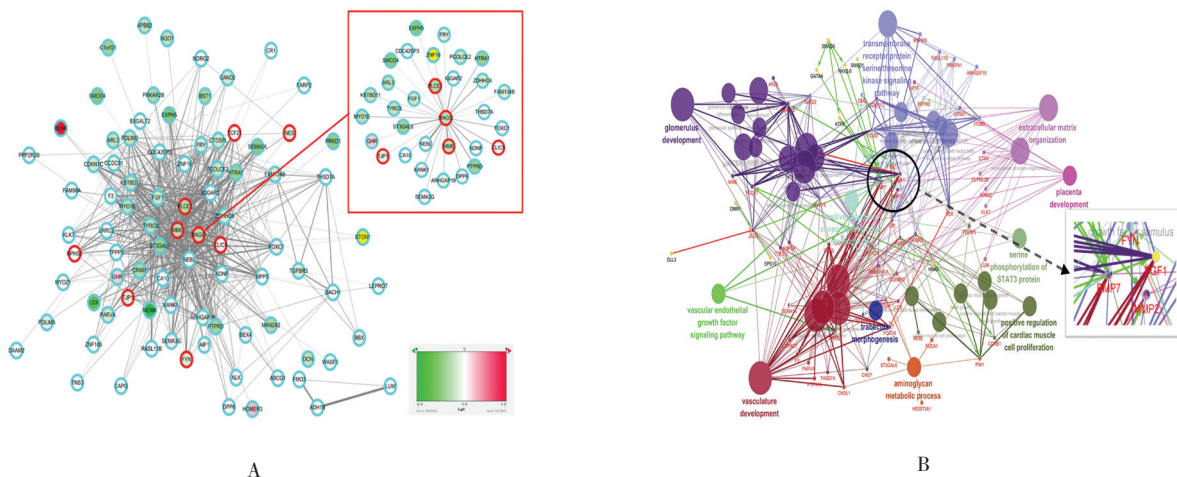


图 3 各模块功能富集结果 (取 P 值最小前 3 项)



A: WGCNA 导出的 green 模块加权基因共表达网络 (连接度  $\geq 0.05$ ), 红色边框表示足细胞标准基因 (PSG), 颜色如图例显示为在 GSE66107 差异基因表达变化 (对数取值); B: green 模块内所有基因的相关的功能注释 (GO)、分子通路 (KEGG) 及蛋白间相关作用网络。

图 4 green 模块内基因功能及相互作用网络可视化

### 3 讨论

CKD 时足细胞损害机制极其复杂, 涉及多种通路及细胞因子<sup>[3]</sup>。但传统的生物学研究以单基因或蛋白为出发点, 因此其仅能对生物系统的局部作出解

释, 难以对系统的整体进行全面的探索。相比之下, WGCNA 是一种从高通量数据中挖掘模块 (module) 信息的算法。在本方法中 module 被定义为一组具有类似表达谱的基因, 如果某些基因在一个生理过程或不同

组织中总是具有相类似的表达变化,那么有理由认为该基因在功能上是相关的,可以把其定义为一个模块(module)。这似乎有点类似于进行聚类分析所得结果,但不同的是,WGCNA的聚类准则具有生物学意义,而非常规的聚类方法(如利用数据间的几何距离),因此该方法所得结果具有更高的可信度,更有利于从生物功能整体入手考虑基因功能以及内在关联<sup>[4]</sup>。

本研究所纳入的GSE66107数据集是PAN诱导的足细胞损伤的经典体外模型,PAN是嘌呤霉素的衍生物,具有选择性的肾损害作用,PAN所致的肾脏损伤与人类肾病损伤一致,早期表现为足细胞足突融合和消失的微小病变(MCD),随着PAN作用时间的延长和蓄积量的增加,足细胞凋亡增加,进一步发展为局灶节段性肾小球硬化(FSGS),而PAN作用的靶细胞即为足细胞。因此通过足细胞体外损伤模型的基因表达变化与另3个CKD患者肾小球基因变化取交集,有利于排除肾脏其他实质细胞(如系膜细胞和内皮细胞)表达变化的干扰,可以更明确探讨CKD时足细胞内可能存在的基因表达异常。最终将4个数据集所得到的DEG结合含有133个CKD样本表达谱数据进行WGCNA分析,并通过GO富集分析发现red和green模块在GO分类上与多种足细胞损伤相关进程较为密切:如机械刺激反应(cell response to mechanical stimulus)<sup>[5]</sup>、内质网应激反应(response to endoplasmic reticulum stress)<sup>[6-7]</sup>、肾小球发育(glomerulus development)、细胞外基质分解<sup>[8]</sup>(extracellular matrix disassembly)等。更有价值的是,笔者发现green模块中的hub基因MAGI2也同为本研究中的coDEG和PSG,同时,网络可视化结果显示MAGI2与多种文献已报道的在足细胞损伤或肾脏病中起到重要“著名”基因相关量,如FGF1<sup>[9]</sup>、BMP7等<sup>[10]</sup>。这意味着MAGI2可能具有作为足细胞损伤早期分子标志物的潜在价值,并且在足细胞损伤中起到重要的调控作用,同时通过共表达网络的构建识别其高度相关的节点基因,对后续MAGI2的调控机制研究提供重要线索。

本研究不足之处在于从GEO数据库中下载的基因表达谱无法得到相应的临床资料,如肾小球滤过率(eGFR)、血肌酐(Scr)、尿蛋白等,如果可以加入该

临床参数作为参考,那么WGCNA分析时更可以将所得模块与该临床信息进行相关性分析,挖掘模块可能所具有的生物学意义,发挥WGCNA的更大作用。

综上所述,本研究通过运用一系列合理的生物信息学手段,分析CKD患者肾小球内与足细胞损伤相关度的15个差异变化基因,同时通过共表达分析构建出足细胞损伤相关基因模块green,研究结果对于CKD,尤其是足细胞损伤的早期诊断提供线索,同时为足细胞保护相关研究提供依据及参考,是WGCNA在CKD足细胞损伤中的一次新型尝试。

#### 参 考 文 献:

- [1] JU W, GREENE C S, EICHINGER F, et al. Defining cell-type specificity at the transcriptional level in human disease[J]. *Genome Res*, 2013, 23(11): 1862-1873.
- [2] LI S, LI B, ZHENG Y, et al. Exploring functions of long noncoding RNAs across multiple cancers through co-expression network[J]. *Sci Rep*, 2017, 7(1): 754.
- [3] 王福珍, 丘昭文, 朱杰, 等. 尿足细胞标志蛋白podocalyxin在肾病综合征合并急性肾损伤诊断中的价值[J]. *中国现代医学杂志*, 2015, 25(28): 36-40.
- [4] ZHAI X, XUE Q, LIU Q, et al. Colon cancer recurrence associated genes revealed by WGCNA coexpression network analysis[J]. *Mol Med Rep*, 2017, 16(5): 6499-6505.
- [5] KRIZ W, LEMLEY K V. A potential role for mechanical forces in the detachment of podocytes and the progression of CKD[J]. *J Am Soc Nephrol*, 2015, 26(2): 258-269.
- [6] CYBULSKY A V. Endoplasmic reticulum stress, the unfolded protein response and autophagy in kidney diseases[J]. *Nat Rev Nephrol*, 2017, 13(11): 681-696.
- [7] YOU Y H, QUACH T, SAITO R, et al. Metabolomics reveals a key role for fumarate in mediating the effects of NADPH oxidase 4 in diabetic kidney disease[J]. *J Am Soc Nephrol*, 2016, 27(2): 466-481.
- [8] ROEDER S S, BARNES T J, LEE J S, et al. Activated ERK1/2 increases CD44 in glomerular parietal epithelial cells leading to matrix expansion[J]. *Kidney Int*, 2017, 91(4): 896-913.
- [9] TOMASZEWSKI M, EALES J, DENNIFF M, et al. Renal mechanisms of association between fibroblast growth factor 1 and blood pressure[J]. *J Am Soc Nephrol*, 2015, 26(12): 3151-3160.
- [10] KAZAMA I, MAHONEY Z, MINER J H, et al. Podocyte-derived BMP7 is critical for nephron development[J]. *J Am Soc Nephrol*, 2008, 19(11): 2181-2191.

(王荣兵 编辑)