原创精选
Selected article

A549 细胞上皮 - 间质转分化中差异表达基因的生物信息学分析

杨果，李超，姚武，夏佳蕊，祁远梦，郝长付

郑州大学公共卫生学院尘肺病教研室，河南 郑州 450000

摘要：

[背景] 在肺纤维化中，上皮 - 间质转分化（EMT）具有重要的作用。肺泡上皮细胞的 EMT 是肌成纤维细胞的一个重要来源。而随着高通量测序技术的发展，利用生物信息学的分析筛选 EMT 过程中的关键基因（Hub 基因）引起了学者们的关注。

[目的] 分析人肺泡 II 型上皮细胞（A549 细胞）相关 EMT 过程中的差异表达基因，以此筛选肺纤维化相关 EMT 过程中的 Hub 基因以及相关的关键信号通路，以期为肺纤维化相关的 EMT 研究提供新的思路。

[方法] 从美国国立信息中心的基因表达综合数据库下载芯片数据集 GSE17708，运用 R 语言中的 limma 包进行差异表达基因的分析。采用 DAVID 6.8 对差异表达基因进行基因本体（GO）富集分析和京都基因与基因组百科全书（KEGG）信号通路分析，并运用相互作用基因库检索工具数据库和 Cytoscape 软件绘制蛋白质 - 蛋白质相互作用网络。利用 Cytoscape 软件中的 CytoHubba 插件筛选 Hub 基因，最后通过实时荧光定量聚合酶链式反应（qRT-PCR）技术对 A549 细胞相关 EMT 过程中的 Hub 基因进行验证。

[结果] 基于高通量测序芯片的筛选，55 个差异表达的基因主要参与的生物学途径包括胶原蛋白分解代谢的过程、细胞迁移正调控、生长负调节、内皮细胞凋亡过程的负调控以及细胞外基质组织；细胞组分的富集结果显示 55 个差异表达基因集中在胞外区和细胞外间隙；分子功能的富集显示差异表达基因主要具有细胞外基质结构组成、肝素结合、磷脂酰丝氨酸绑定、酶抑制剂的活动、纤连蛋白结合等功能。KEGG 结果表明，差异表达基因可能参与细胞外基质受体相互作用、阿米巴类感染、黏着斑的形成、PI3K-Akt 信号通路、矿物质吸收、血小板激活、补体及凝血级联反应、小细胞肺癌等信号通路。随后，通过最大团性中心方法筛选出得分最高的前 8 个基因（*THBS1*、*COL4A1*、*COL5A1*、*COL4A2*、*FGG*、*SERPINE1*、*LAMC2*、*IGFBP5*）作为 EMT 相关的 Hub 基因，qRT-PCR 结果显示 *IGFBP5* 在 A549 细胞中的表达丰度较低，并且与对照组相比，TGF-β 刺激 A549 细胞 24 h 和 72 h 的差异并不明显。除 *FGG* 在 TGF-β 刺激 A549 细胞 24 h 的时候表达情况与筛选结果不一致外，其他 6 个 Hub 基因表达情况与筛选结果一致。

[结论] 筛选出了 7 个与肺泡上皮细胞间质转分化相关的 Hub 基因，其中 *FGG* 表达下调，其他基因表达上调。

关键词：A549 细胞；上皮 - 间质转分化；差异表达基因；生物信息学；肺纤维化

Bioinformatics analysis on differentially expressed genes in epithelial-mesenchymal transition of A549 cells

YANG Guo, LI Chao, YAO Wu, XIA Jiarui, QI Yuanmeng, HAO Changfu (Department of Pneumoconiosis, College of Public Health, Zhengzhou University, Zhengzhou, Henan 450000, China)

Abstract:

[Background] Epithelial mesenchymal transition (EMT) plays an important role in pulmonary fibrosis. EMT of alveolar epithelial cells is a major source of myofibroblasts. Bioinformatic analysis used to screen hub genes in EMT process has attracted the attention of scholars as the high-throughput sequencing technology evolves.

[Objective] This study is conducted to screen potential hub genes and related key signaling pathways in the EMT process related to pulmonary fibrosis by analyzing the genes differentially expressed during the EMT of human alveolar type II epithelial cells (A549 cells), aiming to provide new ideas for the study of EMT related to pulmonary fibrosis.

DOI 10.13213/j.cnki.jeom.2021.20467

基金项目
国家自然科学基金项目（81773404）

作者简介
并列第一作者。
杨果（1994—），女，硕士生；
E-mail：1163256073@qq.com
李超（1996—），男，硕士生；
E-mail：lichao@gs.zzu.edu.cn

通信作者
郝长付，E-mail：haochangfu@126.com

伦理审批 已获取
利益冲突 无申报
收稿日期 2020-10-08
录用日期 2021-01-12

文章编号 2095-9982(2021)03-0245-09
中图分类号 R114
文献标志码 A

▸引用
杨果，李超，姚武，等. A549 细胞上皮 - 间质转分化中差异表达基因的生物信息学分析[J]. 环境与职业医学，2021，38（3）：245-253.

▸本文链接
www.jeom.org/article/cn/10.13213/j.cnki.jeom.2021.20467

Funding
This study was funded.

Correspondence to
HAO Changfu, E-mail: haochangfu@126.com

Ethics approval Obtained
Competing interests None declared
Received 2020-10-08
Accepted 2021-01-12

▸To cite
YANG Guo, LI Chao, YAO Wu, et al. Bioinformatics analysis on differentially expressed genes in epithelial-mesenchymal transition of A549 cells[J]. Journal of Environmental and Occupational Medicine, 2021, 38(3): 245-253.

▸Link to this article
www.jeom.org/article/en/10.13213/j.cnki.jeom.2021.20467

[Methods] An GSE17708 dataset was downloaded from the Gene Expression Omnibus of the National Center for Biotechnology Information, and used to identify differentially expressed genes by limma package in R language. DAVID 6.8 was used to perform Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathway analysis. An interaction gene database and Cytoscape software were used to draw protein-protein interaction (PPI) network. The hub genes were identified by the CytoHubba plugin of Cytoscape software. Finally, the hub genes in the process of EMT of A549 cells were verified by real-time fluorescence quantitative polymerase chain reaction (qRT-PCR).

[Results] A total of 55 differentially expressed genes were screened out based on high-throughput chip screening. The GO enrichment results showed that 5 enriched GO terms fell in the biological process category, namely collagen catabolic process, positive regulation of cell migration, negative regulation of growth, negative regulation of endothelial cell apoptotic process, and extracellular matrix organization; 2 enriched terms in the cellular component category, mainly in extracellular region and extracellular space; and 5 enriched terms in the molecular function category, containing extracellular matrix structural constituent, heparin binding, phosphatidylserine binding, enzyme inhibitor activity, and fibronectin binding. The results of KEGG showed that differentially expressed genes were mainly involved in extracellular matrix-receptor interaction, amoebiasis, focal adhesion, PI3K-Akt signaling pathway, mineral absorption, platelet activation, complement and coagulation cascades, and small lung cell cancer. Subsequently, the top eight genes with the highest scores were determined as hub genes related to EMT through the maximal clique centrality (MCC) method: *THBS1*, *COL4A1*, *COL5A1*, *COL4A2*, *FGG*, *SERPINE1*, *LAMC2*, and *IGFBP5*. The qRT-PCR results showed that the abundance of *IGFBP5* was lower and not significantly different in A549 cells after TGF-β stimulation for 24 h and 72 h compared with the control group. Except the expression of *FGG*, the expressions of the remaining 6 hub genes were consistent with the screening results of A549 cells after TGF-β stimulation for 24 h.

[Conclusion] Seven hub genes related to EMT are identified. The expression of *FGG* is down-regulated, and the expressions of *THBS1*, *COL4A1*, *COL5A1*, *COL4A2*, *LAMC2* and *SERPINE1* are up-regulated.

Keywords: A549 cell; epithelial mesenchymal transition; differentially expressed gene; bioinformatics; pulmonary fibrosis

肺纤维化是一种慢性、进行性和具有破坏性的肺间质性疾病。在我国，肺纤维化的发生常与环境和职业、高龄、吸烟等因素相关。各种致病因素导致肺泡结构破坏、炎症细胞的浸润、成纤维细胞增殖和分化、细胞外基质（extracellular matrix，ECM）异常沉积，从而造成正常肺组织破坏，肺功能受损[1]。肺纤维化的预后极差，治愈率非常低，确诊后患者的平均存活期只有3~5年[2]。近年来，随着研究的不断深入，发现肺上皮细胞的上皮-间质转分化（epithelial-mesenchymal transition，EMT）在肺纤维化发生过程中具有重要的作用[3-4]。EMT是指上皮细胞失去细胞紧密排列、缺乏运动能力、具有基底极性等上皮细胞特性，进而转化为具有细胞连接疏松，获得迁移能力以及丧失细胞极性的间质细胞特性，它参与了肺纤维化相关疾病的发生发展[5]。转化生长因子-β（transforming growth factor-β，TGF-β）是一种促进纤维化的细胞因子，是导致EMT发生的主要诱导剂，它能够通过TGF-β1/Smads通路、丝裂原活化蛋白激酶（mitogen-activated protein kinase，MAPK）通路和Wnt通路等多个信号通路诱导肺部EMT的发生[6]。A549细胞保留了II型肺泡上皮细胞的重要特征，TGF-β能够激活A549细胞并诱导其EMT的发生，而这种转变伴随着与肺纤维化进展有关的功能变化[7]。TGF-β诱导A549细胞发生EMT的体外模型已经成为研究肺纤维化相关EMT的常用体外模型[8]。因此，探究TGF-β刺激A549细胞相关EMT过程中差异表达基因，能够帮助了解肺纤维化相关EMT过程中起到关键作用的基因，本研究选取基因表达综合数据库（Gene Expression Omnibus，GEO）中的基因芯片数据集GSE17708，通过生物信息学的方法对EMT过程中的差异表达基因进行筛选，而后找出EMT过程相关的关键基因（Hub基因）并进行相应实验验证，以检验这些Hub基因的差异表达与芯片中的趋势是否一致，从而进一步认识EMT在肺纤维化过程的作用，为肺纤维化相关EMT的研究提供参考。

1 材料与方法

1.1 芯片数据来源

芯片数据集来自美国国立信息中心（National Center for Biotechnology Information，NCBI）创立的GEO数据库（https://www.ncbi.nlm.nih.gov/geo）中的数据集GSE17708，该芯片被纳入研究的原因：①包含多个时间点的数据，非单一时间点的测序数据；②该数据集中的样本多为3个重复样本，可靠性较强。芯片信息平台为GPL570［HG-U133_Plus_2］Affymetrix Human Genome U133 Plus 2.0 Array。该数据集包含26个样本，分为9组，即用5 ng·mL^{-1} TGF-β刺激A549细胞0、0.5、1、2、4、8、16、24、72 h。除2 h组只有2个重复外，其他组均为3个重复样本。5 ng·mL^{-1} TGF-β刺激A549细胞时间小于24 h，倒置相差显微镜下观察到A549细胞并未发生明显形态学上的变化。考虑到刺

激时间较短，其诱发EMT过程并不理想，因此本研究选择0、24、72 h三组，共9个样本。

1.2 数据预处理及差异表达基因的筛选

运用Perl语言将探针名称转化为基因名称，随后采用R语言中的limma包，以校正后的$P<0.05$和log_2基因表达值倍数变化（log_2FC）的绝对值>1为筛选条件，分别筛选出TGF-β刺激A549细胞24、72 h与对照组（0 h）相比的差异表达基因，随后通过FunRich（http://www.funrich.org/）软件进行交集分析得到共同的差异表达基因。

1.3 基因本体论（Gene Ontology，GO）富集分析和京都基因与基因组百科全书（Kyoto Encyclopedia of Genes and Genomes，KEGG）信号通路分析

采用DAVID 6.8（https://david.ncifcrf.gov）对差异表达基因进行GO富集分析和KEGG信号通路分析。GO富集分析从生物学途径、细胞组分、分子功能三个方面对基因进行分类注释，KEGG信号通路用以寻找EMT相关重要的信号通路。

1.4 蛋白质-蛋白质相互作用（protein-protein interaction，PPI）网络的构建及Hub基因的筛选

运用相互作用基因库检索工具（search tool for the retrieval of interacting genes，STRING）数据库（https://string-db.org）构建差异表达基因的PPI网络，以中等可信度（交互作用评分>0.4）作为筛选的标准[9]。随后采用Cytoscape V3.7.1对差异表达基因的PPI网络进行可视化，并选取其中的CytoHubba插件通过采用最大团性中心（maximal clique centrality，MCC）方法将评分较高的基因作为Hub基因[10-11]。

1.5 实时荧光定量聚合酶链式反应（quantitative real-time PCR，qRT-PCR）验证EMT相关转分化指标和Hub基因

A549细胞购买于中国科学院细胞库。将细胞培养在含体积分数为10%的胎牛血清、100 $mg·L^{-1}$链霉素以及$1.0×10^5$ $U·L^{-1}$青霉素的RIPM 1640（美国Thermo）完全培养基中，放置在体积分数为5%的二氧化碳、37℃恒温无菌细胞培养箱中。A549细胞以$2×10^5$个每孔铺于6孔板中，实验分为对照组和TGF-β处理组，用5 $ng·mL^{-1}$ TGF-β分别刺激24、48、72 h后采用Trizol试剂（日本TaKaRa）提取总RNA，采用PrimeScript™RT reagent Kit with gDNA Eraser试剂盒（日本TaKaRa）进行反转录，并用TB Green™ Premix Ex Taq™ II试剂盒（日本TaKaRa）进行实时荧光定量，用以验证EMT相关转分化指标和Hub基因。EMT相关转分化指标选择上皮细胞的标志物E-钙黏蛋白（E-cadherin，E-cad）和间质细胞的标志物波形蛋白（vimentin，VIM），用以判断TGF-β刺激A549细胞后是否诱发了EMT。首先，qRT-PCR检测24 h和72 h转分化指标以及*THBS1*、*COL4A1*、*COL5A1*、*COL4A2*、*FGG*、*SERPINE1*、*LAMC2*、*IGFBP5*的表达。而后，qRT-PCR检测48 h转分化指标和*THBS1*、*COL4A1*、*COL5A1*、*COL4A2*、*FGG*、*SERPINE1*、*LAMC2*的表达。引物由上海生工生物有限公司合成，引物序列如表1。数据采用$2^{-\Delta\Delta Ct}$法进行处理，并采用SPSS 21.0分析数据，检验水准为α=0.05（双侧）。

表1 qRT-PCR引物序列

Table 1 qRT-PCR primer sequence

基因 Gene	正向引物（5′→3′） Forward	反向引物（5′→3′） Reverse
THBS1 [12]	GCT CCA GTC CTA CCA GTG TC	TCA GTC ACT TGC GGA TGC T
FGG [13]	TTA TTG TCC AAC TAC CTG TGG C	GAC TTC AAA GTA GCA GCG TCT AT
LAMC2 [14]	TAC CAG AGC CAA GAA CGC TG	CGC AGT TGG CTG TTG ATC TG
SERPINE1 [15]	CCG CCT CTT CCA CAA ATC AG	AAT GTT GGT GAG GGC AGA GA
COL4A1 [16]	TGT GGG CCA GCC AGG CAT TG	CAG GGG GTC CGA TCG CTC CA
COL5A1 [17]	GCC CGG ATG TCG CTT ACA G	AAA TGC AGA CGC AGG GTA CAG
COL4A2 [16]	AGG GTC GCA GGG AGA GCT GG	TGG GCC TCG TTC CCC TGG AG
IGFBP5	GGA GGA GCC GAG AAC ACT G	GCG AAG CCT CCA TGT GTC
E-cad	AGG ATG ACA CCC GGG ACA AC	TGC AGC TGG CTC AAG TCA AAG
VIM	CCT TGA ACG CAA AGT GGA ATC	GAC ATG CTG TTC CTG AAT CTG AG
GAPDH	GCA CCG TCA AGG CTG AGA AC	TTG TGA AGA CGC CAG TGG A

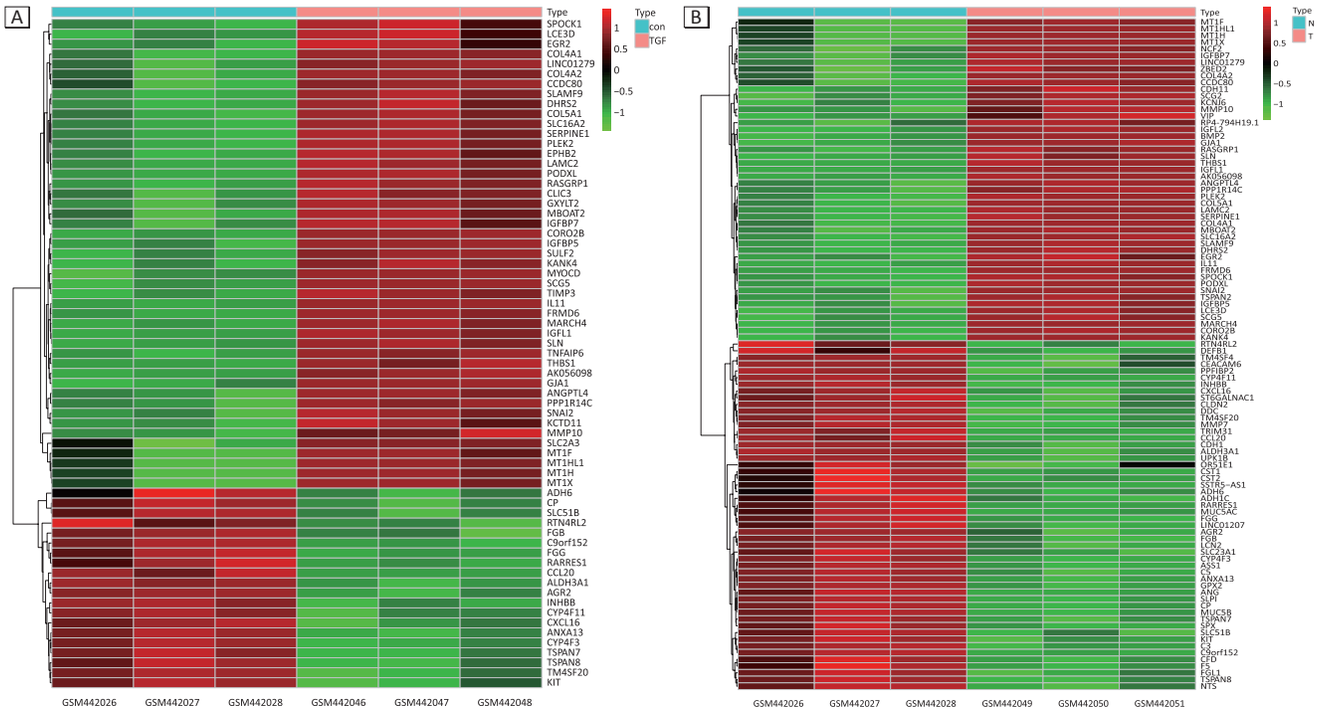
2 结果

2.1 差异表达基因的筛选

TGF-β刺激A549细胞24 h，共筛选出差异表达基因67个，其中上调47个，下调20个，热图如图1A所示。刺激72 h，共筛选出差异表达基因100个，其中上调48个，下调52个，热图见图1B。分别取24 h和72 h上调表达基因以及下调表达基因的交集，去除3个非编码基因等未在数据库中识别的基因名称，最终得到差异表达基因55个，其中上调35个，下调20个，见表2。

2.2 差异表达基因的GO富集分析

对上述得到的55个差异表达基因进行GO富集分析，结果显示：55个差异表达基因主要参与的生物学途径包括胶原蛋白分解代谢的过程、细胞迁移正调控、生长负调节、内皮细胞凋亡过程的负调控以及ECM组织；细胞组分的富集结果显示55个差异表达基因集中在胞外区和细胞外间隙；分子功能的富集显示差异表达基因主要具有ECM结构组成、肝素结合、磷脂酰丝氨酸绑定、酶抑制剂的活动、纤连蛋白结合功能。见图2。



[注] A：24 h；B：72 h。红色为上调基因，绿色为下调基因。标记基因为表达差异有统计学意义的基因。横坐标为基因芯片中样本的编号，纵坐标为基因名称。

[Note] A: 24 h; B: 72 h. Up-regulated genes are red, and down-regulated genes are green. The genes displayed are those with significant differences. The abscissa represents the number of samples in the gene chip, and the ordinate represents the gene symbol with significant differences.

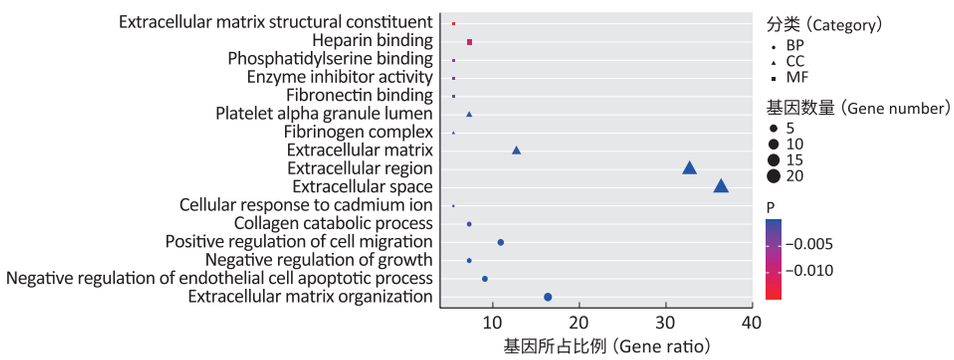
图1 差异表达基因热图

Figure 1 Heat map of differentially expressed genes

表2 共表达的差异表达基因

Table 2 Co-expressed differentially expressed genes

表达趋势 (Regulation)	基因名称 (Gene symbol)
上调 (Up-regulation) (*n*=35)	*IL11*、*CORO2B*、*FRMD6*、*IGFL1*、*SLN*、*SCG5*、*RASGRP1*、*IGFBP5*、*GJA1*、*THBS1*、*SLAMF9*、*LAMC2*、*SLC16A2*、*PODXL*、*ANGPTL4*、*KANK4*、*COL4A1*、*PPP1R14C*、*SERPINE1*、*PLEK2*、*DHRS2*、*COL5A1*、*SNAI2*、*MBOAT2*、*COL4A2*、*MMP10*、*IGFBP7*、*CCDC80*、*SPOCK1*、*LCE3D*、*MT1H*、*MT1X*、*MT1HL1*、*EGR2*、*MT1F*
下调 (Down-regulation) (*n*=20)	*ALDH3A1*、*AGR2*、*INHBB*、*C9orf152*、*CCL20*、*CYP4F3*、*ANXA13*、*TM4SF20*、*CYP4F11*、*TSPAN7*、*CXCL16*、*FGG*、*TSPAN8*、*CP*、*FGB*、*SLC51B*、*RARRES1*、*KIT*、*RTN4RL2*、*ADH6*



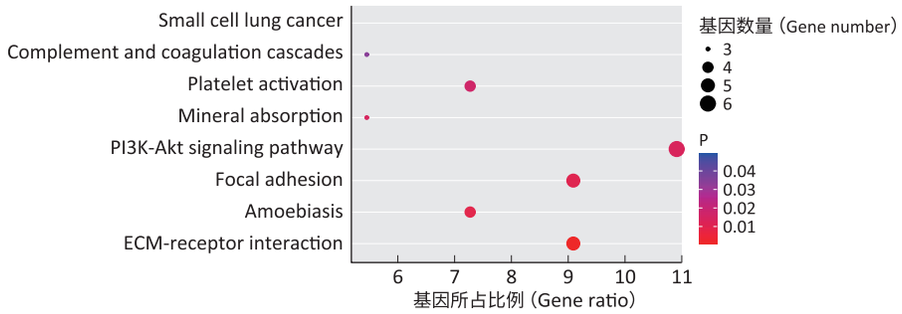
[注 (Note)] Extracellular matrix structural constituent 为细胞外基质结构组成，Heparin binding 为肝素结合，Phosphatidylserine binding 为磷脂酰丝氨酸绑定，Enzyme inhibitor activity 为酶抑制剂的活性，Fibronectin binding 为纤连蛋白结合，Platelet alpha granule lumen 为血小板-阿尔法颗粒管腔，Fibrinogen complex 为纤维蛋白原复杂，Extracellular matrix 为细胞外基质，Extracellular region 为胞外区，Extracellular space 为细胞外间隙，Cellular response to cadmium ion 为细胞对镉离子的反应，Collagen catabolic process 为胶原蛋白分解代谢的过程，Positive regulation of cell migration 细胞迁移正调控，Negative regulation of growth 为生长负调节，Negative regulation of endothelial cell apoptotic process 为内皮细胞凋亡过程的负调控，Extracellular matrix organization 为细胞外基质组织。BP：生物学途径 (Biology process)；CC：细胞组分 (Cellular component)；MF：分子功能 (Molecular function)。

图2 差异表达基因的GO功能注释

Figure 2 GO functional annotation of differentially expressed genes

2.3 差异表达基因的KEGG信号通路分析

对差异表达基因进行KEGG信号通路分析如图3所示。结果显示，55个差异表达基因主要参与的信号通路包括ECM受体相互作用、阿米巴类感染、黏着斑的形成、PI3K-Akt信号通路、矿物质吸收、血小板激活、补体及凝血级联反应、小细胞肺癌。



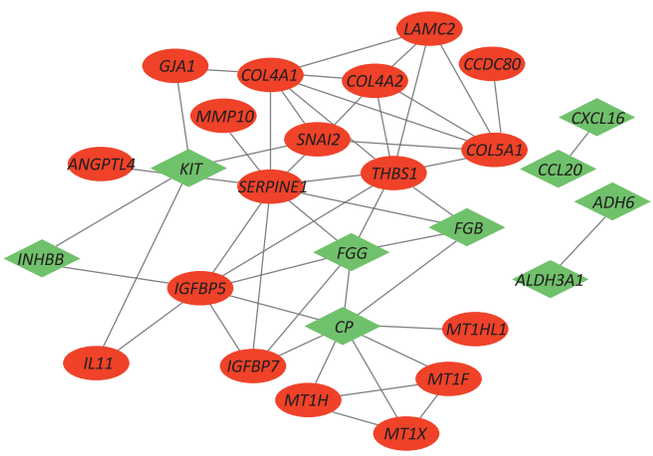
[注 (Note)] ECM-receptor interaction为细胞外基质受体相互作用，Amoebiasis为阿米巴类感染，Focal adhesion为黏着斑的形成，PI3K-Akt signaling pathway为PI3K-Akt信号通路，Mineral absorption为矿物质吸收，Platelet activation为血小板激活，Complement and coagulation cascades为补体及凝血级联反应，Small cell lung cancer为小细胞肺癌。

图3 差异表达基因的KEGG信号通路

Figure 3 KEGG signaling pathway of differentially expressed genes

2.4 差异表达基因的PPI网络构建及Hub基因的筛选

STRING数据库 (https://string-db.org) 用于构建差异表达基因的蛋白互作网络，除去涣散的链接以及孤立点，绘制PPI网络，如图4所示。网络中共包含27个差异表达基因，表达上调的差异表达基因为18个，表达下调的基因为9个。采用的CytoHubba插件中的MCC法筛选PPI网络中的Hub基因，其结果如图5所示。评分较高的基因分别为*THBS1*、*COL4A1*、*COL5A1*、*COL4A2*、*FGG*、*SERPINE1*、*LAMC2*、*IGFBP5*。

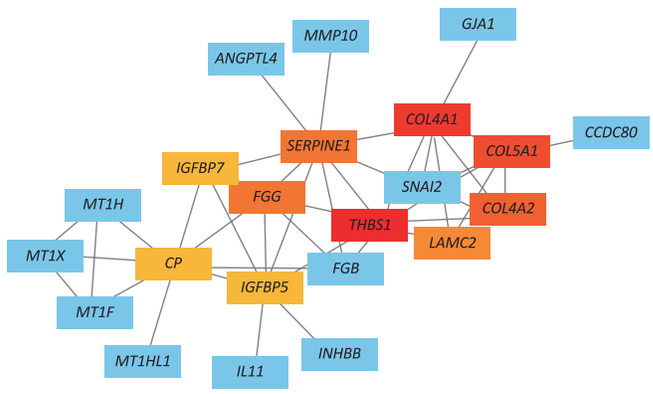


[注] 红色为上调的差异表达基因，绿色为下调的差异表达基因，连线代表基因之间存在相互作用。

[Note] Up-regulated genes are red, and down-regulated genes are green. The lines represent interactions between genes.

图4 差异表达基因的PPI网络

Figure 4 PPI network of differentially expressed genes



[注] 红色为MCC法计算出的得分较高，橙色次之，蓝色为得分较低的基因。

[Note] The score is calculated by the MCC method. Red represents a higher score, followed by orange, and blue represents a lower score.

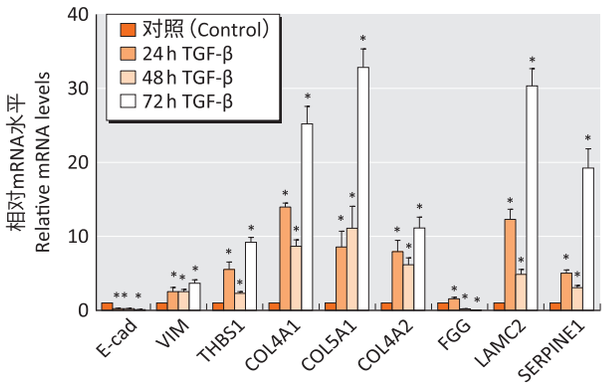
图5 Hub基因相关的PPI网络

Figure 5 PPI network of hub genes

2.5 qRT-PCR验证Hub基因和EMT相关转分化指标

首先，通过qRT-PCR测定了EMT相关转分化指标的表达。E-cad的表达在24、72 h均下调，VIM表达在24、72 h均上调 (如图6)。这说明TGF-β刺激A549细胞后发生了EMT的过程。随后，qRT-PCR结果显示*IGFBP5*在A549细胞中的丰度较低，并且与对照组相比，TGF-β刺激A549细胞24 h和72 h的差异并不明显。因此，对上述前7个Hub基因的表达情况进行了验证，除了*FGG*在24 h表达上调与芯片结果不符，其他基因的表达趋势均与芯片数据一致 (如图6)。除此之外，进一步检测TGF-β刺激A549细胞48 h时EMT相关转

分化指标的表达和上述7个Hub基因的表达情况（如图6）。



［注］*：与对照组相比，$P<0.05$。E-cad为E-钙黏蛋白，VIM为波形蛋白，*THBS1*为血小板反应蛋白-1基因，*COL4A1*为Ⅳ型胶原蛋白α1基因，*COL5A1*为Ⅴ型胶原蛋白α1基因，*COL4A2*为Ⅳ型胶原蛋白α2基因，*FGG*为人纤维蛋白原γ链抗体基因，*SERPINE1*为丝氨酸蛋白酶抑制蛋白E1基因，*LAMC2*为层粘连蛋白γ2基因，*IGFBP5*为胰岛素样生长因子结合蛋白5基因。

［Note］*: Compared with the control group, $P<0.05$. E-cad: E-cadherin, VIM: vimentin, *THBS1*: thrombospondin-1 gene, *COL4A1*: collagen alpha-1(IV) chain gene, *COL5A1*: collagen alpha-1(V) chain gene, *COL4A2*: collagen alpha-2(IV) chain gene, *FGG*: fibrinogen gamma chain gene, *SERPINE1*: serpin family E member 1 gene, *LAMC2*: laminin subunit gamma 2 gene, *IGFBP5*: insulin like growth factor binding protein 5.

图6 qRT-PCR验证EMT相关转分化基因和Hub基因

Figure 6 qRT-PCR verification of EMT-related transition genes and hub genes

3 讨论

肺纤维化是一种慢性、进行性和具有破坏性的肺间质性疾病。环境和职业、高龄、吸烟等因素都会影响肺纤维化的发生。肺纤维化预后差，治愈率低，发病原因并不明确。随着研究的不断深入，上皮间质转化被证实在肺纤维化的发生过程中具有重要的作用[18]。而TGF-β是导致EMT的主要诱导剂[19]。因此本研究选用TGF-β处理的肺泡Ⅱ型上皮细胞的基因芯片数据，以此来筛选肺纤维化相关EMT过程中具有差异表达的基因。

本研究共筛选和验证了7个差异表达基因——*THBS1*、*COL4A1*、*COL5A1*、*COL4A2*、*FGG*、*LAMC2*、*SERPINE1*。THBS1是血小板反应蛋白家族中的一员，它是一种多功能蛋白，具有不同的功能和结构域且具有多种生物活性。在组织生物发生和重塑中，THBS1调节细胞表型和细胞外结构，是被识别的第一构件[20]。已有文献报道THBS1是TGF-β的主要生理激活剂，它是EMT的有效激发剂[21-23]。研究表明，在结直肠癌中，THBS1通过增强癌细胞EMT促进它的迁移和侵袭[24]。而在另一项关于黑色素瘤研究中发现，THBS1能够激活TGF-β1来促进黑色素瘤的EMT[25]。由此可见，THBS1在EMT的过程中发挥着重要的作用，这一点在本研究中得到了相应的证实。

Ⅳ型胶原蛋白是基底膜中最丰富和必需的成分。基底膜是ECM的一种特殊形式，由Ⅳ型胶原蛋白、层粘连蛋白、纤连蛋白组成[26]。在某些器官的纤维化病变中可见Ⅳ型胶原蛋白异常沉积，如在特发性间质肺炎的早期纤维化过程中，Ⅳ型胶原蛋白被证实具有异常的沉积，并能够促进成纤维细胞迁移以及增加黏着斑激酶的磷酸化[27]。在EMT相关研究中，COL4A1被证实参与EMT过程。研究表明在膀胱尿路上皮癌中，COL4A1链激活细胞内蛋白激酶B（protein kinase B，PKB，也称AKT）信号传导途径从而触发EMT[28]。COL4A2链在EMT中的作用并未见相关报道，但在本研究中发现COL4A2链的上调趋势明显。随着生物信息学的发展，更多与EMT相关的Hub基因可能被筛选出来，而COL4A1链和COL4A2链可能会在EMT相关研究中的重要意义。Ⅴ型胶原蛋白是肺中的次要胶原，被隔离在Ⅰ型胶原（主要的肺胶原）的原纤维中[29]。它被认为是正常肺中的一种隔离抗原，位于血管周围和支气管周围的结缔组织中。先前的研究表明特发性肺纤维化期间肺中Ⅴ型胶原蛋白过表达，并且Ⅴ型胶原蛋白相关的免疫反应在肺纤维化的发病机理中起关键作用[30]。在一项关于闭塞性细支气管炎研究中表明，白细胞介素17（interleukin 17，IL17）能够让作为自身抗原的COL5A1表达上调，从而导致其异常沉积，促进EMT的发生[31]。而在另一项研究中发现在特发性肺纤维化患者的肺中，miR-185和miR-186除了能够调节COL5A1，还可以防止TGF-β诱导的Ⅴ型胶原蛋白表达和EMT[32]。因此Ⅳ型和Ⅴ型胶原蛋白都可能在肺纤维化相关的EMT中发挥重要作用。

纤维蛋白原是一种二聚体蛋白，由三对分别称为α、β和γ的多肽链组成，分别由纤维蛋白原α链（FGA）、纤维蛋白原β链（FGB）和纤维蛋白原γ链（FGG）三种不同的基因编码。FGG参与纤维蛋白原的形成和功能的调节。纤维蛋白原在血液凝固、纤维蛋白溶解、细胞和基质相互作用、炎症和伤口愈合中起重要作用[33]。在肝细胞癌中，FGG通过激活EMT增强肝细胞癌细胞的迁移和侵袭[13]。

层粘连蛋白是重要的ECM的组成成分之一，主要

在神经视网膜、骨骼或上皮细胞的ECM中积累。层粘连蛋白5是由α3（LAMA3）、β3（LAMB3）和γ2（LAMC2）三种亚基构成。在ECM中，层粘连蛋白5由上皮细胞特异性表达，并与细胞表面整联蛋白相互作用[34]。LAMC2仅存在于层粘连蛋白5中，它作为层粘连蛋白5分子中的独特链，在层粘连蛋白5的生理功能中起主要作用[35]。在胰腺癌的发生过程中，LAMC2被发现与胰腺癌细胞EMT有关[36]。在肺癌中，LAMC2通过诱导EMT促进肺腺癌细胞中的迁移、侵袭和转移[37]。在胆管癌中，LAMC2表达增加时，胆管癌组织中的EMT被激活[38]。由此可见LAMC2与EMT过程具有相关性。

SERPINE1也称为纤溶酶原激活物抑制剂1型，是参与ECM重组和细胞黏附的重要调控蛋白，它是丝氨酸蛋白酶抑制剂超家族的成员，是纤溶酶原激活系统的主要调节剂[39]。上皮性卵巢癌中，增加的SERPINE1表达能够诱导上EMT，而SERPINE1的低甲基化也参与了这一过程[40]。而在缺氧诱导的乳腺癌细胞EMT过程中SERPINE1表达增加[41]。同样，在胃癌中SERPINE1能够通过调节EMT促进肿瘤细胞的增殖、迁移和侵袭[42]。但其在肺纤维相关EMT中是否具有重要作用未见相关报道，可能由于目前肺纤维化相关的EMT研究还处在一个探索状态，需要借助测序技术等生物信息学的手段去发现更多起到关键作用的基因。此外，在A549细胞中的*IGFBP5*的丰度较低，并且TGF-β刺激A549细胞24 h和72 h后与对照组相比差异并不明显，这可能是*IGFBP5*的丰度较低对检测的准确度造成了影响。

综上，本研究从TGF-β刺激A549细胞的芯片数据中筛选并验证了出了7个Hub基因。*THBS1*、*COL4A1*、*COL5A1*、*COL4A2*、*LAMC2*、*SERPINE1*均具有差异表达并且趋势与芯片结果一致，*FGG*基因在72 h与芯片结果一致，但在TGF-β刺激A549细胞24 h时的趋势与芯片结果不一致。在24 h时，芯片中*FGG*基因的$\log_2$FC仅为–1.07，下调并不明显，这可能是实验结果与芯片趋势不符的重要原因。这些基因能够帮助发现更多在肺纤维化EMT过程中起关键作用的基因。然而，本研究并没有涉及深层次的机制探索，没有发掘Hub基因之间的潜在关联。但是，对于了解在EMT的过程中起到调控作用的基因具有参考意义。

参考文献

[1] MOORE B B，LAWSON W E，OURY T D，et al. Animal models of fibrotic lung disease [J]. Am J Respir Cell Mol Biol，2013，49（2）：167-179.

[2] FERNANDEZ I E，EICKELBERG O. New cellular and molecular mechanisms of lung injury and fibrosis in idiopathic pulmonary fibrosis [J]. Lancet，2012，380（9842）：680-688.

[3] GUO J，YANG Z，JIA Q，et al. Pirfenidone inhibits epithelial-mesenchymal transition and pulmonary fibrosis in the rat silicosis model [J]. Toxicol Lett，2019，300：59-66.

[4] 陈慧婷. 游离二氧化硅诱导大鼠肺泡II型上皮细胞转分化及其时间-效应关系 [D]. 郑州：郑州大学，2017.
CHEN H T. The identification of epithelial mesenchymal transdifferentiation induced by free silicon dioxide in type iialveolar epithelial cells and the time-effect [D]. Zhengzhou：Zhengzhou University，2017.

[5] NIETO M A，HUANG R Y，JACKSON R A，et al. EMT：2016 [J]. Cell，2016，166（1）：21-45.

[6] 姜岩，于功昌，张兴国. 转化生长因子-β1及其介导的信号通路在百草枯致肺上皮间质转化中的作用 [J]. 环境与职业医学，2017，34（8）：729-733.
JIANG Y，YU G C，ZHANG X G. Roles of transforming growth factor-beta 1 and its mediated signaling pathways in pulmonary epithelial mesenchymal transition induced by paraquat [J]. J Environ Occup Med，2017，34（8）：729-733.

[7] KASAI H，ALLEN J T，MASON R M，et al. TGF-β1 induces human alveolar epithelial to mesenchymal cell transition（EMT）[J]. Respir Res，2005，6：56.

[8] ZHANG C，ZHU X，HUA Y，et al. YY1 mediates TGF-β1-induced EMT and pro-fibrogenesis in alveolar epithelial cells [J]. Respir Res，2019，20（1）：249.

[9] SZKLARCZYK D，GABLE A L，LYON D，et al. STRING v11：protein-protein association networks with increased coverage，supporting functional discovery in genome-wide experimental datasets [J]. Nucleic Acids Res，2019，47（D1）：D607-D613.

[10] CHIN C H，CHEN S H，WU H H，et al. *CytoHubba*：identifying hub objects and sub-networks from complex interactome [J]. BMC Syst Biol，2014，8（Suppl 4）：S11.

[11] YANG L，ZHAO X，TANG X. Predicting disease-related proteins based on clique backbone in protein-protein interaction network [J]. Int J Biol Sci，2014，10（7）：677-688.

[12] WANG W，CHEN G，WANG B，et al. Long non-coding RNA

BZRAP1-AS1 silencing suppresses tumor angiogenesis in hepatocellular carcinoma by mediating THBS1 methylation [J]. J Transl Med, 2019, 17 (1): 421.

[13] ZHANG X, WANG F, HUANG Y, et al. FGG promotes migration and invasion in hepatocellular carcinoma cells through activating epithelial to mesenchymal transition [J]. Cancer Manag Res, 2019, 11: 1653-1665.

[14] JIN J, LIU H, JIN M, et al. Silencing of hsa_circ_0101145 reverses the epithelial-mesenchymal transition in hepatocellular carcinoma via regulation of the miR-548c-3p/LAMC2 axis [J]. Aging (Albany NY), 2020, 12 (12): 11623-11635.

[15] AKULA S M, RUVOLO P P, MCCUBREY J A. TP53/miR-34a-associated signaling targets SERPINE1 expression in human pancreatic cancer [J]. Aging (Albany NY), 2020, 12 (3): 2777-2797.

[16] TURNER A W, NIKPAY M, SILVA A, et al. Functional interaction between *COL4A1*/*COL4A2* and *SMAD3* risk loci for coronary artery disease [J]. Atherosclerosis, 2015, 242 (2): 543-552.

[17] LIU W, WEI H, GAO Z, et al. *COL5A1* may contribute the metastasis of lung adenocarcinoma [J]. Gene, 2018, 665: 57-66.

[18] YU X, ZHAI R, HUA B, et al. miR-let-7d attenuates EMT by targeting HMGA2 in silica-induced pulmonary fibrosis [J]. RSC Adv, 2019, 9 (34): 19355-19364.

[19] CAJA L, DITURI F, MANCARELLA S, et al. TGF-β and the tissue microenvironment: relevance in fibrosis and cancer [J]. Int J Mol Sci, 2018, 19 (5): 1294.

[20] ISENBERG J S, MARTIN-MANSO G, MAXHIMER J B, et al. Regulation of nitric oxide signalling by thrombospondin 1: implications for anti-angiogenic therapies [J]. Nat Rev Cancer, 2009, 9 (3): 182-194.

[21] BIGE N, SHWEKE N, BENHASSINE S, et al. Thrombospondin-1 plays a profibrotic and pro-inflammatory role during ureteric obstruction [J]. Kidney Int, 2012, 81 (12): 1226-1238.

[22] MASSAGUÉ J. TGFβ signalling in context [J]. Nat Rev Mol Cell Biol, 2012, 13 (10): 616-630.

[23] SCHULTZ-CHERRY S, RIBEIRO S, GENTRY L, et al. Thrombospondin binds and activates the small and large forms of latent transforming growth factor-*β* in a chemically defined system [J]. J Biol Chem, 1994, 269 (43): 26775-26782.

[24] LIU X, XU D, LIU Z, et al. THBS1 facilitates colorectal liver metastasis through enhancing epithelial-mesenchymal transition [J]. Clin Transl Oncol, 2020, 22 (10): 1730-1740.

[25] JAYACHANDRAN A, ANAKA M, PRITHVIRAJ P, et al. Thrombospondin 1 promotes an aggressive phenotype through epithelial-to-mesenchymal transition in human melanoma [J]. Oncotarget, 2014, 5 (14): 5782-5797.

[26] VAN DER REST M, GARRONE R. Collagen family of proteins [J]. FASEB J, 1991, 5 (13): 2814-2823.

[27] URUSHIYAMA H, TERASAKI Y, NAGASAKA S, et al. Role of α1 and α2 chains of type IV collagen in early fibrotic lesions of idiopathic interstitial pneumonias and migration of lung fibroblasts [J]. Lab Invest, 2015, 95 (8): 872-885.

[28] MIYAKE M, HORI S, MORIZAWA Y, et al. Collagen type IV alpha 1 (COL4A1) and collagen type XIII alpha 1 (COL13A1) produced in cancer cells promote tumor budding at the invasion front in human urothelial carcinoma of the bladder [J]. Oncotarget, 2017, 8 (22): 36099-36114.

[29] LINSENMAYER T F, GIBNEY E, IGOE F, et al. Type V collagen: molecular structure and fibrillar organization of the chicken alpha 1 (V) NH2-terminal domain, a putative regulator of corneal fibrillogenesis [J]. J Cell Biol, 1993, 121 (5): 1181-1189.

[30] VITTAL R, MICKLER E A, FISHER A J, et al. Type V collagen induced tolerance suppresses collagen deposition, TGF-β and associated transcripts in pulmonary fibrosis [J]. PLoS One, 2013, 8 (10): e76451.

[31] VITTAL R, FAN L, GREENSPAN D S, et al. IL-17 induces type V collagen overexpression and EMT via TGF-β-dependent pathways in obliterative bronchiolitis [J]. Am J Physiol Lung Cell Mol Physiol, 2013, 304 (6): L401-L414.

[32] LEI G S, KLINE H L, LEE C H, et al. Regulation of collagen V expression and epithelial-mesenchymal transition by miR-185 and miR-186 during Idiopathic pulmonary fibrosis [J]. Am J Pathol, 2016, 186 (9): 2310-2316.

[33] MOSESSON M W. Fibrinogen and fibrin structure and functions [J]. J Thromb Haemost, 2005, 3 (8): 1894-1904.

[34] COLOGNATO H, YURCHENCO P D. Form and function: the laminin family of heterotrimers [J]. Dev Dyn, 2000, 218 (2): 213-234.

[35] CARTER W G，RYAN M C，GAHR P J. Epiligrin，a new cell adhesion ligand for integrin α3β1 in epithelial basement membranes [J]. Cell，1991，65 (4)：599-610.

[36] WANG H，CAI J，Du S，et al. LAMC2 modulates the acidity of microenvironments to promote invasion and migration of pancreatic cancer cells via regulating AKT-dependent NHE1 activity [J]. Exp Cell Res，2020，391 (1)：111984.

[37] MOON Y W，RAO G，KIM J J，et al. LAMC2 enhances the metastatic potential of lung adenocarcinoma [J]. Cell Death Differ，2015，22 (8)：1341-1352.

[38] PEI Y F，LIU J，CHENG J，et al. Silencing of *LAMC2* reverses epithelial-mesenchymal transition and inhibits angiogenesis in cholangiocarcinoma *via* inactivation of the epidermal growth factor receptor signaling pathway [J]. Am J Pathol，2019，189 (8)：1637-1653.

[39] MCMAHON B，KWAAN H C. The plasminogen activator system and cancer [J]. Pathophysiol Haemost Thromb，2008，36 (3/4)：184-194.

[40] PAN J X，QU F，WANG F F，et al. Aberrant SERPINE1 DNA methylation is involved in carboplatin induced epithelial-mesenchymal transition in epithelial ovarian cancer [J]. Arch Gynecol Obstet，2017，296 (6)：1145-1152.

[41] AZIMI I，PETERSEN R M，THOMPSON E W，et al. Hypoxia-induced reactive oxygen species mediate N-cadherin and SERPINE1 expression，EGFR signalling and motility in MDA-MB-468 breast cancer cells [J]. Sci Rep，2017，7 (1)：15140.

[42] YANG J D，MA L，ZHU Z. SERPINE1 as a cancer-promoting gene in gastric adenocarcinoma：facilitates tumour cell proliferation，migration，and invasion by regulating EMT [J]. J Chemother，2019，31 (7/8)：408-418.

（英文编辑：汪源；责任编辑：王晓宇）

（上接第244页）

[7] FORSTHUBER M，KAISER A M，GRANITZER S，et al. Albumin is the major carrier protein for PFOS，PFOA，PFHxS，PFNA and PFDA in human plasma [J]. Environ Int，2020，137：105324.

[8] CHIESA L M，NOBILE M，PASQUALE E，et al. Detection of perfluoroalkyl acids and sulphonates in Italian eel samples by HPLC-HRMS Orbitrap [J]. Chemosphere，2018，193：358-364.

[9] 孙瑞. 全氟化合物在上海市典型环境介质和非暴露人群血清中赋存及污染特征研究 [D]. 上海：上海大学，2017.
SUN R. Preliminary studies on perfluorinated compounds in various typical environmental media and non-occupational human serum of shanghai [D]. Shanghai：Shanghai University，2017.

[10] KOPONEN J，RUDGE J，KUSHON S，et al. Novel volumetric adsorptive microsampling technique for determination of perfluorinated compounds in blood [J]. Anal Biochem，2018，545：49-53.

[11] SOLLIEC M，ROY-LACHAPELLE A，SAUVÉ S. Quantitative performance of liquid chromatography coupled to Q-Exactive high resolution mass spectrometry (HRMS) for the analysis of tetracyclines in a complex matrix [J]. Anal Chim Acta，2015，853：415-424.

[12] LIN Y，LIU R，HU F，et al. Simultaneous qualitative and quantitative analysis of fluoroalkyl sulfonates in riverine water by liquid chromatography coupled with Orbitrap high resolution mass spectrometry [J]. J Chromatogr A，2016，1435：66-74.

[13] WU M，SUN R，WANG M，et al. Analysis of perfluorinated compounds in human serum from the general population in Shanghai by liquid chromatography-tandem mass spectrometry (LC-MS/MS) [J]. Chemosphere，2017，168：100-105.

[14] KÄRRMAN A，VAN BAVEL B，JÄRNBERG U，et al. Development of a solid-phase Extraction-HPLC/Single quadrupole MS method for quantification of perfluorochemicals in whole blood [J]. Anal Chem，2005，77 (3)：864-870.

[15] KOPONEN J，RANTAKOKKO P，AIRAKSINEN R，et al. Determination of selected perfluorinated alkyl acids and persistent organic pollutants from a small volume human serum sample relevant for epidemiological studies [J]. J Chromatogr A，2013，1309：48-55.

（英文编辑：汪源；责任编辑：丁瑾瑜）