

赵为民, 方晓敏, 涂 枫, 等. 猪单核源性巨噬细胞受 FSL-1 刺激后 lncRNAs 的鉴定与特征分析[J]. 江苏农业学报, 2019, 35(2): 346-356.

doi: 10.3969/j.issn.1000-4440.2019.02.015

猪单核源性巨噬细胞受 FSL-1 刺激后 lncRNAs 的鉴定与特征分析

赵为民^{1,2}, 方晓敏^{1,2}, 涂 枫^{1,2}, 周李生^{1,2}, 李碧侠^{1,2}, 王学敏^{1,2}, 付言峰^{1,2}, 任守文^{1,2}

(1.江苏省农业科学院畜牧研究所/江苏省农业种质资源保护与利用平台, 江苏 南京 210014; 2.农业部种养殖结合重点实验室, 江苏 南京 210014)

摘要: lncRNA 是新近发现的一种具有重要功能的 RNA 分子, 在各种生理活动中发挥着重要作用。为鉴定猪单核源性巨噬细胞受 FSL-1 刺激后相关的 lncRNA, 利用二代测序技术结合生物信息学对 lncRNA 进行了组装与特征分析。结果显示 FSL-1 刺激成功地构建了细胞炎症模型, 试验组与对照组共鉴定到 1 056 个 lncRNA 转录本, 其对应于 831 个基因座, 这些 lncRNA 的平均长度为 1 643 bp, 平均外显子数为 2.4 个。相对于对照组, 试验组有 51 个 lncRNA 上调表达, 44 个 lncRNA 下调表达。在上调 lncRNA 两侧相邻的蛋白质编码基因参与免疫反应、炎症反应、病毒反应等通路, 而在下调 lncRNA 两侧相邻的蛋白质编码基因参与的通路较少, 不参与上述通路。定量 PCR 结果显示挑选的 4 个 lncRNA 的差异倍数与 RNA-Seq 的结果相似, 其中上调的 lncRNA 呈现一定的组织特异表达, 而下调的 lncRNA 呈现多组织表达。这些结果为进一步研究 lncRNA 在 FSL-1 介导的炎症反应中的作用奠定了基础。

关键词: 猪; lncRNA; 单核源性巨噬细胞; FSL-1; 炎症

中图分类号: S828 **文献标识码:** A **文章编号:** 1000-4440(2019)02-0346-11

Identification and characterization of lncRNA in porcine monocyte-derived macrophage stimulated by FSL-1

ZHAO Wei-min^{1,2}, FANG Xiao-min^{1,2}, TU Feng^{1,2}, ZHOU Li-sheng^{1,2}, LI Bi-xia^{1,2}, WANG Xue-min^{1,2}, FU Yan-feng^{1,2}, REN Shou-wen^{1,2}

(1. Institute of Animal Science, Jiangsu Academy of Agricultural Sciences/Jiangsu Germplasm Resources Protection and Utilization Platform, Nanjing 210014, China; 2. Key Laboratory of Crop and Livestock Integration, Ministry of Agriculture, Nanjing 210014, China)

Abstract: lncRNA is a newly discovered RNA molecule with important functions and plays an important role in various physiological activities. To identify lncRNA in porcine monocyte-derived macrophage stimulated by FSL-1, lncRNA was

收稿日期: 2018-06-25

基金项目: 国家自然科学基金项目(31601928); 国家生猪现代产业技术体系南京综合试验站(CARS-35); 江苏省农业科技自主创新基金项目[CX(17)2002]

作者简介: 赵为民(1983-), 男, 湖北钟祥人, 博士, 副研究员, 研究方向为猪的抗病育种。(Tel) 025-84391941; (E-mail) zhao_weimin1983@aliyun.com

通讯作者: 任守文, (Tel) 025-84390349; (E-mail) shouwenren@163.com

assembled and characterized using next generation sequencing technology combined with bioinformatics. The results showed that the model of cell inflammation was successfully constructed with FSL-1 stimulation. A total of 1 056 lncRNA transcripts were identified in the experimental and control groups, which corresponded to 831 gene loci. The average length of these lncRNA was 1 643 bp, and these lncRNA span 2.4 exon on average. Compared to the control group, 51 and 44 lncRNAs were up-regulated and

down-regulated in the experimental group, respectively. Protein encoding genes adjacent to both sides of up-regulated lncRNA participated in immune response, inflammatory response, and viral response pathways, while protein encoding genes adjacent to both sides of down-regulated lncRNAs participated in fewer pathways and did not participate in the above pathway. Quantitative PCR results showed that the differential folds of the four selected lncRNAs were similar to those of RNA-Seq, in which the up-regulated lncRNAs exhibited tissue-specific expression, while the down-regulated lncRNAs exhibited multi-tissue expression. These results lay the foundation for further study of the role of lncRNA in FSL-1 mediated inflammatory response.

Key words: pig; lncRNA; monocyte-derived macrophage; FSL-1; inflammation

长链非编码 RNA (Long non-coding RNA, lncRNA) 是一类不具有编码蛋白质能力而本身长度超过 200 bp 的 RNA 分子。基于其在基因组上相对于蛋白质编码基因的位置主要可分为 3 种: 基因间长链非编码 RNA (Long intervening non-coding RNA, lincRNA), 自然反义链 (Natural antisense transcript, NAT) 和内含子长链非编码 RNA (Intronic long non-coding RNA)。lncRNA 由于其较低的物种间保守性在最初大量发现时被认为是转录噪音, 不具有生物学功能^[1]。随着近年对 lncRNA 研究的不断深入, lncRNA 受到越来越多的关注, 其功能从最初参与的 X 染色失活^[2]到后来的细胞亚小体生成^[3]、细胞重编程^[4]、胚胎发育^[5]、干细胞多能性^[6]、肌与脂肪生成^[7-8]、肿瘤癌症^[9-10]等各种细胞生命活动中。

天然免疫系统是宿主抵御病原体入侵的第一道防线, 其介导宿主的早期炎症反应对机体在抵抗病原体感染过程中发挥着关键作用^[11]。然而炎症反应中炎症因子的紊乱表达会影响宿主对病原体感染的抵抗, 研究表明, 肺部炎症因子的紊乱表达与宿主对支原体肺炎的易感性密切相关^[12-13]。炎症因子的释放由宿主 PRRs 与 PAMP 的识别以及细胞内复杂的信号转导等一系列事件构成^[11], 越来越多的研究表明炎症因子的释放不仅受到一些经典的相关蛋白质基因和 miRNA 调控^[11, 14], lncRNA 在其中也发挥着重要的调控作用^[15-16], 因而探索 lncRNA 在炎症反应中的作用对深入了解炎症因子的精确调控具有重要意义。

巨噬细胞作为宿主与外界病原体之间天然免疫反应的关键防线, 其与病原体之间的互作情况对宿主的感染进程有重要的决定作用^[17-20]。来源于外周血单个核细胞 (Peripheral blood mononuclear cell, PMBC) 在体外培养时通过添加粒细胞-巨噬细胞集落刺激因子 (Granulocyte-macrophage colony-stimulating factor, GM-CSF) 可形成单核源性巨噬细胞 (Monocyte-derived macrophage, MDM)^[21]。相对于屠宰猪

获取的巨噬细胞, MDM 的获取简单方便, 因而可对不同品种、不同年龄阶段猪种进行持续稳定的研究。

本研究通过 RNA-Seq 结合生物信息学分析对猪单核源性巨噬细胞受支原体脂蛋白 FSL-1 刺激后的 lncRNA 进行鉴定与特征分析, 为进一步研究 lncRNA 在 FSL-1 介导的炎症反应中的作用奠定基础。

1 材料和方法

1.1 试剂

RNA 提取试剂盒 Direct-zol™ RNA Kits 购于 Zymo research 公司, DNA marker、cDNA 反转录试剂盒、TB Green™ Premix Ex Taq™ II 定量试剂盒购于 TaKaRa 公司, DMEM 培养基、0.25% 胰酶、1640 培养基、青链霉素、胎牛血清购于 Thermo fisher 公司, 猪外周血单个核细胞分离液 KIT 购于天津 TBD 公司, GM-CSF 购于 RD system 公司, FSL-1 购于 InvivoGen 公司。

1.2 猪单核巨噬细胞的培养及 FSL-1 刺激

猪来源于江苏省农业科学院六合动物科学基地, 苏山猪种, 采用猪外周血单个核细胞分离液 KIT 分离 3 头健康猪外周血单核细胞, 全培养基 (10% FBS + 90% RPMI 1640) 中添加终质量浓度 20 ng/ml GM-CSF 培养 7 d, 每 2.5 d 换 1 次液, 得到猪单核源性巨噬细胞^[21-22], 设置对照组与试验组, 每组 3 个生物学重复。试验组加入终质量浓度 100 ng/ml 的 FSL-1, 对照组加入相应的 PBS, 刺激 6 h 后收集细胞提取 RNA。

1.3 RNA 测序的文库构建

RNA 提取参照试剂盒说明书操作, 并进行 DNaseI 处理, 纯化及回收。DNaseI 处理效果用 PCR 检测, 引物见表 1, Agilent 2100 分析 RNA 的完整性, 当 RNA 完整性相关参数 $RIN \geq 8$, $28S/18S \geq 1.5$ 时可进行测序文库的构建。RNA 测序文库依据 NEBNext Ultra™ RNA Library Prep Kit for Illumina 的说明操作, 简单如下: 提取总 RNA 先进行 poly+A

的纯化,然后进行 cDNA 第一链及第二链的合成,再对 cDNA 产物进行末端修复,3'末端加 A 和加 A-dapter,最后进行连接产物的纯化回收,上样测序,产出原始数据 Raw reads。

1.4 转录本的构建及 lncRNA 的鉴定流程

Raw reads 经过去除 Adapter、poly-N 以及低质量的 Reads,得到 Clean reads。然后通过 HISAT2 将 Clean reads 比对猪基因组 Sscrofa11.1,基于比对结果用 StringTie 组装转录本,从 NCBI 和 Ensemble 上分别下载猪的蛋白质编码基因的位置信息(染色体及其起始终止位置),版本分别为 NCBI Annotation Release:106(猪的蛋白质编码基因以 NM_和 XM_开头)和 Ensemble release 91(Protein_coding)。采用以下步骤来鉴定 lncRNA:(1)过滤小于 200 bp 长度以及表达量极低的转录本($FPKM < 0.01$);(2)去除和 NCBI 以及 Ensemble 中猪蛋白质编码基因位置重叠的转录本;(3)去除 Coding potential calculator (CPC)^[23] tool>0 的转录本;(4)去除 NCBI blastX 中有对应蛋白质数据库 UniProtKB/Swiss-Prot (swissprot)的转录本($E\text{-value} < 10^{-5}$);(5)将剩余的转录本翻译成相应氨基酸序列,使用隐马尔可夫模型(Hidden markov models, HMMs)对 Pfam 蛋白质库进行比对,去除在 Pfam 中有对应的转录本;最终剩余的转录本为 lncRNA。

1.5 差异 lncRNA 的保守性与重复序列分析

用 HTseq 对组装的转录本进行表达水平的定量分析,然后使用 DEseq2 对组间的 lncRNA 进行表达差异

分析。将变化倍数(Fold change) ≥ 2 以及错误发现率 $FDR < 0.05$ 作为差异表达的筛选标准。通过 Ensemble 数据库下载人和小鼠的 lncRNA 序列,利用 BLASTN 比对程序,分析差异 lncRNA 与人和小鼠 lncRNA 之间的保守性。利用 RepeatMasker 网站对差异 lncRNA 序列中可能存在的重复序列元件类别进行在线分析。在 Search engine 中选择 Cross_match,在 DNA source 中选择物种 Pig,其它选择参数为默认。

1.6 差异 lncRNA 两侧蛋白质编码基因的 GO 与 KEGG 分析

以上述 NCBI Annotation Release:106 和 Ensemble Release 92 的蛋白质编码基因为数据库,选择差异 lncRNA 两侧相邻的蛋白质编码基因,使用 DAVID (Database for annotation, visualization and integrated discovery) 对这些蛋白质编码基因进行 GO (Gene ontology) 和 KEGG pathway (Kyoto encyclopedia of genes and genomes) 分析,研究这些蛋白质编码基因参与的生物学功能与途径。筛选条件选择 $Count = 2$, $EASE = 0.05$, $Fold\ enrichment \geq 1.5$ 。

1.7 差异 lncRNA 定量 PCR 验证

对照组与试验组 RNA 反转录成 cDNA,分别挑选 2 个上调和下调的 lncRNA 进行 qPCR 验证,以 HPRT 作为内参基因,在 ABI 7500 仪器上进行定量 PCR 反应,对照组与试验组各 3 个生物学重复,每个样品技术重复 3 次。荧光定量 PCR 按照试剂盒说明书操作,定量数据用 $2^{-\Delta\Delta Ct}$ 法处理,差异比较采用独立样本 t 检验。引物序列见表 1。

表 1 引物序列信息

Table 1 The sequence information of primer

基因	引物序列(5'→3')	扩增长度 (bp)	退火温度 (°C)
<i>IL1B</i>	F: ACCCAAAACCTTGACCTC R: GACGGATGAAAAGATGCT	180	61
<i>LncRNA-MSTRG.8350</i>	F: CCATCATGGCCAGGAAAGGA R: TCCCCAAGCGGATAGTCAGA	215	60
<i>LncRNA-MSTRG.8244</i>	F: GTTGCTCGTTTAACGCTGCA R: AGTGGATCTCCTAGCCAGCA	154	60
<i>LncRNA-MSTRG.12242</i>	F: GCTGGGGATGAGGATTGGAG R: CTGCGTGTGGTTCCCTACAT	214	60
<i>LncRNA-MSTRG.3552</i>	F: CCGAGCTGACAGGTGAAAGA R: CTTTCCCGGGGATGAACAC	243	60
<i>HPRT</i>	F: CCCAGCGTCCTGATTAGTGA R: TTGAGCACACAGAGGGCTAC	191	60

2 结果与分析

2.1 RNA 质量及其测序文库数据分析

经过 Agilent 2100 bioanalyzer 检测后发现 6 个样本 RNA 的 RIN 为 9.0~9.4, $28S/18S$ 为 2.1~2.5, 符合 RNA 测序文库的构建要求。通过在 $IL1B$ 外显子上设计不跨内含子的扩增引物来确认 RNA 中是否有 DNA 污染。我们对 RNA DNase I 处理前后的结果进行了验证,直接以 DNase I 处理前后的 RNA

为模板进行 PCR 扩增,同时以猪基因组和 Pegfp-N1 作为阳性与阴性对照。图 1A 和图 1B 显示在 DNase I 处理前,对照组与试验组的 RNA 里均能扩增出 $IL1B$ 产物, DNase I 处理后均扩增不出 $IL1B$ 产物,说明 DNase I 处理能消除 RNA 样品中的 DNA。RNA 测序方面,6 个样本中经过 RNA-Seq 产出的 Clean data 在 6.52~7.35 Gb, GC 含量都在 53% 左右,其 Q20 百分比都在 95% 以上, Q30 百分比都在 90% 以上。

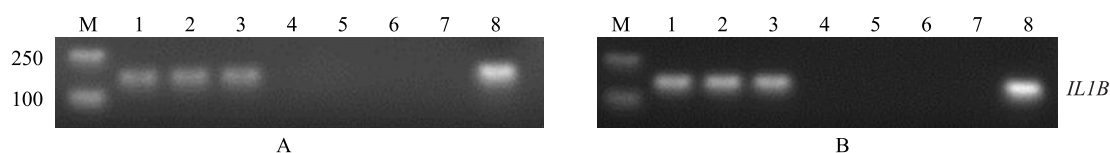


图 1 RNA 的 DNase I 处理前后比较
A:对照组;B:试验组。M:DL2000; 1~3:1~3 RNA 样品在 DNase I 处理前 $IL1B$ 的 PCR 扩增;4~6:1~3 RNA 样品在 DNase I 处理后 $IL1B$ 的 PCR 扩增;7:阴性对照;8:猪基因组。

图 1 RNA 的 DNase I 处理前后比较

Fig.1 Comparison of RNA before and after DNase I treatment

2.2 差异蛋白质编码基因参与的生物学过程

对猪的蛋白质编码基因分析发现,相对于对照组,试验组有 1 295 个蛋白质编码基因呈现变化倍数。对这些差异表达的蛋白质编码基因进行 GO 的生物学过程分析发现,这些差异基因一共参与 327 个生物学进程,其中最显著富集的第 1 到第 10 个进程依次排序为 GO: 0006954 (炎症反应)、GO:0006955 (免疫反应)、GO: 0032496 (脂多糖反应)、GO: 0071222 (细胞对脂多糖反应)、GO:0050729 (炎症反应的正调节)、GO:0030335 (细胞迁移的正调节)、GO: 0006915 (凋亡过程)、GO: 0009615 (病毒反应)、GO:0050731 (肽酪氨酸磷酸化的正调节)与 GO:0051607 (病毒防疫反应)。这些进程涉及到炎症反应、免疫反应、脂多糖反应、炎症反应调节、细胞迁移调节、细胞凋亡、病毒响应、肽酪氨酸磷酸化等,而在这前 10 个进程中有 7 个涉及到免疫、炎症及抗病毒反应(图 2)。

2.3 lncRNA 的鉴定与结构分析

经过 HISAT2 比对和 StringTie 组装后的转录本经过层层过滤后最终得到 1 056 个 lncRNA 转录本,其对应于 831 个基因座,这些 lncRNA 的平均长度为 1 643 bp,平均外显子数为 2.4 个。对 lncRNA 的长度区间进一步分析发现(图 3),在 200~500 bp 与 501~1 000 bp 的 lncRNA 占总数的 52%,随着长度

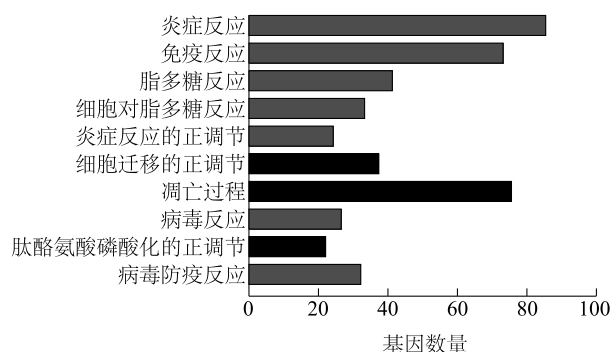


图 2 差异表达蛋白编码基因的参与的生物学过程分析

Fig.2 Analysis of the differentially expressed protein-coding genes involved in biological processes

的不断增加,lncRNA 所占的比例逐渐减小。而对于外显子分布的进一步分析发现,只有 1 个和 2 个外显子的 lncRNA 占总数的 64%,随着外显子数目的不断增加,lncRNA 所占的比例也逐渐减小。lncRNA 在不同的染色体上都有不同程度的分布,在 Y 染色体上的数量最少。

2.4 差异表达 lncRNA 的鉴定

通过分析比较,在对照组与试验组之间共有 95 个差异表达的 lncRNA,相对于对照组,试验组有 51 个 lncRNA 上调表达(表 2),44 个 lncRNA 下调表达(表 3)。

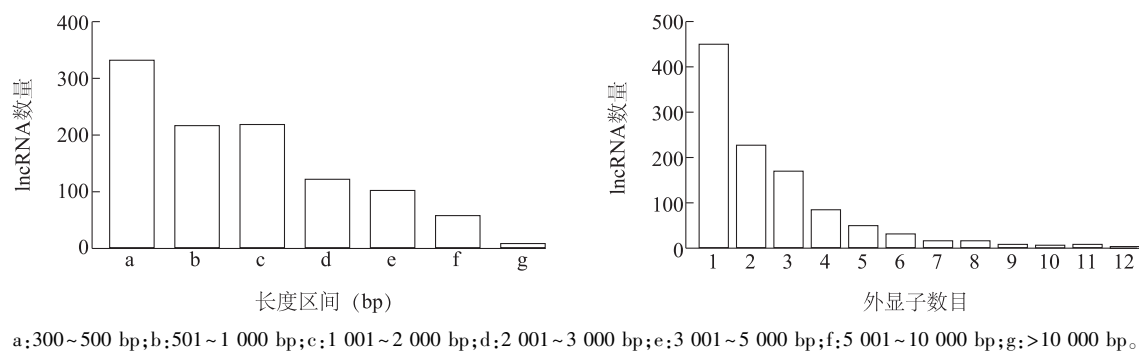


图 3 lncRNA 的长度区间与外显子数目的数量分布

Fig.3 The number distribution of length interval and exon number of lncRNA

表 2 上调表达的 lncRNA 位置信息

Table 2 The position information of up-regulated lncRNA

染色体	转录起始	转录终止	上调 lncRNA	染色体	转录起始	转录终止	上调 lncRNA
1	255 433 056	255 434 142	MSTRG.1106	8	69 878 233	69 879 532	MSTRG.8349
1	139 804 580	139 805 000	MSTRG.641	8	69 875 987	69 880 547	MSTRG.8348
1	7 505 933	7 507 364	MSTRG.69	8	31 412 535	31 416 912	MSTRG.8244
1	375 368	379 482	MSTRG.5	8	45 569 152	45 571 497	MSTRG.8289
1	159 076 233	159 096 880	MSTRG.691	8	69 994 814	69 996 501	MSTRG.8350
1	26 955 386	26 955 610	MSTRG.154	9	25 585 353	25 637 111	MSTRG.8826
1	92 517 518	92 517 938	MSTRG.365	11	21 048 617	21 051 648	MSTRG.9912
1	216 537 376	216 537 709	MSTRG.898	11	21 046 868	21 047 595	MSTRG.9911
1	236 365 727	236 366 164	MSTRG.975	12	40 698 426	40 703 022	MSTRG.10615
2	6 895 835	6 897 032	MSTRG.1470	12	54 054 878	54 057 499	MSTRG.10897
3	102 187 836	102 195 738	MSTRG.3673	12	40 698 426	40 700 470	MSTRG.10615
3	38 842 519	38 842 796	MSTRG.3165	12	40 823 086	40 823 757	MSTRG.10620
3	128 879 804	128 880 856	MSTRG.3826	12	22 629 869	22 630 155	MSTRG.10437
3	4 113 677	4 124 237	MSTRG.2726	13	204 834 894	204 840 384	MSTRG.11913
3	31 875 002	31 875 616	MSTRG.3126	15	79 172 657	79 174 090	MSTRG.13160
4	99 357 468	99 357 879	MSTRG.4483	15	68 748 385	68 751 458	MSTRG.13128
4	129 073 744	129 094 202	MSTRG.4717	15	39 015 258	39 016 369	MSTRG.13018
4	57 631 346	57 631 705	MSTRG.4098	16	63 034 306	63 061 203	MSTRG.13720
5	64 062 177	64 062 738	MSTRG.5306	18	6 463 176	6 463 488	MSTRG.14306
5	87 453 864	87 454 103	MSTRG.5460	18	50 458 437	50 463 100	MSTRG.14525
5	29 383 736	29 384 018	MSTRG.5122	X	99 741 549	99 741 771	MSTRG.14936
6	455 856	489 505	MSTRG.5543	X	40 588 040	40 596 461	MSTRG.14705
6	49 214 322	49 214 573	MSTRG.5967	X	40 569 859	40 596 461	MSTRG.14705
7	75 002 731	75 003 927	MSTRG.7780	X	40 593 019	40 681 904	MSTRG.14705
7	29 803 953	29 804 271	MSTRG.7401	X	40 588 263	40 596 427	MSTRG.14705
7	3 906 075	3 906 733	MSTRG.7113				

表 3 下调表达的 lncRNA 位置信息

Table 3 The position information of down-regulated lncRNA

染色体	转录起始	转录终止	下调 lncRNA	染色体	转录起始	转录终止	下调 lncRNA
1	233 321 288	233 321 498	MSTRG.957	8	936 943	938 409	MSTRG.8115
1	228 198 039	228 199 556	MSTRG.939	8	3 795 605	3 797 229	MSTRG.8145
1	269 575 516	269 607 453	MSTRG.1233	9	71 853 945	71 866 210	MSTRG.9079
2	1 208 237	1 224 112	MSTRG.1332	9	27 360 752	27 361 172	MSTRG.8841
2	45 629 310	45 629 594	MSTRG.1772	9	99 910 400	99 911 773	MSTRG.9176
2	1 214 174	1 214 463	MSTRG.1333	9	131 891 028	131 903 257	MSTRG.9370
2	131 722 277	131 749 921	MSTRG.2483	10	66 336 579	66 343 082	MSTRG.9730
2	103 664 365	103 667 114	MSTRG.2407	12	3 802 518	3 805 352	MSTRG.10177
2	1 218 873	1 219 944	MSTRG.1334	12	142 866	157 739	MSTRG.10056
2	137 670 118	137 670 798	MSTRG.2543	13	147 252 041	147 252 616	MSTRG.11761
2	1 616 996	1 618 105	MSTRG.1341	13	160 411 969	160 416 621	MSTRG.11815
2	1 613 791	1 616 215	MSTRG.1340	14	139 605 802	139 606 351	MSTRG.12873
3	41 576 291	41 576 514	MSTRG.3298	14	40 925 727	40 943 857	MSTRG.12242
3	77 638 016	77 645 290	MSTRG.3552	14	140 735 464	140 736 467	MSTRG.12889
4	79 369 825	79 370 516	MSTRG.4173	16	322 155	339 504	MSTRG.13487
5	67 917 852	67 925 172	MSTRG.5347	16	39 958 613	40 084 728	MSTRG.13612
6	4 656 195	4 668 006	MSTRG.5587	17	30 272 583	30 272 990	MSTRG.13899
6	69 634 999	69 638 356	MSTRG.6427	17	30 257 471	30 259 162	MSTRG.13901
7	98 809 934	98 810 345	MSTRG.7980	18	3 062 425	3 063 710	MSTRG.14263
7	97 721 466	97 723 622	MSTRG.7958	X	110 384 151	110 387 654	MSTRG.14975
8	135 420 840	135 428 488	MSTRG.8618	X	83 200 516	83 203 618	MSTRG.14880
8	19 815 386	19 817 717	MSTRG.8204	Y	9 238 340	9 239 060	MSTRG.15098

2.5 差异表达 lncRNA 的保守性分析

Ensemble 数据库中人有 7 856 个 lncRNA 基因,小鼠有 5 438 个 lncRNA 基因。进行 BLASTN 比对后,在 95 个差异的猪 lncRNA 中有 10 个 lncRNA 与人 lncRNA 同源;而只有 1 个 lncRNA 与小鼠 lncRNA 同源 ($E\text{-value} < 10^{-5}$),分别占到了其总数的 10.50% 和 1.05%,同时与人和小鼠 lncRNA 同源的有 1 个 lncRNA。在与小鼠 lncRNA 同源的猪 lncRNA 中,其同源序列长度跨度为 327~620 bp,平均长度为 514 bp;在与人类 lncRNA 同源的猪 lncRNA 中,其同源序列长度跨度为 31~1 301 bp,平均长度为 490 bp。

2.6 差异表达 lncRNA 的重复序列分析

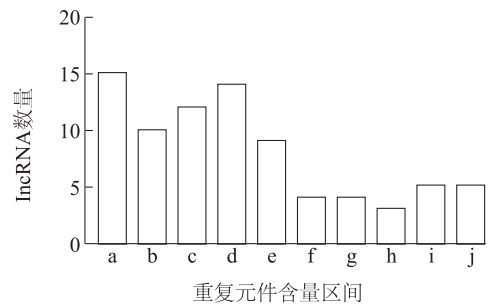
通过 RepeatMasker 分析发现,95 个差异 lncRNA 中有 81 个 lncRNA 序列中含有重复元件,占

到了总数的 85%。对 lncRNA 序列中重复元件的种类分析发现,短散在重复序列(SINE)和长散在重复序列(LINE)分别占到了重复元件总数的 40.4% 和 17.7%,长末端重复序列(LTR)占到了 11.6%,简单重复占 16.9%,这几种常见的重复元件占到了总重复元件类型的 86.6%,剩余的是一些不常见的类型。进一步对不同含量区间的 lncRNA 分析发现,以 10% 为 1 个区间,在含 0~10.0%,10.1%~20.0%,20.1%~30.0%,30.1%~40.0% 重复元件的 lncRNA 占总数的 10% 以上,含 50% 以上重复元件的 lncRNA 数量较少(图 4)。

2.7 差异表达 lncRNA 两侧蛋白质编码基因的 GO 与 KEGG 分析

通过对 95 个差异表达的 lncRNA 与猪 NCBI 和 Ensemble 蛋白质数据库染色体位置比较,163 个有

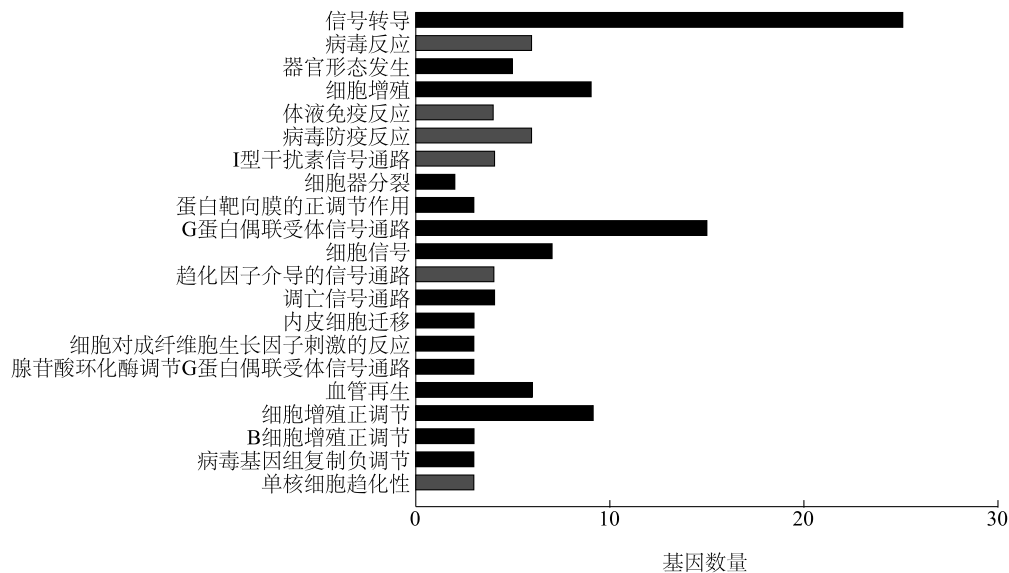
官方名称的蛋白质编码基因处于差异 lncRNA 两侧。我们对这 163 个蛋白质编码基因进行 GO 与 KEGG 分析。在 GO 分析中,这些蛋白质编码基因一共参与了 22 类生物学进程(图 5)。这 22 类进程中涉及到信号转导、抗病毒反应、器官形态发生、细胞增殖、体液免疫等,其中 6 个进程与病毒反应、免疫反应相关,分别为 GO:0009615 (病毒反应)、GO:0006959 (体液免疫反应)、GO:0051607 (病毒防疫反应)、GO:0060337 (I 型干扰素信号通路)、GO:0070098 (趋化因子介导的信号通路)、GO:0002548 (单核细胞趋化性)。



a: 0 ~ 10.0%; b: 10.1% ~ 20.0%; c: 20.1% ~ 30.0%; d: 30.1% ~ 40.0%; e: 40.1% ~ 50.0%; f: 50.1% ~ 60.0%; g: 60.1% ~ 70.0%; h: 70.1% ~ 80.0%; i: 80.1% ~ 90.0%; j: 90.1% ~ 100.0%。

图 4 差异表达 lncRNA 含有的重复元件比例分析

Fig.4 Analysis of the proportion of repeat elements in differentially expressed lncRNA



灰色条框代表与病毒反应、免疫反应、炎症因子相关。

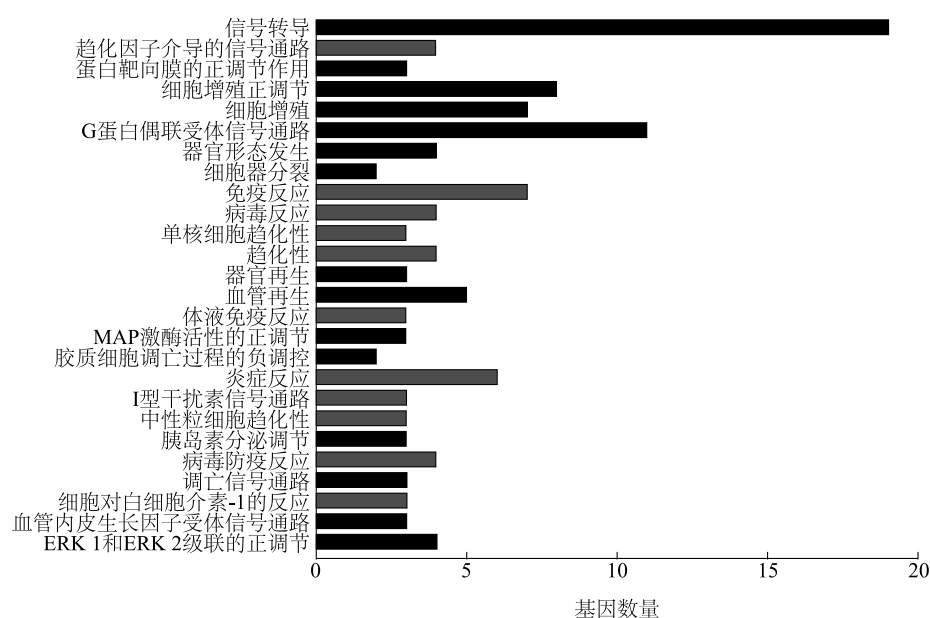
图 5 差异表达 lncRNA 两侧蛋白质编码基因的 GO 分析

Fig.5 GO analysis of the neighboring protein-coding genes of differentially expressed lncRNA

对 51 个上调表达的 lncRNA 两侧的蛋白质编码基因进行 GO 分析发现,这些蛋白质编码基因一共参与了 26 类生物学进程(图 6)。这 26 类进程中涉及到信号转导、趋化因子信号通路、细胞增殖、G 蛋白受体信号、器官形态发生等,其中有 11 个进程涉及到病毒反应、免疫反应、炎症因子,分别为 GO:0070098 (趋化因子介导的信号通路)、GO:0006955 (免疫反应)、GO:0009615 (病毒反应)、GO:0002548 (单核细胞趋化性)、GO:0006935 (趋化性)、

GO:0006959 (体液免疫反应)、GO:0006954 (炎症反应)、GO:0060337 (I 型干扰素信号通路)、GO:0030593 (中性粒细胞趋化性)、GO:0051607 (病毒防疫反应)、GO:0071347 (细胞对白细胞介素-1 的反应)。

对 44 个下调表达的 lncRNA 两侧的蛋白质编码基因进行 GO 分析发现,这些蛋白质编码基因一共参与了 3 类生物学过程(图 7)。这 3 个过程涉及到 DNA 重组、RNA 聚合酶 II 转录调控、心内膜垫层形成等。



灰色条框代表与病毒反应、免疫反应、炎症因子相关。

图 6 上调表达 lncRNA 两侧蛋白质编码基因的 GO 分析

Fig.6 GO analysis of the neighboring protein-coding genes of up-regulated expressed lncRNA

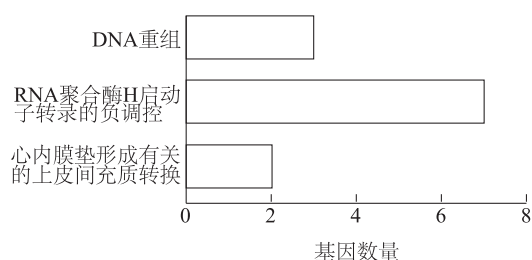


图 7 下调表达 lncRNA 两侧蛋白质编码基因的 GO 分析

Fig.7 GO analysis of the neighboring protein-coding genes of down-regulated expressed lncRNA

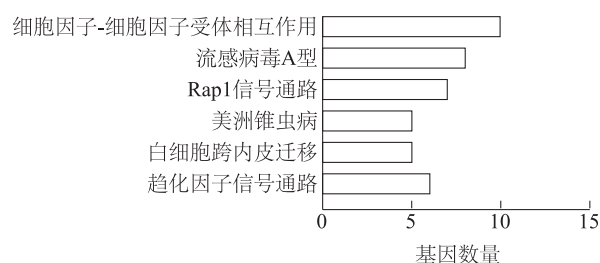


图 8 差异表达 lncRNA 两侧蛋白质编码基因参与的 KEGG pathway 分析

Fig.8 KEGG pathway analysis of the neighboring protein-coding genes of differentially expressed lncRNA

对差异表达的 lncRNA 两侧的 163 个蛋白质编码基因进行 KEGG pathway 分析发现,这些蛋白质编码基因总共参与了 6 个 Pathway,这些 Pathway 涉及到细胞因子-受体结合、A 型流感、Rap1 信号通路、美洲锥虫病、白细胞跨内皮迁移与趋化因子信号通路(图 8),其中绝大部分与免疫反应、炎症反应相关。对 51 个上调表达的 lncRNA 两侧蛋白质编码基因进行 KEGG pathway 分析发现,这些蛋白质编码基因总共参与了 7 个 Pathway,这些 Pathway 涉及到细胞因子-受体结合、A 型流感、Rap1 信号通路、类风湿性关节炎、美洲锥虫病、白细胞跨内皮迁移与趋化因子信号通路(图 9),其中绝大部分也与免疫反应、炎症反应相关。然而对 44 个下调表达的 lncRNA

两侧的 75 个蛋白质编码基因进行 KEGG pathway 分析时发现没有 Pathway 被显著富集。

2.8 差异表达 lncRNA 定量 PCR 验证

我们上调和下调的 lncRNA 中分别选取 2 个 lncRNA 进行定量 PCR 验证。在上调的 2 个 lncRNA 中,lncRNA-MSTRG.8350 的两侧蛋白质编码基因为 *CXCL8* 与 *CXCL6*,通过上述的 GO 与 KEGG 注释显示它们与免疫反应、趋化因子通路相关;lncRNA-MSTRG.8244 其两侧蛋白质编码基因 *RHOH* 与 *CHRNA9* 分别与信号传导和白细胞跨内皮迁移相关,与炎症反应间接相关。下调中的 lncRNA 两侧蛋白质编码基因在 GO 与 KEGG 注释中均没有涉及

到免疫反应、炎症反应、趋化因子、病毒响应,故随机选取了 2 个 lncRNA,分别为 lncRNA-MSTRG.3552 与 lncRNA-MSTRG.12242。定量 PCR 结果显示,与对照组相比 lncRNA-MSTRG.8350 和 lncRNA-MSTRG.8244 显著上调($P<0.01$), lncRNA-MSTRG.12242 和 lncRNA-MSTRG.3552 显著下调($P<0.01$) (图 10),上调和下调差异倍数与 RNA-seq 的差异倍数相似,表 4 为 qPCR 差异倍数与 RNA-Seq 差异倍数比较。

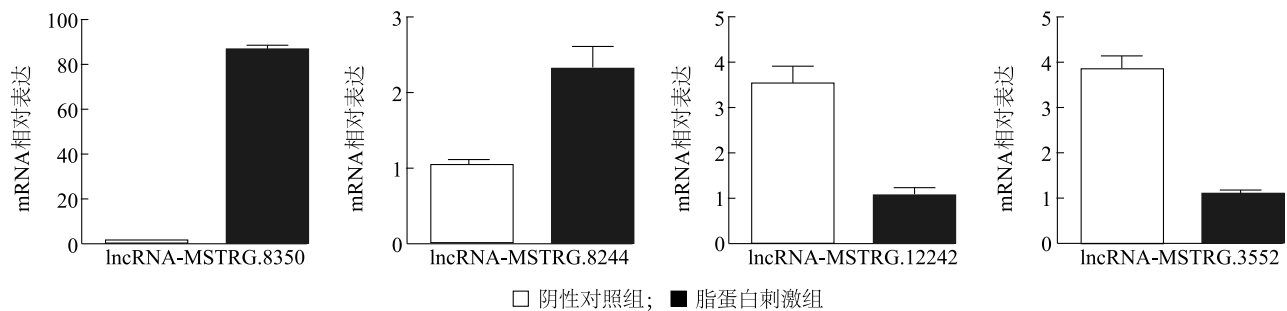


图 10 差异 lncRNA 定量 PCR 验证

Fig.10 Validation of differentially expressed lncRNA by qPCR

表 4 差异表达 lncRNA 的 RNA-Seq 测序与 qPCR 比较

Table 4 Comparison of differentially expressed lncRNA between RNA-Seq sequencing and qPCR

差异 lncRNA	表达	差异倍数	
		RNA-Seq 测序	qPCR
lncRNA-MSTRG.8350	上调	55.9	86.8
lncRNA-MSTRG.8244	上调	3.8	2.2
lncRNA-MSTRG.12242	下调	5.9	3.3
lncRNA-MSTRG.3552	下调	3.4	3.6

2.9 lncRNA 组织表达谱

以成年猪的心脏、肝、脾、肺、肾、小肠、背肌等组织的 cDNA 为模板,对上述上调和下调的 4 个 lncRNA 进行组织表达谱分析。结果(图 11)显示在上调的 2 个 lncRNA 中, lncRNA-MSTRG.8350 仅在肺中有微弱的表达,不表达于其他组织, lncRNA-MSTRG.8244 在脾中表达相对较高,在肝、小肠中表达较弱。在下调的 2 个 lncRNA 中, lncRNA-MSTRG.12242 与 lncRNA-MSTRG.3552 均呈现多组织表达。

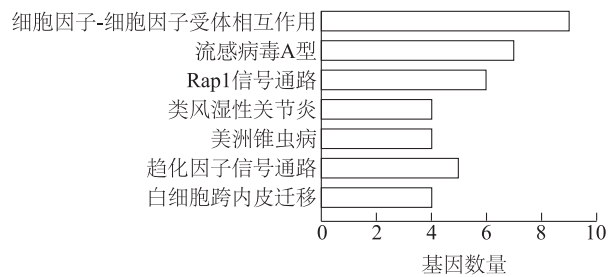
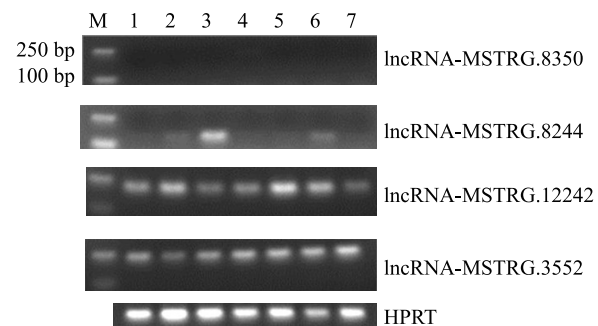


图 9 上调表达 lncRNA 两侧蛋白质编码基因参与的 KEGG pathway 分析

Fig.9 KEGG pathway analysis of the neighboring protein-coding genes of up-regulated expressed lncRNA



M: DL2000 marker; 1~7 分别代表为心脏、肝、脾、肺、肾、小肠、背肌。

图 11 lncRNA 的组织表达谱

Fig.11 Tissue expression profile of lncRNA

3 讨论

lncRNA 可通过本身的复杂结构与 DNA、RNA 或蛋白质之间互作,从而在多种层面上调控炎症因子的表达水平^[16,24]。从 lncRNA 这一新的调控分子去研究炎症因子的表达调控能更加全面地理解炎症反应的机理。基于此,本研究通过 RNA-Seq 测序并结合生物信息学分析对猪单核源性巨噬细胞受支原

体脂蛋白 FSL-1 刺激后的 lncRNA 进行了鉴定与特征分析。

支原体细胞膜上的脂质相关膜蛋白是其介导宿主发生天然免疫的重要抗原^[25],由于绝大部分支原体的脂蛋白结构是双酰基而不是三酰基^[26],因而在脂蛋白上选取了双酰基结构的 FSL-1 脂蛋白,这更好地模拟支原体脂蛋白对细胞的炎症反应。我们发现在脂蛋白 FSL-1 刺激细胞后,其差异基因参与的最显著的通路都集中在炎症与免疫反应上,这说明 FSL-1 刺激成功地构建了细胞炎症模型。我们在 lncRNA 的鉴定中由于 lncRNA 基因与蛋白质编码基因的起始位置不重叠,实际上得到的是 lincRNA。lincRNA 由于不与其他蛋白质编码基因重叠而具有相对独立的转录单元,这也便于对其表达和调控进行更好的研究。lncRNA 基因一般长度在 1 kb 左右,外显子数目大多数为 1~3^[27]。本研究结果显示 lncRNA 平均长度为 1 643 bp,平均外显子数为 2.4 个,而且长度在 1 000 bp 以下和外显子数在 2 以下的 lncRNA 都占到了总数的 50% 以上,表现出较短的长度以及较少的外显子数,这也与报道相符合,表明我们得到的 lncRNA 是可信赖的。lncRNA 基因由于部分和经典的蛋白质编码基因 mRNA 结构相似,也显示出一定的转录剪接^[5,28],我们鉴定的 1 056 个 lncRNA 转录本,其对应于 831 个基因座,这也说明其中有些 lncRNA 基因发生了选择性剪接,产生了 2 个以上的转录本。我们发现在差异 lncRNA 基因中,分别有 10.50% 和 1.05% 的 lncRNA 与人和小鼠同源,这表明相对于小鼠,猪的 lncRNA 与人的更为保守。研究结果表明基因组的重复序列可能是 lncRNA 产生的主要来源^[29]。我们对差异 lncRNA 的序列分析发现绝大多数的差异 lncRNA 中都含有重复序列,这也与上述结论相符。进一步分析发现短散在重复序列(SINE)占到了差异 lncRNA 中总重复序列种类的 40%,这说明差异 lncRNA 可能大部分来源于基因间的短散在重复序列区间。

lncRNA 通过顺式作用^[30]与反式作用^[16]来发挥其重要功能,虽然目前还未有较好的方法来准确预测 lncRNA 的功能,然而通常可根据 lncRNA 的这些调控方式来预测其功能,一般对 lncRNA 两侧蛋白质编码基因参与的通路来预测 lncRNA 可能的功能。在本研究中,我们对差异 lncRNA 两侧的蛋白

质编码基因进行 GO 与 KEGG 分析发现,在 GO 分类中,差异 lncRNA 包括上调与下调 lncRNA 两侧的蛋白质编码基因参与了 22 个生物学过程,其中 6 个与免疫反应、炎症反应直接相关,而对上调的 lncRNA 两侧的蛋白质编码基因分析发现,它们参与 26 个生物学过程,其中 11 个与免疫反应、炎症反应直接相关,占到了大约一半的比例;而对下调的 lncRNA 两侧的蛋白质编码基因分析发现,没有参与免疫反应、炎症反应的通路。这说明上调的 lncRNA 更倾向于富集在与免疫反应、炎症反应相关的基因两侧,可能具有重要作用。KEGG 的通路分析结果也与 GO 分析结果相符合,上调的 lncRNA 两侧蛋白质编码基因几乎都参与免疫反应、炎症反应相关的通路,而下调的 lncRNA 两侧的蛋白质编码基因没有参与生物学进程。我们进一步对上调和下调的 lncRNA 进行组织表达谱分析,结果表明位于与炎症反应相关的蛋白质编码基因两侧的 2 个上调 lncRNA,都呈现出一定的组织特异表达,其中一个表现出严格的组织表达,而在下调的 lncRNA 中与此相比都表现出多组织表达。研究结果表明,严格的组织表达或时序特异性表达的 lncRNA 具有重要功能^[30-32],这也在一定程度上说明了上调的 lncRNA 尤其是位于与炎症反应相关的蛋白质编码基因两侧的 lncRNA 值得进一步深入研究。

参考文献:

- [1] WANG J, ZHANG J, ZHENG H, et al. Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs[J]. *Nature*, 2004, 431(7010): 1-2.
- [2] BROWN C J, HENDRICH B D, RUPERT J L, et al. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus[J]. *Cell*, 1992, 71(3): 527-542.
- [3] CLEMONS C M, HUTCHINSON J N, SARA S A, et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles[J]. *Molecular Cell*, 2009, 33(6): 717-726.
- [4] LOEWER S, CABILI M N, GUTTMAN M, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells[J]. *Nat Genet*, 2010, 42(12): 1113-1117.
- [5] ULITSKY I, SHKUMATAVA A, JAN C H, et al. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution[J]. *Cell*, 2011, 147(7): 1537-1550.
- [6] GUTTMAN M, DONAGHEY J, CAREY B W, et al. LincRNAs

- act in the circuitry controlling pluripotency and differentiation[J]. *Nature*, 2011, 477(7364): 295-300.
- [7] WANG J, GONG C, MAQUAT L E. Control of myogenesis by rodent SINE-containing lncRNAs[J]. *Genes Dev*, 2013, 27(7): 793-804.
- [8] WEI S, DU M, JIANG Z, et al. Long noncoding RNAs in regulating adipogenesis: new RNAs shed lights on obesity[J]. *Cell Mol Life Sci*, 2016, 73(10):2079-2087.
- [9] PRENSNER JR, CHINNAIYAN AM. The emergence of lncRNAs in cancer biology[J]. *Cancer Discov*, 2011, 1(5):391-407.
- [10] LIN C, WANG Y, WANG Y, et al. Transcriptional and posttranscriptional regulation of HOXA13 by lncRNA HOTTIP facilitates tumorigenesis and metastasis in esophageal squamous carcinoma cells[J]. *Oncogene*, 2017, 36(38):5392-5406.
- [11] HAN J, ULEVITCH R J. Limiting inflammatory responses during activation of innate immunity[J]. *Nat Immunol*, 2005, 6(12): 1198-1205.
- [12] FAULKNER C B, SIMECKA J W, DAVIDSON M K, et al. Gene expression and production of tumor necrosis factor alpha, interleukin 1, interleukin 6, and gamma interferon in C3H/HeN and C57BL/6N mice in acute *Mycoplasma pulmonis* disease[J]. *Infect Immun*, 1995, 63:4084-4090.
- [13] FONSECA-ATEN M, RÍOS A M, MEJÍAS A, et al. *Mycoplasma pneumoniae* induces host-dependent pulmonary inflammation and airway obstruction in mice[J]. *Am J Respir Cell Mol Biol*, 2005, 32(3):201-210.
- [14] O'CONNELL R M, RAO D S, BALTIMORE D. MicroRNA regulation of inflammatory responses[J]. *Annu Rev Immunol*, 2012, 30:295-312.
- [15] CARPENTER S, AIELLO D, ATIANAND M K, et al. A long noncoding RNA mediates both activation and repression of immune response genes[J]. *Science*, 2013, 341(6147):789-792.
- [16] LI Z, CHAO T C, CHANG K Y, et al. The long noncoding RNA THRIL regulates TNF α expression through its interaction with hnRNPL[J]. *Proc Natl Acad Sci USA*, 2014, 111(3):1002-1007.
- [17] 韦显凯,覃 晴,祝 健,等.小鼠 *TLR3* 基因双敲除打靶载体构建及巨噬细胞 RAW264.7^{TLR3-/-}细胞系的建立[J]. *南方农业学报*, 2018, 49(5):993-999.
- [18] HICKMAN-DAVIS J M, MICHALEK S M, GIBBS-ERWIN J, et al. Depletion of alveolar macrophages exacerbates respiratory mycoplasmosis in mycoplasma-resistant C57BL mice but not mycoplasma-susceptible C3H mice[J]. *Infect Immun*, 1997, 65(6): 2278-2282.
- [19] 倪黎纲,赵旭庭,王宵燕,等.姜曲海猪肺泡巨噬细胞的分离培养与鉴定[J]. *江苏农业科学*, 2017, 45(23):163-165.
- [20] BYERS D E, HOLTZMAN M J. Alternatively activated macrophages and airway disease[J]. *Chest*, 2011, 140(3):768-774.
- [21] LACEY D C, ACHUTHAN A, FLEETWOOD A J, et al. Defining GM-CSF- and macrophage-CSF-dependent macrophage responses by in vitro models[J]. *J Immunol*, 2012, 188(11):5752-5765.
- [22] CHITKO-MCKOWN C G, CHAPES S K, MILLER L C, et al. Development and characterization of two porcine monocyte-derived macrophage cell lines[J]. *Results Immunol*, 2013, 3:26-32.
- [23] KONG L, ZHANG Y, YE Z Q, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine[J]. *Nucleic Acids Res*, 2007, 35:345-349.
- [24] GOMEZ J A, WAPINSKI O L, YANG Y W, et al. The nest long ncRNA controls microbial susceptibility and epigenetic activation of the interferon- γ locus[J]. *Cell*, 2013, 152:743-754.
- [25] BROWNING G F, MAREDA M S, NOORMOHAMMADI A H, et al. The central role of lipoproteins in the pathogenesis of mycoplasmoses[J]. *Vet Microbiol*, 2011, 153(1/2):44-50.
- [26] SASAKI Y, ISHIKAWA J, YAMASHITA A, et al. The complete genomic sequence of *Mycoplasma penetrans*, an intracellular bacterial pathogen in humans[J]. *Nucleic Acids Res*, 2002, 30: 5293-5300.
- [27] CABILI M N, TRAPNELL C, GOFF L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses[J]. *Genes Dev*, 2011, 25(18):1915-1927.
- [28] NUMATA K, KANAI A, SAITO R, et al. Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection[J]. *Genome Research*, 2003, 13: 1301-1306.
- [29] KAPUSTA A, KRONENBERG Z, LYNCH V J, et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs[J]. *PLoS Genet*, 2013, 9(4):e1003470.
- [30] ANDERSON K M, ANDERSON D M, MCANALLY J R, et al. Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development[J]. *Nature*, 2016, 539(7629):433-436.
- [31] CHEN J F, MANDEL E M, THOMSON J M, et al. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation[J]. *Nat Genet*, 2006, 38(2):228-233.
- [32] GROTE P, WITTLER L, HENDRIX D, et al. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse[J]. *Dev Cell*, 2013, 24(2):206-214.

(责任编辑:陈海霞)