

• 论 著 •

基于 GEO 数据库的胎儿发育异常时羊水游离 RNA 中
组织特异性差异表达基因分析*

任建宇¹, 司艳梅¹, 刘欣², 杨树法^{1△}

1. 首都医科大学附属北京妇产医院产前诊断中心, 北京 100026; 2. 首都医科大学
基础医学院医学遗传学与发育生物学系, 北京 100069

摘要:**目的** 通过生物信息学分析发现羊水游离 RNA 中能够反映胎儿发育异常的组织特性基因。**方法** 从 Human Protein Atlas 数据库下载基因在正常组织表达数据, 从 Gene Expression Omnibus 数据库下载唐氏综合征及爱德华综合征胎儿羊水游离 RNA 芯片检测数据。利用 R 语言统计分析正常组织表达数据中的组织特异性基因。利用 limma 程序包分析唐氏综合征及爱德华综合征的差异表达基因。取两者交集获取组织特异性差异表达基因。**结果** 与正常胎儿比较, 唐氏综合征胎儿有 717 个差异基因, 爱德华综合征胎儿有 1 038 个差异基因, 71 个基因为共同差异基因。DOK7、ARHGEF39、FAM111B、CCHCR1、R3HDML、WNK3、FIBCD1、SMIM10L2B 及 SMIM10L2A 9 个基因为组织特异性差异表达基因, 这些基因参与大脑、甲状腺、睾丸、消化道等多个组织的发育过程。**结论** 差异表达基因分析和组织特异性基因相结合的方法是筛选羊水游离 RNA 中胎儿发育异常标志物的可行方法。

关键词: 羊水游离 RNA; 产前诊断; 差异表达基因; 组织特异性基因
DOI:10. 3969/j. issn. 1673-4130. 2021. 04. 009 **中图法分类号:**R722. 11
文章编号:1673-4130(2021)04-0426-05 **文献标志码:**A

Analysis of tissue-specific differentially expressed genes in aberrant
fetal amniotic fluid cell free RNA based on GEO database*

REN Jianyu¹, SI Yanmei¹, LIU Xin², YANG Shufa^{1△}

1. Prenatal Diagnostic Center, Beijing Obstetrics & Gynecology Hospital, Capital
Medical University, Beijing 100026, China; 2. Department of Genetics & Developmental
Biology, Capital Medical University, Beijing 100069, China

Abstract:**Objective** To discover tissue specific genes in correlation to fetal aberration in amniotic fluid cell free RNA via bioinformatics analysis. **Methods** Data of gene expression in normal tissues were downloaded from Human Protein Atlas database. Amniotic fluid free RNA array data of Down's syndrome, Edwards syndrome, and normal fetus were downloaded from Gene Expression Omnibus database respectively. R language was employed to analyze the tissue-specific genes in the normal tissue expression data. Limma package was used to analyze the differentially expressed genes of Down's syndrome and Edward's syndrome. The tissue-specific differentially expressed genes were obtained from the intersection of above two sets. **Results** Compared with normal fetus, 717 differential genes were screened from Down's syndrome, 1 038 differential genes were found in Edwards syndrome, 71 genes were common differential genes. DOK7, ARHGEF39, FAM111B, CCHCR1, R3HDML, WNK3, FIBCD1, SMIM10L2B and SMIM10L2A were tissue-specific differentially expressed genes, which involve multiple tissues development such as brain, thyroid, testis and digestive tract. **Conclusion** Combination of differential expression gene analysis and tissue-specific gene is a feasible method to screen the markers of fetal dysplasia in amniotic fluid free RNA.

Key words: amniotic fluid cell free RNA; prenatal diagnosis; differentially expressed genes; tissue specific genes

* 基金项目:北京市教育委员会科技计划项目(KM201910025027)。
作者简介:任建宇,男,技师,主要从事遗传病实验室诊断的相关研究。△ 通信作者, E-mail: sfyang@mail. ccmu. edu. cn。
本文引用格式:任建宇,司艳梅,刘欣,等. 基于 GEO 数据库的胎儿发育异常时羊水游离 RNA 中组织特异性差异表达基因分析[J]. 国际检验医学杂志, 2021, 42(4): 426-429.

出生缺陷是胎儿出生前发生的身体结构、功能或代谢异常,严重影响患儿和患儿家庭的生活质量,产前诊断是预防出生缺陷的主要手段。染色体微阵列分析技术、全基因组测序及全外显子测序等分子生物学技术在产前诊断中的应用,极大地提高了致病性变异检出率,同时也检出大量可疑致病性、不确定性、可疑良性的基因组结构性变异^[1-3]。如何获取更多胎儿发育信息,结合 B 超等影像学检查,综合判断胎儿发育情况,对胎儿做出合理的诊治,成为产前诊断中亟待解决的问题。羊水是产前诊断的主要标本,羊水上清液通常作为医疗废物处理。研究发现羊水上清液中包含各种游离 RNA,并且羊水游离 RNA 比羊水脱落细胞中 RNA 有代表性,更能反映胎儿全身的发育状态^[4-5]。胎儿发育异常能够影响羊水游离 RNA 中的基因表达,羊水游离 RNA 基因变化与器官组织发育异常密切相关^[6]。本研究通过对比分析发育异常胎儿和正常胎儿羊水游离 RNA 的芯片检测结果,获取差异表达基因,进而分析这些差异表达基因的组织表达特点,以期发现能够反映胎儿发育异常的组织特异性基因。

1 资料与方法

1.1 一般资料 以唐氏综合征(DS)及爱德华综合征(ES)胎儿作为系统发育异常的研究对象。从 GEO 数据库中下载 GSE16176^[7]和 GSE25634^[8]芯片检测数据。GSE16176 中包括 7 例正常胎儿和 7 例 DS 胎儿;GSE25634 包括 6 例正常胎儿和 5 例 ES 胎儿。所用的芯片为 Affymetrix Human Genome U133 Plus 2.0 Array。

1.2 方法

1.2.1 组织特异性基因的筛选 正常组织的基因表达量数据来源于 Human Protein Atlas 数据库^[9],下载基因表达数据表(https://www.proteinatlas.org/download/rna_tissue_consensus.tsv.zip)到本地。编写 R 脚本统计该数据表所涵盖的组织类型及所包括的基因种类,生成不同器官来源的组织列表。按照组织特异性基因的定义筛选神经相关组织的特异性基因。组织特异性基因定义:基因在某个组织中的表达量,高于该基因在所有组织中表达量均值的 8 倍以上,则该基因因为这个组织的特异性表达基因。

1.2.2 差异表达基因的筛选 对芯片结果进行背景校正、归一化,获取基因表达量数据。利用 limma 程序包对异常胎儿和正常胎儿的芯片检测结果进行差异表达分析,获取胎儿发育异常时发生变化的差异表达基因。以表达量变化倍数的 $\log_2$ 转化值($\log FC$) > 0.5 和 $P < 0.05$ 作为阈值筛选差异表达基因。

1.2.3 组织特异性差异表达基因的表达特性分析

取组织特异性基因集及差异性基因集得交集,获取组织特异性差异表达基因。统计组织特异性差异表达基因在正常胎儿羊水游离 RNA 中的变异系数(CV)值和均值,同时计算芯片检测基因在正常胎儿羊水游离 RNA 中的 CV 值和均值。

1.3 统计学处理 数据的分析和处理借助 R 语言完成。使用 *oligo* 程序包读取芯片原始结果,并对结果进行背景校正、归一化。差异基因的筛选使用 *limma* 程序包,通过 *t* 检验完成,根据设定阈值筛选差异有统计学意义的差异表达基因。CV 值为基因在正常胎儿表达量的标准差除以均值乘以 100% 得到。

2 结果

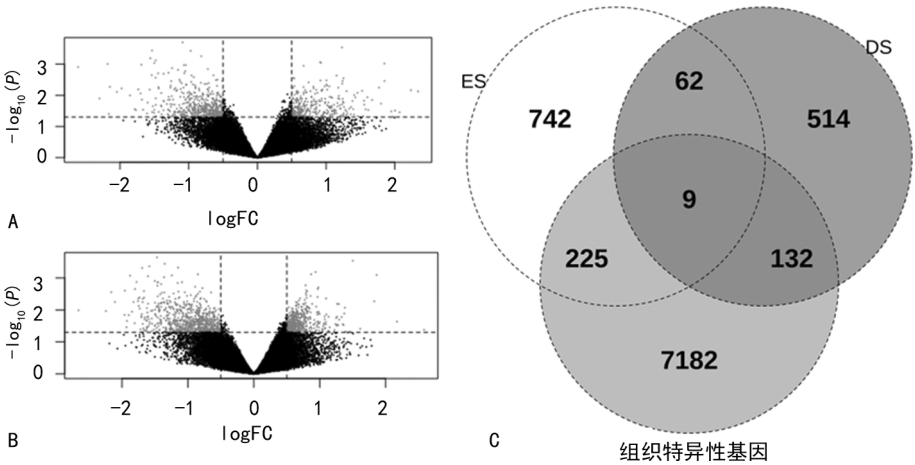
2.1 组织特异性基因的筛选 在组织表达量数据表中记录了 62 种组织的 19 651 个基因的表达量。62 种组织涵盖了神经系统、免疫和血液系统、生殖系统、腺体、泌尿系统及消化道等人体的几乎所有组织,见表 1。以基因表达量高于所有组织中表达量均值的 8 倍以上作为组织特异性基因的筛选标准,共筛选到 7 548 个基因。

2.2 DS 和 ES 差异表达的组织特异性基因 将 DS 与正常胎儿的羊水游离 RNA 中基因表达量进行比较,共发现 717 个差异表达基因,差异表达基因的火山图见图 1A。将 ES 与正常胎儿进行比较,共发现 1 038 个差异基因,差异基因的火山图见图 1B。将 DS 差异基因、ES 差异基因及组织特异性基因进行比较发现,71 个基因为 DS 与 ES 共同差异基因,9 个基因既是 DS 与 ES 共同差异基因又是组织特异性基因,见图 1C。

2.3 组织特异性差异表达基因 将正常胎儿的标本,进行背景校正和归一化,不同芯片的检测值的中位数以及上下四分位数值基本一致,所有芯片的结果具有可比性。统计分析 13 例正常胎儿标本的芯片检测结果,所有检测基因的相对表达量均值为 3.11, CV 为 19.90%。9 个组织特异性差异表达基因(DOK7、ARHGEF39、FAM111B、CCHCR1、R3HDML、WINK3、FIBCD1、SMIM10L2B 及 SMIM10L2A)中,5 个基因(CCHCR1、FAM111B、FIBCD1、ARHGEF39 及 SMIM10L2B)的表达量高于芯片所有检测基因的平均表达量并低于所有检测基因的 CV,3 个基因(DOK7、WINK3 及 R3HDML)的表达量低于芯片所有检测基因的平均表达量并低于所有检测基因的 CV,见图 2。在组织表达数据库中,9 个组织特异性差异表达基因特异对应的特异表达组织包括大脑皮质、海马、睾丸、小肠、十二指肠、肾上腺、心肌、骨骼肌、甲状腺、睾丸及附睾等。基因在这些组织的表达量,高于基因在所有组织中表达量均值的 8 倍以上。

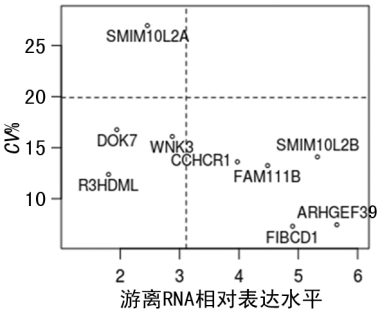
表 1 各系统中包含的组织

系统	组织名称
神经系统	大脑皮层、胼胝体、海马结构、下丘脑、杏仁核、中脑、嗅觉区、垂体、脑桥和延髓、脊髓、基底神经节、黑质、丘脑、小脑
免疫和血液系统	骨髓、树突状细胞、淋巴结、NK 细胞、单核细胞、脾脏、T 细胞、胸腺、B 细胞、血液中多页核细胞、扁桃体、粒细胞
生殖系统	输精管、附睾、精囊、睾丸、子宫颈、子宫内膜、输卵管、卵巢、阴道
腺体	肾上腺、胰腺、甲状旁腺、前列腺、唾液腺、甲状腺
泌尿系统	肾、膀胱
消化道	胃、结肠、十二指肠、食道、阑尾、直肠、小肠
其他	肺、胆囊、肝脏、视网膜、骨骼肌、皮肤、平滑肌、舌头、乳房、心肌、脂肪组织、胎盘



注：A、B 分别表示 DS 胎儿(A)、ES 胎儿(B)与正常胎儿羊水游离 RNA 中差异表达基因的火山图，纵轴代表 limma 程序包基因差异分析 P 值经 $-\log_{10}$ 转化得到的值，灰色点代表 $\log FC > 0.5$ 和 $P < 0.05$ 的差异表达基因；C 为 DS 差异基因、ES 差异基因及组织特异性基因比较的维恩图，Specific gene 代表组织特异性基因集合，DS 代表 DS 胎儿与正常胎儿差异表达基因集，ES 代表 ES 胎儿与正常胎儿差异表达基因集。

图 1 DS 和 ES 差异表达的组织特异性基因



注：垂直虚线为基因游离 RNA 水平在正常胎儿羊水标本中的平均值；横向虚线为所有检测变异系数。

图 2 组织特异性差异表达基因在正常胎儿标本中的表达

3 讨 论

羊水游离 RNA 的变化受到胎儿状态的影响，孕期内胎儿是一个变化的过程；这种变化包括正常的胎儿发育和异常发育^[5-6,10-11]。区分是胎儿正常生长还是病理性发育导致的羊水游离 RNA 中基因的变化，是其临床应用的根本。孕中期是行羊水穿刺检测的主要阶段，也是能够获取羊水上清液的主要阶段。分析这一阶段羊水游离 RNA 中基因变化特点，将是利用羊水游离 RNA 进行产前诊断的基础。利用每个基

因在孕中期的 CV 可以对该基因在孕中期的表达稳定性进行有效的区分。通过对筛选到的 9 个组织特异性差异表达基因的 CV 分析发现，8 个基因的 CV 均低于基因整体的 CV 的中位数，结果表明，这些基因在孕中期具有较好的稳定性，较少受到了胎儿正常发育的影响。

羊水游离 RNA 来源于胎儿细胞的凋亡或死亡，凋亡细胞释放的 RNA 可以通过胎儿血液经由肾脏排泄入羊水。皮肤也是 RNA 进入羊水的重要途径，研究表明，在孕 20 周前胎儿皮肤尚未角质化，胎儿和羊水间可以通过皮肤进行物质双向交换^[12]。作为能够反映胎儿发育状态的基因，其在羊水游离 RNA 中应具备一定的表达量，许多研究通过制订筛选阈值来筛选羊水游离 RNA 中“表达的”基因^[13]，另外，还有利用芯片进行自身判定的方法^[14]。本研究将筛选到的组织特异性差异表达基因与芯片所有基因的均值进行比较，结果表明，5 个基因的表达量在平均表达量之上。通过对基因在孕中期 CV 和平均表达量进行分析，CCHCR1、FAM111B、FIBCD1、ARHGEF39 及 SMIM10L2B 的游离 RNA 在羊水中具有可靠的水

平,同时在孕中期表达稳定,具备作为特异性组织差异基因的条件。

筛选到的组织性特异性差异表达基因涉及大脑、睾丸、甲状腺、心肌、骨骼肌及消化道等多个器官组织,这些组织在 ES 和 DS 患儿中存在不同程度的发育异常^[15]。CCHCR1 在睾丸高表达,定位于细胞质、细胞核、线粒体或中心体等,可以调节细胞的各种功能,是精子生成和雄性配子发育过程中的重要分子^[16-17],结果表明,CCHCR1 在 DS 和 ES 患儿中的表达水平变化与患儿生殖功能障碍密切相关。DOK7 的突变与先天性肌无力综合征密切相关^[18];肌张力异常是 ES 和 DS 患儿共同症状,这也解释了 DOK7 游离 RNA 在羊水水中的异常表达。表明研究中筛选到的组织特异性差异表达基因的功能与 DS 和 ES 的临床症状存在较好的相关性。

基因的特异表达类型与目前研究不一致的情况,例如,FIBCD1 在 Human Protein Atlas 数据库中是海马组织特异性基因,而目前关于神经系统中 FIBCD1 的研究较少,现有文献表明 FIBCD1 可以位于胃肠道上皮细胞,参与皮肤及胃肠黏膜的固有免疫,与肝癌的预后密切相关^[19-20]。ARHGEF39 在 Human Protein Atlas 数据库中是甲状腺的组织特异性基因,但目前对该基因的研究集中在肺癌与肝癌领域^[21-22]。出现这种现象的可能原因与胚胎发育中的基因表达与成年人存在差异有关,也可能与基因研究历史短有关。由于发育中的人类胚胎组织难以获取,缺少胎儿发育过程中的组织表达数据集,研究中多采用成人的组织表达数据集替代^[23]。

本研究利用基因在组织中的表达量对变异基因进行解释,发现了与组织异常发育密切相关的组织特异性基因,为利用羊水游离 RNA 判定胎儿发育异常提供了思路。

参考文献

- [1] WANG J, CHEN L, ZHOU C, et al. Prospective chromosome analysis of 3 429 amniocentesis samples in China using copy number variation sequencing[J]. Am J Obstet Gynecol, 2018, 219(3): 287. e1-287. e18.
- [2] YANG S, LV J, SI Y, et al. Diagnostic differences between patients opting for non-invasive prenatal testing and patients having traditional prenatal diagnosis[J]. Int J Clin Exp Pathol, 2018, 11(5): 2831-2838.
- [3] RIGGS E R, ANDERSEN E F, CHERRY A M, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource

- (ClinGen)[J]. Genet Med, 2020, 22(2): 245-257.
- [4] KANG J H, PARK H J, JUNG Y W, et al. Comparative transcriptome analysis of cell-free fetal RNA from amniotic fluid and RNA from amniocytes in uncomplicated pregnancies[J]. PLoS One, 2015, 10(7): e0132955.
- [5] TARCA A L, ROMERO R, PIQUE-REGI R, et al. Amniotic fluid cell-free transcriptome: a glimpse into fetal development and placental cellular dynamics during normal pregnancy[J]. BMC Med Genomics, 2020, 13(1): 25.
- [6] ZWEMER L M, BIANCHI D W. The amniotic fluid transcriptome as a guide to understanding fetal disease[J]. Cold Spring Harb Perspect Med, 2015, 5(4): a023101.
- [7] SLONIM D K, KOIDE K, JOHNSON K L, et al. Functional genomic analysis of amniotic fluid cell-free mRNA suggests that oxidative stress is significant in Down syndrome fetuses[J]. Proc Natl Acad Sci U S A, 2009, 106(23): 9425-9429.
- [8] KOIDE K, SLONIM D K, JOHNSON K L, et al. Transcriptomic analysis of cell-free fetal RNA suggests a specific molecular phenotype in trisomy 18[J]. Hum Genet, 2011, 129(3): 295-305.
- [9] THUL P J, AKESSON L, WIKING M, et al. A subcellular map of the human proteome[J]. Science, 2017, 356(6340): eaal3321.
- [10] YONG W J, JUNG I S, SO H S, et al. Global gene expression analysis of cell-free RNA in amniotic fluid from women destined to develop preeclampsia [J]. Medicine (Baltimore), 2019, 98(3): e13971.
- [11] CHO H Y, CHO Y, SHIN Y J, et al. Functional analysis of cell-free RNA using mid-trimester amniotic fluid supernatant in pregnancy with the fetal growth restriction [J]. Medicine (Baltimore), 2018, 97(2): e9572.
- [12] LILLIAN M Z, HUI L S, HEATHER C W, et al. RNA-Seq and expression microarray highlight different aspects of the fetal amniotic fluid transcriptome[J]. Prenat Diagn, 2014, 34(10): 1006-1014.
- [13] JUNG Y W, SHIM S S, PARK J E, et al. Analysis of the cell-free amniotic fluid transcriptome expressed during the euploid mid-trimester of pregnancy[J]. Eur J Obstet Gynecol Reprod Biol, 2016, 203(1): 94-98.
- [14] HUI L, SLONIM D K, WICK H C, et al. The amniotic fluid transcriptome: a source of novel information about human fetal development[J]. Obstet Gynecol, 2012, 119(1): 111-118.
- [15] 陆国辉, 徐湘民. 临床遗传咨询[M]. 北京: 北京大学医学出版社, 2007.
- [16] HUANG M T, MENG Z, JIANG T, et al. Fine mapping the MHC region identified rs4997052 as a new variant associated with nonobstructive azoospermia in Han Chinese males[J]. Fertil Steril, 2019, 111(1): 61-68. (下转第 433 页)

照组,可能与稽留流产患者胎盘系统应激反应及自噬的激活有关^[10-11]。单因素分析和多因素 Logistic 回归分析显示,人工流产次数、自然流产次数、不孕时间、PSG1 mRNA、BMP-2、Beclin-1 mRNA 水平与稽留流产具有一定的相关性。有研究发现,首次流产患者子宫蜕膜组织 PSG1 mRNA、BMP-2、Beclin-1 mRNA 水平与再次妊娠时子宫蜕膜组织上述指标的水平具有一定的相关性^[12-14],这意味着若首次妊娠时上述指标的水平异常则下次妊娠时这些指标水平异常风险会增加,这对于想再次妊娠的患者具有一定的意义。

综上所述,人工流产次数、自然流产次数、不孕时间、PSG1 mRNA、BMP-2、Beclin-1 mRNA 水平与稽留流产具有一定的相关性,可为稽留流产的诊断提供参考。

参考文献

[1] 龙取云,汤欣欣,顾莹,等.稽留流产细胞遗传学病因分析及生化指标检测的临床意义[J].中国计划生育学杂志,2019,27(7):902-904.

[2] 吴小青,李英,安刚,等.88例稽留流产组织的 SNP-array 检测结果分析[J].中华医学遗传学杂志,2018,35(6):923-925.

[3] 刘丹,丁翔,唐国栋,等.200例稽留流产绒毛染色体的间期 FISH 分析[J].生殖医学杂志,2019,28(5):548-552.

[4] 张羽,张钗红,唐丹,等.稽留流产高危因素的相关研究进展[J].临床误诊误治,2017,30(12):113-116.

[5] 范小玲,刘姿.孕早期稽留流产的病因分析及清宫术后用药评价[J].湖南师范大学学报(医学版),2017,14(1):72-74.

(上接第 429 页)

[17] LING Y H, WONG C C, LI K W, et al. CCHCR1 interacts with EDC4, suggesting its localization in P-bodies [J]. Exp Cell Res, 2014, 327(1): 12-23.

[18] LORENZONI P J, KAY C S K, AMDT R C, et al. Congenital myasthenic syndrome due to DOK7 mutation in a cohort of patients with 'unexplained' limb-girdle muscular weakness[J]. J Clin Neurosci, 2020, 75: 195-198.

[19] VON HUTH S, MOELLER J B, SCHLOSSER A, et al. Immunohistochemical localization of fibrinogen C domain containing 1 on epithelial and mucosal surfaces in human tissues[J]. J Histochem Cytochem, 2018, 66(2): 85-97.

[20] YAN W, SUN M J, LIU J B, et al. FIBCD1 overexpression predicts poor prognosis in patients with hepatocellular carcinoma[J]. Oncol Lett, 2019, 19(1): 795-804.

[6] 罗树玲,郑春艳,曹锦慧.肝素结合表皮生长因子的高表达与稽留流产病因的相关性研究[J].中国性科学,2016,25(9):101-104.

[7] BESPALOVA O, BAKLEICHEVA M, KOVALEVA I, et al. Expression of vitamin D and vitamin D receptor in chorionic villous in missed abortion[J]. Gynecol Endocrinol, 2019, 35(sup1): S49-S55.

[8] ROONEY J. Abortion in Northern Ireland: a missed opportunity to consider article 3 [J]. J Soc Welf Fam Law, 2019, 41(2): 225-228.

[9] JIN Z, YAO X Y, TANG Y H, et al. Analysis of clinicopathologic classification features of hydatidiform mole misdiagnosed as missed abortion preoperatively[J]. Rep Dev Med, 2019, 3(2): 97.

[10] 薛慧琴,周岩,卢洪涌,等. NOG 基因与不明原因自然流产相关性的初步探讨[J].中国优生与遗传杂志,2018,26(2):71-73.

[11] 王鹤, REDDY R, NGUYEN N M P, et al. 骨形态发生蛋白 15 基因多态性与复发性葡萄胎和习惯性流产的相关性研究[J].中国妇幼保健,2017,32(11):2432-2435.

[12] ALI M K, BOTROS H A, MOSTAFA S A. Foley's catheter balloon for induction of mid-trimester missed abortion with or without traction applied: a randomized controlled trial[J]. J Matern Fetal Neonatal Med, 2020, 33(2): 198-205.

[13] 谢冰,陈霞,于骏,等. Beclin-1 在稽留流产中的表达模式及临床意义[J].医学研究生学报,2019,32(1):78-81.

[14] WOOD S L, BRAIN P H. Medical management of missed abortion: a randomized clinical trial[J]. Obstet Gynecol, 2002, 99(4): 563-566.

(收稿日期:2020-04-02 修回日期:2020-09-05)

[21] ZHOU H J, LIN C, ZHANG X P, et al. ARHGEF39 promotes tumor progression via activation of Rac1/P38 MAPK/ATF2 signaling and predicts poor prognosis in non-small cell lung cancer patients[J]. Lab Invest, 2018, 98(5): 670-681.

[22] XU B J, WU L, LI X Y, et al. Prognostic genes of hepatocellular carcinoma based on gene coexpression network analysis[J]. J Cell Biochem, 2019, 120(7): 11616-11623.

[23] NEETA L V, HUI L S. Next-generation sequencing and prenatal 'omics: advanced diagnostics and new insights into human development[J]. Genet Med, 2018, 20(8): 791-799.

(收稿日期:2020-06-12 修回日期:2020-12-18)