

转录因子 THAP1 在人肺泡 II 型上皮细胞中结合靶基因的谱系分析

丁 蓉, 胡毓华*, 何思思, 陆 超

(南京医科大学第一附属医院儿科, 江苏 南京 210036)

[摘要] 目的: 确定转录因子 THAP1 的靶基因, 为新生儿肺损伤的治疗寻找新策略。方法: 利用 THAP1 抗体进行染色质免疫共沉淀实验(chromatin immuno-precipitation, ChIP), 将富集到肺泡 II 型上皮细胞来源的肺腺癌细胞 A549 上的 DNA 样品进行高通量测序, 对得到的核苷酸序列进行生物信息学分析。结果: 共得到 20 417 个 THAP1 结合位点的核苷酸序列片段, 对应 9 688 个结合的靶基因, THAP1 结合的主要区域分布在基因间区、内含子区和启动子区。其中包含 1 051 个基因的启动子区。将 THAP1 结合于启动子区的基因进行基因功能(gene ontology, GO)注释, 表明 THAP1 参与包括蛋白质二聚化、同源蛋白质结合、蛋白质磷酸化调节、类固醇激素受体的激活等多种细胞生物学过程。THAP1 结合的基因主要涉及 9 个代谢路径, 以胞吞作用、多糖降解、硫酸软骨素或硫酸皮肤素的生物合成等路径为主。结论: 初步确定了转录因子 THAP1 在肺泡 II 型上皮细胞中的靶基因结合序列, 提示 THAP1 广泛参与了多种生物学过程, 为研究 THAP1 的功能及作用机制研究提供了线索, 从而为新生儿肺损伤的治疗奠定了基础。

[关键词] 高通量测序; A549 细胞; THAP1; 染色质免疫共沉淀实验

[中图分类号] R393

[文献标志码] A

[文章编号] 1007-4368(2015)09-1200-05

doi: 10.7655/NYDXBNS20150903

Profiling analysis of target genes binded by transcription factor THAP1 in human alveolar type II epithelial cells

Ding Rong, Hu Yuhua*, He Sisi, Lu Chao

(Department of Pediatrics, the First Affiliated Hospital of NJMU, Nanjing 210036, China)

[Abstract] **Objective:** To determine target genes of transcription factors THAP1 for a new therapeutic strategy of neonatal lung injury. **Methods:** Chromatin DNA in alveolar type II epithelial cells was immunoprecipitated with THAP1 antibody. The obtained DNA samples were then subjected to high-throughput sequencing and bioinformatic analysis. **Results:** A total 20 417 enrichment peaks corresponding 9 688 genes was found. The mapping of the peaks revealed that the targeted nucleotide sequences located mainly within intergenic regions, intron regions and promoter regions, which contained 1 051 gene promoter regions. Gene Ontology (GO) analysis suggested that THAP1 regulated a set of genes, which were involved in various biological processes including protein dimerization activity, identical protein binding, protein phosphatase regulator activity, steroid hormone receptor activity. Signal pathway analysis showed that these genes were associated with endocytosis, glycan degradation, and biosynthesis-chondroitin sulfate/dermatan sulfate. **Conclusion:** The targeted genes of the transcription factor THAP1 in the alveolar type II epithelial cells were determined. These results show that THAP1 is widely involved in various biological processes, which might provide a useful insight to reveal the function and mechanism of THAP1 gene on neonatal lung injury.

[Key words] high-throughput sequencing; A549 cells; THAP1; ChIP

[Acta Univ Med Nanjing, 2015, 35(09):1200-1204, 1215]

THAP 蛋白是细胞因子家族, 其最明显的特征是蛋白质的 N 端有进化上保守的 THAP 结构域, 其

内部含有新型的锌指模序, 因为能特异性地识别 DNA、RNA, 所以这一类细胞因子可广泛参与细胞增殖、细胞周期、细胞凋亡、转录的调控, 以及染色体修饰和重构等细胞生命活动。

[基金项目] 国家自然科学基金(81170487); 江苏省妇幼系统重点学科项目(FXK 201212)

*通信作者(Corresponding author), E-mail: wqm403-19h@163.com

新生儿肺损伤是由多种肺内与肺外原因引起

的弥漫性肺部损害,主要会出现顽固性低氧血症、进行性呼吸困难、肺顺应性严重降低等表现。其损伤机制涉及炎性水肿、组织异常修复和细胞凋亡、血管生成、细胞外基质重建等多种因素。THAP1 是 THAP 家族中第一个被发现的成员,作为一种凋亡因子在调控新生儿肺损伤方面是否有作用,作用机制如何目前尚不清楚,为了寻找可能的机制,本课题组进行了染色质免疫共沉淀(chromatin immuno-precipitation, ChIP)^[1]与高通量测序(high-throughput sequencing)^[2]结合的(ChIP-Seq)实验,这是鉴定转录因子及其调控靶基因转录的有效方法^[3-5],为揭示 THAP1 在新生儿肺损伤发生过程中的调控作用及机制提供了线索。

1 材料和方法

1.1 材料

胎牛血清、DMEM 培养基(Gibco 公司,美国),肺泡 II 型上皮细胞来源的肺腺癌 A549 细胞(本院儿科实验室保存),甲醛、甘氨酸、蛋白酶抑制剂(上海康成生物工程有限公司),THAP1 抗体(Sant Cruz 公司,美国)。

1.2 方法

1.2.1 A549 细胞的培养

从液氮罐中取出细胞冻存管,迅速置于预温的 37℃ 水浴中,迅速晃动直至细胞冻存液彻底融解(1 min 内完成),然后将细胞悬液转移至 15 mL 无菌离心管内,加入 5 mL DMEM 培养液,轻轻吹吸,使之混匀,1 500 r/min 离心 5 min,弃上清,再加入 1 mL 含 10% 胎牛血清的 DMEM 细胞培养液,再次吹吸混匀,最后转移至 10 cm 培养皿,并加入适量的含 10% 胎牛血清的培养液,置于 37℃、5%CO₂ 恒温恒湿的细胞培养箱中培养。为保证结果可靠性,共培养了 3 组细胞进行重复试验。

1.2.2 细胞交联

将 10 cm 培养皿中旧培养基吸弃,添加 10 mL 新鲜 DMEM 培养基,加入 270 μL 37% 的甲醛,轻轻混匀,室温孵育 10 min,加入 1 mL 1.25 mol 的甘氨酸轻轻混匀,室温孵育 5 min,中和甲醛,将培养皿中的混合液吸掉,用预冷的 1×PBS 洗细胞 2 遍,加入 1 mL 含有 1×protease inhibitor cocktail 的 1×PBS,用细胞刮将培养皿底的细胞刮下,收集到 1.5 mL 离心管中,2 000 g 离心 5 min,弃上清,收集到的样品进行下一步试验。

1.2.3 染色质免疫共沉淀与高通量测序(ChIP-Seq)

细胞交联后固定了细胞内所有与 DNA 结合的蛋

白的活动,然后裂解细胞,使染色体分离,用超声将染色质随机切割,再将蛋白质-DNA 复合物与目的蛋白抗体孵育,将与之相结合的蛋白-DNA 复合物洗脱下来,再通过反交联释放结合蛋白的 DNA 片段,纯化 DNA 后进行末端修复、加 polyA 并连接测序接头,然后用琼脂糖凝胶电泳进行片段大小选择,然后进行聚类合成,建好的测序文库用 Illumina HiSeq™2000 进行高通量测序。测序由上海康成生物工程有限公司完成。

用比对软件 BOWTIE software 将测序数据与人类基因组序列 hg19 进行比对,获得唯一比对的读长(reads)。序列比对之后产生的文件可以上传到 UCSC 基因组浏览器上实现数据的可视化^[6]。再用 MACS v1.4.0 (Model-based Analysis of ChIP-Seq) 软件进行比较^[7],样品比对到基因组上不同区域的读序密度,并找出差异性的比对区域,这些区域被称为峰(peak),从而获得 THAP1 富集峰,统计 THAP1 富集峰在全基因组的分布。根据 Gene ontology(<http://www.geneontology.org>)得到基因功能(gene ontology, GO)注释,基于最新的 KEGG 数据库(http://www.genome.jp/kegg/tool/map_pathway2.html)。得到不同 mRNAs 表达的路径分析(图 1)。

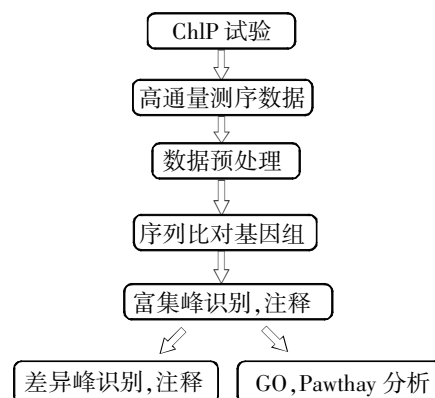


图 1 信息分析流程

Figure 1 Flow diagram of information analysis

2 结果

2.1 测序产量与峰信息统计结果

测序以 100 个碱基为 1 个 read 长度,测序结果显示 THAP1 获得的读长总数为 17 106 163,产量为 17 106 163 bp,将得到的总产量与全基因组染色体进行比对,比对结果显示:THAP1 唯一比率为 6.71% (表 1)。扣除背景和噪音后,得到峰信息统计结果:THAP1 共得到了 20 417 个峰数目,代表了 9 688 个 THAP1 结合的基因,峰总长度为 6 791 200 bp,峰长度主要分布在 200~600 bp 之间,峰平均长度为 333 bp,

表 1 测序统计与序列比对统计结果

Table 1 Clean data statistics results and sequence alignment results

样品	读长	读长总数	产量(bp)	唯一比对数量	唯一比对率(%)
THAP1	100	17 106 163	1 710 616 300	1 148 543	6.71
Input	100	21 398 892	2 139 889 200	2 651 395	12.40

峰总长度占总产量的 0.397%。

2.2 THAP1 在染色体上结合位点分析结果

分析 THAP1 结合位点在染色体上的分布情况,发现 THAP1 结合位点富集在 1、2、3、5、9、10、11、12、17 号染色体上较多,其他染色体上相对较少,Y 染色体上最少(图 2)。

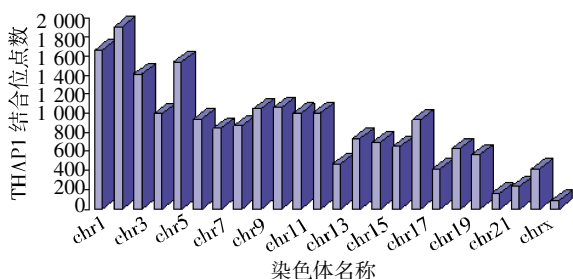


图 2 每条染色体上含有 THAP1 结合点的数量

Figure 2 The number of the THAP1 combining site on each chromosome

2.3 THAP1 结合位点分布区域分析结果

以基因转录起始位点 (transcription start site, TSS) 为参考,距离 TSS -2~+2 kb 的为候选启动子区域,距离 TSS -20~-2 kb 的为上游区域,距离 TSS +2 kb 处的为内含子或外显子区域,除外上述 4 个区域以外区域为基因间区(图 3)。THAP1 主要分布在基因间区(46.57%)和内含子区(33.12%),而外显子区分布最少(仅 1.67%),在启动子区分布比例为 5.20%(图 4)。

2.4 THAP1 结合位点与转录起始位点距离的分布

在调控基因表达的作用上,转录因子与启动子区结合,在 RNA 聚合酶 II 作用下形成转录起始复合物,还需要多种蛋白因子共同协助,一起参与转录过程。随转录起始表达的进行,转录因子与 RNA 聚合酶 II 形成的复合物从转录起始位点开始沿转录方向向下游移动,所以理论上说,转录起始位点附近的转录因子与 DNA 结合的数量较多,随着与转录起始位点距离增大而逐渐减少。

在这里,计算了数据中结合位点到最近的转录起始位点之间的距离,并统计了转录起始位点下游 20 000 bp 范围内结合位点到转录起始位点距离的频数分布情况,从图 5 中可以看出转录起始位点附近的 THAP1 与 DNA 结合的分布符合上述理论。

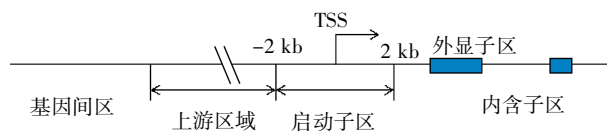


图 3 峰的 5 种分布区域

Figure 3 Five types of peaks

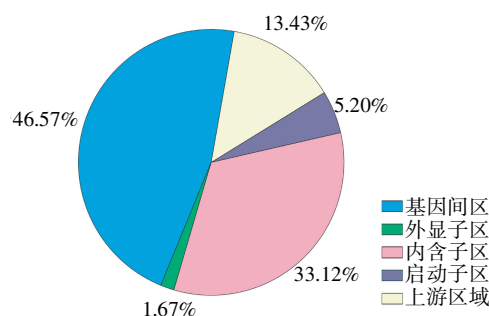


图 4 在 A549 细胞中峰的分布

Figure 4 Peaks distribution in A549

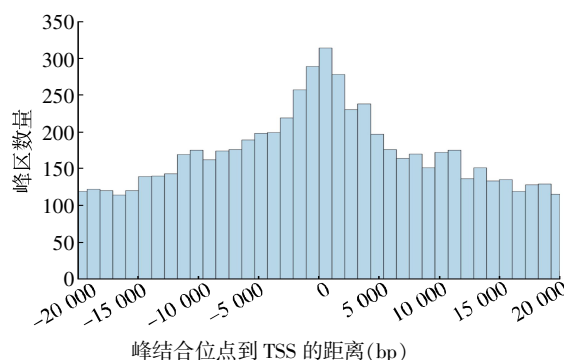


图 5 在 A549 细胞转录起始位点 ± 20 kb 内峰的分布

Figure 5 Distribution of peaks within TSS ± 20 kb in A549

2.5 THAP1 结合于启动子区的基因进行 GO 注释的分析结果

为揭示 THAP1 结合基因的功能,将 THAP1 结合于启动子区的基因进行 GO 注释,如图 6 结果所示,THAP1 结合的靶基因参与了多种生物学过程,包括蛋白质二聚化、同源蛋白质结合、蛋白质磷酸化调节、类固醇激素受体的激活、序列特异性 DNA 绑定 RNA 聚合酶 II 转录因子的活化、二磷酸尿核苷-糖基转移酶活化等多种生物学功能。

2.6 THAP1 调控的基因参与的主要代谢路径

THAP1 调控的基因参与主要代谢路径涉及 9 个方面(图 7),与细胞内吞、多糖降解、黏多糖-硫酸软骨素生物合成、药物代谢等途径有关。其中还涉

及丝裂原活化蛋白激酶 (mitogen-activated protein kinase, MAPK) 通路, MAPK 是丝氨酸/苏氨酸蛋白激酶, 该信号通路主要参与细胞增殖、分化、转化、凋亡等过程, 而 TGF- β 1 可活化该通路, 有很强的致纤维化效应。

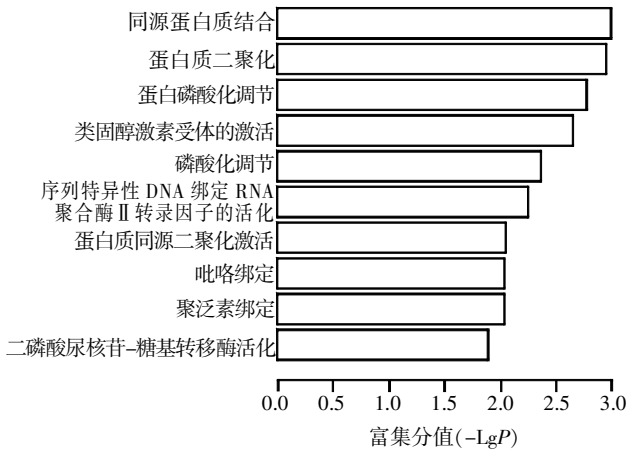


图 6 GO 注释 ChIP 富集基因的分子功能

Figure 6 GO terms of ChIP-enriched molecular function

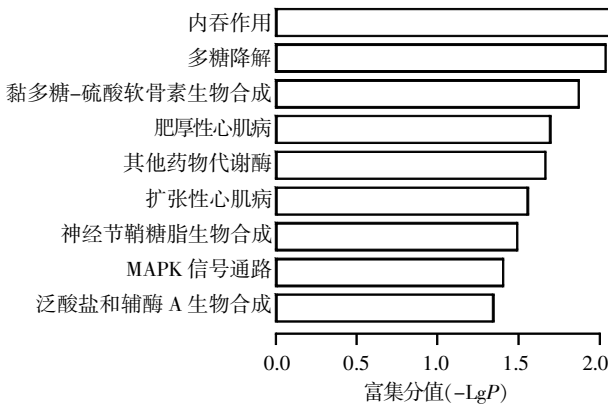


图 7 基因代谢途径

Figure 7 Genetic metabolic pathways

2.7 THAP1 结合于启动子区的部分基因

把 THAP1 结合于启动子区的基因按富集倍数排序, 发现了下列基因, 这些基因可能涉及听觉神经信息传递、肿瘤抑制、减数分裂起始的调控、甲状腺肿瘤等(表 2)。

在这些基因中还发现了一些基因与液体转运、抗菌抗炎、抗蛋白酶活性, 细胞增殖、分化、迁移、凋亡, 血管发生和免疫功能调节有关, 极有可能与肺损伤的发生相关, 为将来的研究提供了方向。

3 讨论

新生儿肺损伤发病机制尚不明确, 多与炎症反应或炎症因子的“瀑布式”级联反应^[8]、血管生成、细胞外基质重建、细胞凋亡、组织异常修复等多种因

素有关。目前尚缺乏判断肺损伤及其程度的特异性生物标志物。有研究表明脓毒症休克导致急性肺损伤时, 血清中 SP-A 与 SP-D 的浓度值可作为急性肺损伤的早期诊断指标^[9]。在俞敏等^[10]研究中发现 LPS 诱导的肺损伤与 RAGE 及 sRAGE 的水平密切相关。另一个回顾性病例对照研究显示在 192 例患者中, 可以用 RAGE 水平升高来区分急性肺损伤与非急性肺损伤的创伤疾病^[11]。还有人提出血管生成素-1(Ang-1)和 Ang-2 作为血管生长因子可作为急性肺损伤的生物标志物。但是由于肺损伤涉及多个结构肺单位与多因素, 目前发现的肺损伤生物标志物特异性不强, 为此有必要对影响肺损伤的相关基因进行研究, 探索肺损伤的发病机制, 为早期预防与诊断提供依据。

THAP 蛋白家族是具有 C2CH 手指状结构域的转录因子, 共有 12 个成员, THAP1 是该家族中第一个被发现的, 是核转录因子, 可以调节内皮细胞增殖^[12]。为了解它在新生儿肺损伤方面是否有调控作用, 于是利用 ChIP-Seq 技术对 THAP1 的结合位点进行了鉴定, 以寻找相关的靶基因。

本实验利用 ChIP-Seq 技术寻找转录因子的结合位点, 共获得 20 417 个峰数目, 代表了 9 688 个 THAP1 结合的基因。发现 THAP1 主要分布在基因间区(46.57%), 而外显子区分布最少(仅约 1.67%)。在启动子区分布比例大约为 5.20%。为揭示 THAP1 结合基因的功能, 将 THAP1 结合于启动子的基因进行 GO 注释, 发现 THAP1 参与了多种生物学过程。

在结合于启动子区的基因中, 发现 THAP1 与 IL24、水通道蛋白 AQP12B、AQP3、基质金属蛋白酶 MMP8、CABIN1、血管生成素样蛋白 3(ANGPTL3)、白细胞蛋白酶抑制因子(SLP1)、转化生长因子(TGF β 3)、TGF β 1、BUB1 等基因相结合。已有研究指出 IL-24 具有刺激免疫细胞分泌细胞因子的免疫调节功能。Caudell 等^[13]也发现 IL-24 可刺激人外周血单核细胞产生高水平的 IL-6、TNF- α 、IFN- β 以及少量的 IL-1、IL-10, 可抑制 Th1 型细胞因子的诱导作用。水通道蛋白(AQPs)除参与肺组织中液体的经上皮转运外, 还可能参与上皮细胞及内皮细胞的迁移和增殖。基质金属蛋白酶(MMPs)可以影响内皮细胞增殖、迁移以及促进血管生成因子或血管生成抑制因子的表达。在 Craig 等^[14]的研究中发现, 博来霉素诱导的急性肺损伤中, MMP8 有抗炎但促进纤维化的作用。TGF- β 超家族成员也广泛调控多种生物活性, 人类多种疾病都与 TGF- β 信号紊乱有关, 例

表 2 与 THAP1 结合于启动子区的基因

Table 2 The gene in the promoter region combined with THAP1

基因名	染色体	峰起始位置	峰终止位置	长度(bp)	富集倍数	峰中心到 TSS 的距离(bp)
AKAP5	chr14	64 933 056	64 933 487	432	118.83	1 055
MLIP	chr6	53 883 155	53 883 547	393	118.83	-362
EMX2OS	chr10	119 303 742	119 304 132	391	101.85	642
SCARNA15	chr15	83 423 129	83 423 376	248	101.85	-1 444
MIR4719	chr16	76 900 905	76 901 159	255	101.85	-1 800
TPO	chr2	1 417 016	1 417 403	388	101.85	-23
AC104134.2	chr2	88 837 808	88 838 204	397	101.85	-231
SOWAHC	chr2	110 370 789	110 371 153	365	101.85	-939
WASF1	chr6	110 499 499	110 499 796	298	101.85	1 560
FLJ41200	chr9	13 430 896	13 431 149	254	101.85	306
TYSND1	chr10	71 905 119	71 905 367	249	098.04	1 253
CHCHD5	chr2	113 340 518	113 340 892	375	098.04	-1 330
ANGPTL3	chr1	63 064 386	63 064 716	331	067.90	1 394
MMP8	chr11	102 596 584	102 596 788	205	067.90	-1 001
TGFβ3	chr14	76 448 711	76 448 916	206	067.90	-721
TGFβR1	chr9	101 866 816	101 867 159	344	049.02	-424
AQP12B	chr2	241 623 327	241 623 792	466	024.51	-1 242
BUB1	chr2	111 434 567	111 434 854	288	067.90	974
SLPI	chr20	43 882 588	43 882 824	237	049.02	499
IL24	chr1	207 070 067	207 070 426	360	019.61	-541

如癌症和纤维症。BUB1 是有丝分裂检测点的重要蛋白之一，是一个对着丝粒张力有反应的蛋白激酶，不但监测有丝分裂过程中微管与着丝点的结合，还对 DNA 损伤应答、肿瘤发生等方面起到重要的调控作用^[15]。上述这些基因可通过参与炎性水肿、血管生成、细胞外基质重建、细胞凋亡、组织异常修复等多种途径与肺损伤的发生密切相关，那么本实验发现 THAP1 能够富集结合于这些基因的启动子区，初步推测 THAP1 可能通过调控这些基因的表达而发挥功能，那么 THAP1 在新生儿肺损伤中的具体调控机制是怎样的需要进一步整合数据，并进行实验验证。

综上，本研究利用 ChIP-Seq 对 THAP1 在全基因组上的结合位点进行了鉴定，共获得 20 417 个峰数目，代表了 9 688 个 THAP1 结合的基因。这些基因广泛参与了多种生物学过程，显示了 THAP1 调控基因的广泛性及多功能性，还发现了部分基因对肺组织中液体的转运、抗菌抗炎、抗蛋白酶活性，细胞增殖、分化、迁移、凋亡，血管发生和免疫功能起着调节作用，这为进一步研究 THAP1 对新生儿肺损伤的调控机制奠定了重要基础。

[参考文献]

[1] Komata M, Katou Y, Tanaka H, et al. Chromatin immunoprecipitation protocol for mammalian cells[J]. *Methods Mol Biol*, 2014, 1164: 33-38

[2] Shendure J, Ji H. Next-generation DNA sequencing [J]. *Nat Biotechnol*, 2008, 26(10): 1135-1145

[3] Park PJ. ChIP-Seq: advantages and challenges of a maturing technology[J]. *Nat Rev Genet*, 2009, 10(10): 669-680

[4] Johnson DS, Mortazavi A, Myers RM, et al. Genome-wide mapping of *in vivo* protein-DNA interactions[J]. *Science*, 2007, 316(5830): 1497-1502

[5] Mardis ER. ChIP-seq: welcome to the new frontier[J]. *Nat Methods*, 2007, 4(8): 613-614

[6] Karolchik D, Barber GP, Casper J, et al. The UCSC Genome Browser database: 2014 update[J]. *Nucleic Acids Res*, 2014, 42: 764-770

[7] Feng J, Liu T, Qin B, et al. Identifying ChIP-seq enrichment using MACS[J]. *Nat Protoc*, 2012, 7(9): 1728-1740

[8] Tang JR, Michaelis KA, Nozik-Grayck E, et al. The NFκB inhibitory proteins IκBα and IκBβ mediate disparate responses to inflammation in fetal pulmonary endothelial cells[J]. *J Immunol*, 2013, 190(6): 2913-2923

[9] 宋 杨, 张 铮, 秦海东, 等. 多巴胺和去甲肾上腺素对脓毒症休克兔急性肺损伤的保护作用[J]. *南京医科大学学报: 自然科学版*, 2014, 34(5): 617-620

[10] 俞 敏, 赵 赛, 程怀平, 等. RAGE 与腹腔注射脂多糖诱导新生大鼠急性肺损伤的相关性研究[J]. *南京医科大学学报: 自然科学版*, 2014, 34(8): 1051-1055