

常卫华, 杨富贵, 张桥桥, 等. 绵羊妊娠早期 lncRNA 生物标志物的筛选与鉴定 [J]. 畜牧与兽医, 2024, 56 (11): 1-9.

CHANG W H, YANG F G, ZHANG Q Q, et al. Screening and identification of lncRNA biomarkers in early pregnancy of sheep [J]. Animal Husbandry & Veterinary Medicine, 2024, 56 (11): 1-9.

## 绵羊妊娠早期 lncRNA 生物标志物的筛选与鉴定

常卫华<sup>1,2,3</sup>, 杨富贵<sup>1</sup>, 张桥桥<sup>1</sup>, 王娟红<sup>1</sup>, 倪国超<sup>1</sup>, 赵影<sup>1</sup>, 严光文<sup>1\*</sup>

(1. 西昌学院动物科学学院, 四川 西昌 615013;

2. 四川攀西动物疫病检测与防控实验室, 四川 西昌 615013;

3. 四川地方特色山羊重点实验室, 四川 西昌 615013)

**摘要:** 旨在探讨长链非编码 RNA (lncRNA) 在绵羊妊娠早期中的调控作用, 研究其对绵羊生殖生理的调控机制。通过高通量技术对不同生理时期绵羊卵巢进行转录组测序, 结合生物信息学获得差异 lncRNA; 利用顺式 (cis) 和反式 (trans) 关系探讨差异 lncRNA 靶基因, 对靶基因进行基因本体 (GO) 和京都基因与基因组百科全书 (KEGG) 通路分析, 通过实时荧光定量 PCR (RT-qPCR) 研究 lncRNA 表达模式。结果: 试验成功构建了绵羊妊娠早期 RNA 文库, 每个样品平均产出 11.50 G 数据, 并且每个样品与基因组平均比对率为 93.88%; 试验共检测到 23 219 个差异基因, 其中 mRNA 418 个, lncRNA 49 个 (已知 17 个, 候选 32 个); 对 49 个差异 lncRNA 通过 cis 和 trans 关系分析获得 341 个靶基因; GO 和 KEGG 通路研究表明, 5 个 lncRNA (LOC105612389、BGIG9940\_27655、RPS28、BGIG9940\_27548 和 BGIG9940\_27563) 在妊娠期表达显著或极显著升高 ( $P < 0.05$  或  $P < 0.01$ ), 均参与早期胎儿发育及相关调控过程; RT-qPCR 结果表明, 5 个 lncRNA 基因表达量达到一定临界值时, 预示该母羊已经妊娠。提示: 5 个 lncRNA 单独或联合使用有望作为早期妊娠检查的候选生物标志物, 对绵羊早期妊娠诊断及提高经济效益具有十分重要的意义。

**关键词:** 高通量测序; 长链非编码 RNA; 生物标志物; 妊娠

**中图分类号:** S826.3 **文献标志码:** A **文章编号:** 0529-5130(2024)11-0001-09

## Screening and identification of lncRNA biomarkers in early pregnancy of sheep

CHANG Weihua<sup>1,2,3</sup>, YANG Fugui<sup>1</sup>, ZHANG Qiaoqiao<sup>1</sup>, WANG Juanhong<sup>1</sup>, NI Guochao<sup>1</sup>,

ZHAO Ying<sup>1</sup>, YAN Guangwen<sup>1\*</sup>

(1. College of Animal Science, Xichang University, Xichang 615013, China;

2. Laboratory for Testing and Prevention of Animal Diseases in Panxi Region, Xichang 615013, China;

3. Key Laboratory of Local Goat in Sichuan, Xichang 615013, China)

**Abstract:** This study was conducted to explore the regulatory role of long non-coding RNA (lncRNA) in early pregnancy of sheep, and to provide reference for studying the regulatory mechanism of lncRNA in reproductive physiology of sheep. Transcriptome sequencing of ovine ovaries of the sheep at different physiological stages were performed by high-throughput technology, and different lncRNAs were obtained by combining bioinformatics. Then, the differential lncRNA target genes were explored by cis- and trans- relationships, and the GO and KEGG pathways were analyzed for target genes. Also, the expression pattern of lncRNA was explored by RT-qPCR. The results showed that an early pregnancy RNA library of sheep was successfully constructed and each sample produced an average of 11.50 G of data. The average comparison rate of the genome was 93.88%. A total of 23 219 genes were detected, including 418 differential mRNA and 49 differential lncRNA (17 known lncRNAs and 32 candidate new lncRNAs). 341 target genes were obtained from 49 different lncRNAs by cis- and trans- relationship analysis. The differential expression, GO and KEGG pathway studies showed that all 5 lncRNAs (LOC105612389, BGIG9940\_27655, RPS28, BGIG9940\_27548, and BGIG9940\_27563) significantly increased during pregnancy ( $P < 0.05$  or  $P < 0.01$ ), and they were involved in the early fetal development and related regulatory processes in the sheep. Moreover, the RT-qPCR results showed that, the expression level of the five genes reaching a certain critical value indicated that the ewes were pregnant. These results also suggested that the 5 lncRNAs

收稿日期: 2024-01-18; 修回日期: 2024-09-18

基金项目: 国家自然科学基金地区项目 (32360905); 中国农业大学对口支援西昌学院联合基金项目 (ZXL202403); 2023 年西昌市技术研究开发与推广项目 [ (2023) 53-11 ]

第一作者: 常卫华, 男, 博士, 副教授

\* 通信作者: 严光文, 博士, 教授, 研究方向为动物疫病防控及兽医临床疾病诊治, E-mail: ygwldky@126.com。

alone or in combination could be used as candidate biomarkers for early pregnancy detection, which is of great significance for the diagnosis of early pregnancy in sheep and improving economic benefits of sheep breeding.

**Keywords:** high-throughput sequencing; lncRNA; biomarker; pregnancy

母羊的繁殖力直接关系到羊场的生产效益, 高效且精准的妊娠诊断技术可以有效减少羊群的空怀天数以及空怀导致的经济损失。目前的妊娠诊断技术仍有操作繁琐、对仪器设备要求和人员专业化要求高等许多弊端, 急需开发出一种简便、快速、准确的早期妊娠诊断技术。应用生物技术对羊实行早期妊娠诊断, 可以在母羊怀孕早期阶段快速检测出母羊是否怀孕, 从而对母羊进行分群精心管理和对空怀母羊进行及时补配, 同时也是提高母羊繁殖效率, 促进集约化管理, 提高养殖效益的重要技术措施。

近年来, 随着基因组学、蛋白组学、代谢组学等多组学技术和生物信息学方法的发展, 在血清、血浆、全血及组织中多种类型的生物标志物被鉴定并应用于生产实践。生物标志物 (biomarker) 是指可以标记系统、器官、组织、细胞及亚细胞结构或功能改变或可能发生改变的生化指标, 具有非常广泛的临床应用前景; 通过对生物标志物的测定可以获知机体当前所处的生物学过程中的进程; 检查一种或多种疾病特异性的生物标志物, 对于疾病的鉴别, 早期诊断及预防, 治疗过程中的监控可能起到帮助作用<sup>[1-4]</sup>。寻找和发现有价值的生物标志物已经成为目前研究的一个重要热点。随着高通量测序技术的快速发展, 数字基因表达谱 (DGE) 测序、蛋白组测序、转录组测序、代谢组测序、小 RNAs 测序、降解组测序、DNA 甲基化测序、染色质免疫共沉淀 DNA 测序等新方法为科学家们开展分子生物学相关研究提供了更多选择, 给基因组学研究带来了一个高效的新平台和巨大的发展机遇<sup>[5-7]</sup>, 是发现生物标志物的重要手段之一。长链非编码 RNA (long non-coding RNA, lncRNA) 是一类本身不编码蛋白, 转录本长度超过 200 nt 的 lncRNA 分子<sup>[8-9]</sup>。lncRNA 通过与染色质修饰蛋白或转录因子与其特定的蛋白结合基序相互作用来发挥调控作用, 继而维持基因表达的活跃或不活跃状态<sup>[8-9]</sup>; lncRNA 还可以直接结合到与其互补的 DNA 序列, 形成一个 RNA-DNA 三重结构, 由此可以阻断转录过程。据报道, 这种三聚体结构可以导致靶基因启动子中特定序列的甲基化<sup>[10-12]</sup>。lncRNA 转录自身的行为甚至可以通过提供一个更开放的染色质区域来触发附近基因的表达, 它也可以通过占据转录因子结合位点来沉默附近的基因, 对生物生命活动起着极其重要的调控作用<sup>[13]</sup>, 而高通量测序及生物信息学分析是发现 lncRNA 的重要途径之一。鉴于此,

试验应用转录组学技术鉴定羊妊娠早期的 lncRNA 差异表达, 挖掘早期妊娠诊断生物标志物, 为建立转录组水平的妊娠诊断技术提供素材, 同时为制定人工干预策略阻止胚胎丢失提供技术指导。

## 1 材料与方法

### 1.1 主要药品与试剂

孕酮阴道栓 (CIDR) 购自微球生命科学研究 (成都) 有限公司, 孕马血清促性腺激素 (PMSG)、前列腺素 (PG) 购自宁波第二激素厂, TRIzol 试剂盒购自宝生物工程 (大连) 有限公司, Taq DNA 聚合酶、T4 RNA 连接酶等购自美国 Invitrogen 生物公司, Illumina 高通量测序在华大基因科技服务有限公司 (BGI) 完成。

### 1.2 试验动物及样品的采集

于四川省布拖县某绵羊良种繁育场购买 12 只雌性、产羔 2 次以上、体重 ( $65 \pm 2.5$ ) kg 的本地成年健康绵羊为试验绵羊, 隔离观察饲养 30 d, 自由采食, 发情期正常; 随机分 4 组, 每组均通过 CIDR+PMSG+PG 法进行同期发情。同期发情程序: 放栓当天记为 0 d, 7 d 换栓, 14 d 早上 8:00 撤栓, 同时注射 PMSG 400 IU 和 PG 0.1 g, 15 d 和 16 d 早晚试情。在卵泡期、黄体期 (发情后 10 d)、妊娠 30 d 和妊娠 45 d 分别屠宰, DEPC 水冲洗卵巢, 然后液氮快速冷冻并转移到  $-80$  °C 冰箱保存备用。

### 1.3 RNA 提取

从  $-80$  °C 冰箱中取出黄体期 (LP)、卵泡期 (FP)、妊娠 30 d (P30)、妊娠 45 d (P45) 卵巢, 通过液氮预冷研钵, 用 TRIzol 试剂提取总 RNA, 溶于适量无 RNA 酶水, 取少量总 RNA 用微量核酸分析仪 (Implen 公司, 德国) 测定其浓度及纯度, 并通过 1% 琼脂糖凝胶电泳检测其样品完整性, 合格总 RNA 样品保存于  $-80$  °C 超低温冰箱备用。

### 1.4 文库构建及高通量测序

取 0.1 g 总 RNA 样品, 利用 Oligo-dT 磁珠和 Ampure XP Beads 进行纯化后合成二链 cDNA, 通过 PCR 扩增建立文库, 然后在 Illumina 2000 测序系统上进行高通量测序, 每个时期样品 3 个重复。

### 1.5 测序数据分析

使用 SOAPnuke (v1.5.2) 对测序数据进行过滤, 具体方法: 去除含有测序接头的数据; 去除碱基质量不佳 (碱基质量小于或等于 5) 大于 20% 的数

据；去除未知碱基（N 碱基）比率大于 5% 的数据，然后得到过滤后数据，并以 FASTQ 格式存储。使用 HISAT2 (v2.0.4) 将过滤后数据映射到参考基因组 ([https://www.ncbi.nlm.nih.gov/datasets/genome/GCF\\_002742125.1/](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_002742125.1/))。然后用 EricScript (v0.5.5) 和 rMATS (v3.2.5) 分别检测融合基因和差异剪接基因 (DSG)。用 Bowtie2 (v2.2.5) 将过滤后数据与华大基因 (深圳) 建立的数据库进行比对，其中包括已知及新的、编码和非编码转录本，然后用 RSEM (v1.2.12) 计算基因的表达水平。根据不同样本的基因表达情况，用 Pheatmap (v1.0.8) 绘制热图。使用 DESeq 2.0 (v1.4.5) 进行差异表达分析，且  $P \leq 0.05$ 。

### 1.6 lncRNA 靶基因预测

对获得的差异 lncRNA 进行靶基因预测，lncRNA 和 mRNA 的斯皮尔曼相关系数、皮尔逊相关系数，要求均  $\geq 0.6$ ，完成相关系数分析后，将 lncRNA 在 mRNA 上游 100 kb 内或下游 100 kb 内的判定为顺式 (cis) 关系，超出这个范围，则用 RNAplex 分析二者结合能，如果结合能小于 -30，则判定为反式 (trans) 关系。参数：最小自由能 (max MFE)：-100；最大数量限制 (max num; Limit)：50。

使用基因本体 (GO) (<http://www.geneontology.org/>)

和京都基因与基因组百科全书 (KEGG) (<https://www.kegg.jp/>) 通路分析在超几何检验的基础上，用 Phyper ([https://en.wikipedia.org/wiki/Hypergeometric\\_distribution](https://en.wikipedia.org/wiki/Hypergeometric_distribution)) 对靶基因进行富集分析，通过靶基因判断 lncRNA 的功能作用。

### 1.7 利用实时荧光定量 PCR (RT-qPCR) 临床检测 lncRNA 的表达

临床采集配种 30~60 d 的母羊血液 30 份，屠宰场采集妊娠母羊卵巢 10 个，以未妊娠母羊的血液和卵巢为对照，以甘油醛-3-磷酸脱氢酶 (GAPDH) 为内参，进行 RT-qPCR 检测，每个试验重复 3 次。按照试剂盒操作步骤进行血液 RNA 的提取，组织按照 TRIzol 法进行 RNA 提取，之后进行反转录。无 RNA 酶管中加入总 RNA 4  $\mu\text{L}$ ，gDNA Eraser 1  $\mu\text{L}$ ，RNase-free  $\text{H}_2\text{O}$  5  $\mu\text{L}$ ，吹打混匀后 42  $^\circ\text{C}$  孵育 5 min。反应结束后置冰上加入 2 $\times$ SPARKscript II RT Plus Master Mix 10  $\mu\text{L}$ ，轻轻混匀离心，50  $^\circ\text{C}$  孵育 15 min，85  $^\circ\text{C}$  加热 5 min，反转录成 cDNA。RT-qPCR 反应体系 20  $\mu\text{L}$ ：2 $\times$ SYBR qPCR Mix 10  $\mu\text{L}$ ，上、下游引物 (2  $\mu\text{mol/L}$ ) 各 0.4  $\mu\text{L}$ ，cDNA 1.2  $\mu\text{L}$ ，RNase-free dd $\text{H}_2\text{O}$  8  $\mu\text{L}$ 。RT-qPCR 反应条件：94  $^\circ\text{C}$  3 min；94  $^\circ\text{C}$  10 s，60  $^\circ\text{C}$  30 s，40 个循环。引物序列见表 1。试验通过  $2^{-\Delta\Delta\text{Ct}}$  方法计算基因相对表达量。

表 1 RT-qPCR 引物序列及参数

基因	引物序列 (5'→3')	产物长度/bp
LOC105612389	F: TGAGGTCTGTTCATCCCATT R: GCTTTGCCGCCTCAAATACAA	175
BGIG9940_27655	F: TCCGAAACACGTGACTCCTC R: CCATGAATTCTACGCCGACC	154
BGIG9940_27548	F: TTTGAGAAACCTCTGCGCCA R: AGGCATGCAGTCCACACTTT	183
BGIG9940_27563	F: CCATCGAGTCAAGTATGCCA R: TTCTTCCCTACAAATAGGTTTCATCT	178
RPS28	F: ACTGCCATTTCTGAGCCGAT R: GTTTGTATCTCGAAGCAGCAGG	121
GAPDH	F: GTCGGAGTGAACGGATTTGG R: CATTGATGACGAGCTTCCCG	196

### 1.8 数据统计与分析

试验数据通过 edgeR (以 RSEM 进行定量)、SPSS 22.0 软件 ( $t$  检验) 进行差异显著性分析， $P < 0.05$  表示差异显著， $P > 0.05$  表示差异不显著。

## 2 结果与分析

### 2.1 文库构建和高通量测序

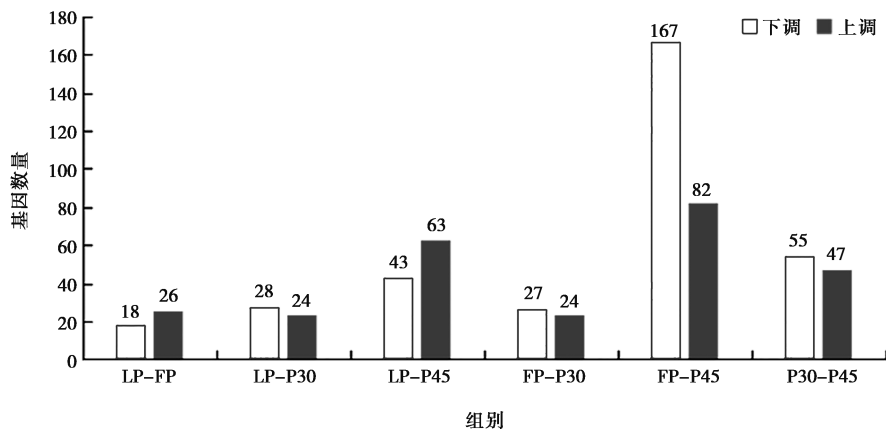
试验使用 DNBSEQ 平台对卵泡期 (FP)、黄体期 (LP)、妊娠 30 d (P30) 和妊娠 45 d (P45) 一共 12 个样品进行了建库测序，每个样品平均产出

11.50 G数据。由表 2 可见, 样品比对基因组的平均比率为 93.88%。一共检测到 23 219 个差异基因, 其中 mRNA 418 个, lncRNA 49 个 (已知 17 个, 候选 32 个)。以 LP 期为对照, FP 期基因上调 26 个, P30 期上调 24 个, P45 期上调 63 个; 以 FP 期为对照, P30 期基因上调 24 个, P45 期上调 82 个; 以 P30 期为对照, P45 期基因上调 47 个 (图 1)。差异

表达分析表明 LOC105612389 在 P30 期显著上调 (以 LP 期为对照), BGIG9940\_27655 在 P45 期极显著上调 (以 FP 期为对照), RPS28 在 P45 期极显著上调 (以 LP 期为对照), BGIG9940\_27548 在 P45 期极显著上调 (以 LP 期为对照), BGIG9940\_27563 在 P45 期极显著上调 (以 P30 期为对照) (表 3)。

表 2 文库测序数据信息

样品	冗余数目/Mb	非冗余片段数/Mb	平均非冗余片段数/Mb	干净读段 Q30/%	非冗余序列比对率/%	总比对率/%
FP_1	117.44	114.47		95.00	97.47	93.85
FP_2	119.94	117.99	115.25	94.51	98.37	94.65
FP_3	114.94	113.29		94.52	98.57	94.80
LP_1	117.44	115.65		94.52	98.47	94.80
LP_2	119.94	117.94	116.38	94.51	98.34	94.14
LP_3	117.44	115.56		94.52	98.40	94.70
P30_1	117.44	115.36		95.07	98.22	94.34
P30_2	119.94	117.92	117.02	94.63	98.31	94.59
P30_3	119.94	117.79		94.70	98.21	94.35
P45_1	112.44	109.28		92.23	97.18	91.49
P45_2	112.44	110.90	111.21	92.22	98.63	92.35
P45_3	114.94	113.44		92.52	98.69	92.51



LP-FP, LP-P30, LP-P45 分别表示与 LP 比较, FP、P30、P45 基因数目变化情况; FP-P30, FP-P45 分别表示与 FP 比较, P30、P45 基因数目变化情况; P30-P45 表示与 P30 比较, P45 基因数目变化情况。

图 1 不同时期差异基因数目

## 2.2 样品相关性分析

为了反映样本间基因表达的相关性, 计算了每 2 个样品之间所有基因表达量的皮尔逊相关系数, 并将

这些系数以热图的形式反映出来。由图 2 可见, 样品间相关系数平均在 0.98 以上, 样品及数据具有较高的可靠性。

表3 5个 lncRNA 差异表达比较分析

基因 ID	基因符号	FP/LP		P30/LP		P45/LP		P30/FP		P45/FP		P45/P30	
		log <sub>2</sub> 值	P 值	log <sub>2</sub> 值	P 值	log <sub>2</sub> 值	P 值	log <sub>2</sub> 值	P 值	log <sub>2</sub> 值	P 值	log <sub>2</sub> 值	P 值
105612389	LOC105612389	0.688	0.998	1.987	0.033	1.332	0.706	1.299	0.718	0.656	0.100	-0.651	0.995
BGIG9940_27655	BGIG9940_27655	0.307	0.998	-0.794	0.997	-1.192	0.680	-1.102	0.608	-1.485	0.000	-0.388	0.995
101103316	RPS28	-0.557	0.242	-0.412	0.199	-0.839	0.000	0.143	0.100	-0.268	0.820	-0.416	0.101
BGIG9940_27548	BGIG9940_27548	-0.470	0.987	-0.168	0.997	-0.818	0.001	0.301	0.100	-0.335	0.926	-0.640	0.062
BGIG9940_27563	BGIG9940_27563	0.497	0.998	0.911	0.997	-0.442	0.995	0.413	0.100	-0.925	0.252	-1.343	0.006

注: log<sub>2</sub> 值正数表示上调, 负数则表示下调。

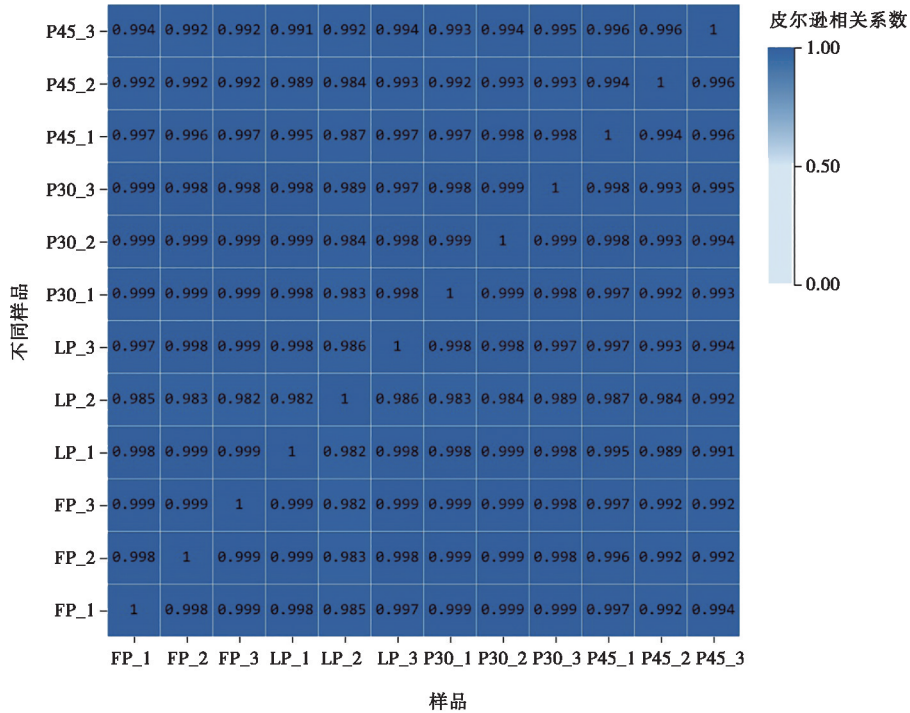


图2 样品相关性热图

### 2.3 lncRNA 靶基因预测与 GO 注释和 KEGG 通路分析

对 49 个 lncRNA 通过 cis 关系和 trans 关系判断标准, 试验获得 trans 关系靶基因 153 个, cis 关系靶基因 (上下游 100 kb) 筛选到 194 个, 合计 347 个靶基因, 其中 cis 关系和 trans 关系筛选的靶基因重复 6 个, 最终筛选到的 lncRNA 靶基因为 341 个。对 341 个靶基因做 GO 和 KEGG 通路分析。生物过程方面, 341 个靶基因富集到 25 条生物过程条目, 14 条细胞组成条目和 8 条分子功能条目, 其中排名列在前 5 的 GO 条目见表 4。通路分类发现 lncRNA 靶基因通路主要分类到 41 类通路, 其中排名列在前 10 的 KEGG 条目为: 信号转导 (signal transduction)、癌症概述 (cancers: overview)、免疫系统 (immune system)、内分泌系统 (endocrine system)、传染病: 病毒性 (infectious diseases: viral)、全局与概述图谱 (global

and overview maps)、信号分子与相互作用 (signaling molecules and interaction)、神经退行性疾病 (neurodegenerative diseases)、内分泌和代谢疾病 (endocrine and metabolic diseases)、癌症: 特殊类型 (cancers: specific types)。

结合差异表达模式, 通过 GO 和 KEGG 通路分析发现: LOC105612389 在 P30 期显著上调, 其靶基因 NOTCH1 (ID: 101115020) 参与宫内胚胎发育、肝脏发育、心内膜发育、表皮发育、内胚层发育、远端小管和集合管发育、Notch 信号通路等过程; BGIG9940\_27655 在 P45 期极显著上调, 靶基因 BCL2L11 (ID: 100505402) 参与脾脏、肾脏、雄性性腺、乳腺等器官及胚胎发育过程; RPS28 在 P45 期极显著上调, 有靶基因 BCL2L11 和 MED12, 靶基因 BCL2L11 (ID: 100505402) 参与脾脏、肾脏、雄性性腺、乳腺等器官及胚胎发育过程, 靶基因

MED12 (ID: 101123015) 参与内胚层发育、心脏发育、脊髓发育、神经管发育、胚胎器官发育、胚胎脑发育、Wnt 信号通路等过程; BGIG9940\_27548 在 P45 期极显著上调, 靶基因 ERBB3 (ID: 101103889) 参与心内膜垫发育、外周神经系统发育、颅神经发育等过程; BGIG9940\_27563 在 P45 期极显著上调, 靶基因 MED12 (ID: 101123015) 参与内胚层发育、心脏发育、脊髓发育、神经管发育、胚胎器官发育、胚胎脑发育、Wnt 信号通路等过程 (表 3, 表 5)。

## 2.4 RT-qPCR 检测

由图 3 可见, LOC105612389、RPS28、BGIG9940\_27655、BGIG9940\_27548 和 BGIG9940\_27563 的 mRNA 在未孕母羊血液和卵巢中表达量很低, 在妊娠母羊血液和卵巢中的表达量极显著升高 ( $P < 0.01$ ) (图 3)。

表 4 排名列在前 5 的 GO 条目

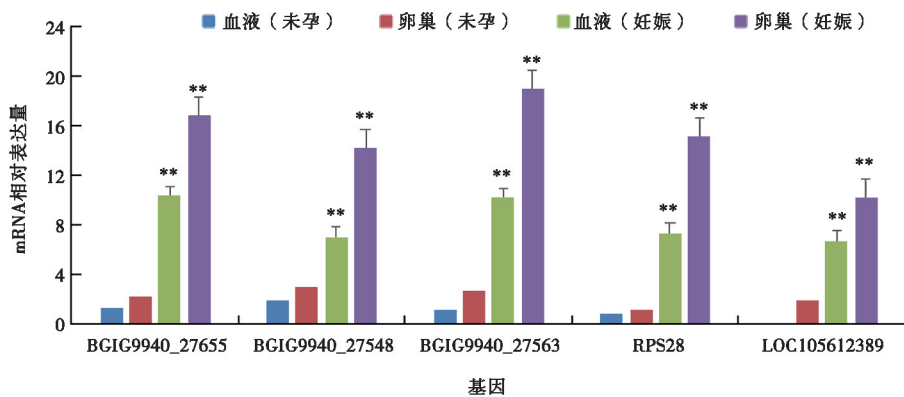
GO 水平分类 1	GO 水平分类 2	数目
生物过程	细胞过程	122
	代谢过程	91
	生物调控	80
	生物过程调控	78
	刺激反应	68
细胞组成	细胞	116
	细胞部分	115
	细胞器	99
	细胞膜	67
分子功能	细胞膜部分	58
	结合	112
	催化活性	52
	转录调节活性	15
	分子传感活性	7
	结构分子活性	7

表 5 5 个 lncRNA 靶基因部分 GO 与 KEGG 通路

lncRNA	靶基因	GO 分类	KEGG 通路
LOC105612389	NOTCH1	GO: 0001701 (宫内胚胎发育); GO: 0001889 (肝脏发育); GO: 0003157 (心内膜发育); GO: 0007219 (Notch 信号通路); GO: 0007492 (内胚层发育); GO: 0008544 (表皮发育); GO: 0021915 (神经管发育); GO: 0030324 (肺发育); GO: 0030326 (胚胎肢体形态发生); GO: 0030900 (前脑发育); GO: 0061314 (Notch 信号参与心脏发育); GO: 0072017 (远端小管发育); GO: 0072044 (集合管发育)	01522 (内分泌抵抗); 04320 (背腹轴形成); 04330 (Notch 信号通路); 04919 (甲状腺激素信号通路); 05020 (朊病毒病)
BGIG9940_27655	BCL2L11	GO: 0001701 (宫内胚胎发育); GO: 0001822 (肾脏发育); GO: 0008584 (雄性性腺发育); GO: 0009791 (胚后期发育); GO: 0030879 (乳腺发育); GO: 0035148 (管形成); GO: 0043583 (耳朵发育); GO: 0048066 (发育性色素沉着); GO: 0048536 (脾脏发育); GO: 0048538 (胸腺发育); GO: 1902263 (胚胎趾形态发生中涉及的凋亡过程)	04068 (FoxO 信号通路); 04151 (PI3K - Akt 信号通路); 04210 (细胞凋亡); 04215 (多物种细胞凋亡); 05169 (EB 病毒感染); 05200 (癌症通路)
RPS28	BCL2L11	GO: 0001701 (宫内胚胎发育); GO: 0001822 (肾脏发育); GO: 0008584 (雄性性腺发育); GO: 0009791 (胚后期发育); GO: 0030879 (乳腺发育); GO: 0035148 (管形成); GO: 0043583 (耳朵发育); GO: 0048066 (发育性色素沉着); GO: 0048536 (脾脏发育); GO: 0048538 (胸腺发育); GO: 1902263 (胚胎趾形态发生中涉及的凋亡过程)	04068 (FoxO 信号通路); 04151 (PI3K - Akt 信号通路); 04210 (细胞凋亡); 04215 (多物种细胞凋亡); 05169 (EB 病毒感染); 05200 (癌症通路)
	MED12	GO: 0001756 (体节发生); GO: 0007492 (内胚层发育); GO: 0007507 (心脏发育); GO: 0014003 (少突胶质细胞发育); GO: 0014044 (施万细胞发育); GO: 0021510 (脊髓发育); GO: 0021915 (神经管发育); GO: 0048568 (胚胎器官发育); GO: 0060070 (经典 Wnt 信号通路); GO: 1990403 (胚胎脑发育)	04919 (甲状腺激素信号通路)

续表5

lncRNA	靶基因	GO 分类	KEGG 通路
BGIG9940_27548	ERBB3	GO: 0003197 (心内膜垫发育); GO: 0007422 (外周神经系统发育); GO: 0021545 (颅神经发育); GO: 0051402 (神经元细胞凋亡过程); GO: 0055025 (心肌组织发育的正向调节); GO: 0097192 (无配体时的外在细胞凋亡信号通路)	01521 (EGFR 酪氨酸激酶抑制剂耐药性); 04010 (MAPK 信号通路); 04012 (ErbB 信号通路); 04020 (钙信号通路); 04151 (PI3K - Akt 信号通路); 05205 (癌症中的蛋白聚糖)
BGIG9940_27563	MED12	GO: 0001756 (体节发生); GO: 0007492 (内胚层发育); GO: 0007507 (心脏发育); GO: 0014003 (少突胶质细胞发育); GO: 0014044 (施万细胞发育); GO: 0021510 (脊髓发育); GO: 0021915 (神经管发育); GO: 0048568 (胚胎器官发育); GO: 0060070 (经典 Wnt 信号通路); GO: 1990403 (胚胎脑发育)	04919 (甲状腺激素信号通路)



\*\* 表示同一基因在相同组织妊娠期和未孕期相比差异极显著 ( $P < 0.01$ )。

图3 不同 lncRNA 基因在未孕和妊娠母羊血液和卵巢中表达

### 3 讨论

加快肉羊品种改良、提高肉羊繁殖效率、增加肉羊养殖数量已成为肉羊产业发展中亟需解决的突出问题。科学、准确地进行早期妊娠诊断是提高母羊繁殖效率的一项关键技术，对空怀母羊进行及时再孕处理，是提高母羊繁殖效率的一项重要措施。

高灵敏度的检验技术和特异性的检验标志物对于母畜早期妊娠检测具有重要的意义。转录组、蛋白组等高通量测序技术目前已成为筛选小分子生物标志物的一种重要方式<sup>[14-17]</sup>。Zul 等<sup>[18]</sup>利用高通量测序技术对临床乳房炎组、亚临床乳房炎组及健康乳牛组血液样本进行测序，经过研究证实白细胞介素 (IL-8) 和结合珠蛋白能作为早期检测隐性乳房炎的潜在生物标志物。本试验运用 Illumina 高通量测序技术分析卵泡期、黄体期、妊娠 30 d 和妊娠 45 d 绵羊卵巢 mRNA 和 lncRNA 表达，探讨 lncRNA 在绵羊妊娠早期中的调控作用。试验成功构建绵羊妊娠早期 RNA 文库，每个样品平均产出 11.50 G 数据，样品比对基

因组的平均比对率为 93.88%。吴垚群等<sup>[5]</sup>测序量为 10 G，平均比对率为 81.22%；李耀坤等<sup>[8]</sup>测序量为 10.4 G，平均比对率在 82% 以上。本试验数据产出率均高于二者的测序结果，说明测序数据质量可靠，为后续试验的深入分析提供了保障。该试验检测到 23 219 个基因，其中 418 个差异 mRNA 和 49 个差异 lncRNA。差异表达分析发现在动物不同生理期，基因表达模式明显不同，试验结果进一步证实了基因参与动物生殖生理活动的调控，与一些研究相一致<sup>[5-6,8-9,14]</sup>。Forde 等<sup>[19]</sup>利用蛋白质组学与转录组学分析发现 ACO2 等基因在妊娠期母牛、非妊娠期母牛体内表达差异显著，推测 ACO2 等基因具有作为妊娠鉴定的潜在价值；黄芳<sup>[20]</sup>发现，胚胎停育患者与正常早孕女性蜕膜中 Wnt4 表达水平差异显著，认为在子宫内膜蜕膜化及维持妊娠中 Wnt4 发挥重要作用，可维持胚胎发育。本研究分析发现其中 LOC105612389 在 P30 期显著上调，BGIG9940\_27655、RPS28、BGIG9940\_27548、BGIG9940\_27563 在 P45 期极显著上调，5 个 lncRNA 在妊娠期表达显

著升高,说明其可能参与了早期胎儿组织器官的发育,为进一步成为候选早期妊娠诊断生物标志物提供了支撑。

对 49 个差异 lncRNA 通过 cis 和 trans 关系分析获得 341 个靶基因,对 341 个靶基因做 GO 和 KEGG 通路分析。通过 GO 和 KEGG 通路分析结合差异表达分析发现,LOC105612389 在妊娠期显著上调,其靶基因 NOTCH1 参与宫内胚胎发育、肝脏发育、心内膜发育、表皮发育、内胚层发育、远端小管和集合管发育、Notch 信号通路等过程。蒋思涵等<sup>[21]</sup>研究证实 Notch1 为性别特征基因,与蝗卵发育密切相关,在胚胎早期发育阶段参与诱导性别向雄性发育;王雪淞等<sup>[22]</sup>研究证实 Notch 信号通路具有调控间充质干细胞分化成不同靶细胞的功能,可见本论文研究结果与上述研究结果相一致。GO 和 KEGG 通路分析发现靶基因 BCL2L11 参与脾脏、肾脏、雄性性腺、乳腺等器官及胚胎发育过程;靶基因 MED12 参与内胚层发育、心脏发育、脊髓发育、神经管发育、胚胎器官发育、胚胎脑发育、Wnt 信号通路等过程;靶基因 ERBB3 参与心内膜垫发育、外周神经系统发育、颅神经发育等过程。刘玉芳等<sup>[23]</sup>研究证实, BCL2L11 是调控卵巢发育的重要基因及调控元件,尤其在胚胎干细胞分化方面发挥重要作用,而 Notch 信号通路、Wnt 信号通路等都是参与动物生殖生理调控的重要通路<sup>[24-25]</sup>。杨帆等<sup>[26]</sup>研究发现,多个差异 lncRNA 均靶向 Wnt 通路相关基因,提示 lncRNA 可能通过调控 Wnt 通路关键基因来影响绵羊繁殖力。本试验数据及一些研究<sup>[23-25]</sup>分析证实 LOC105612389、BGIG9940\_27655、RPS28、BGIG9940\_27548、BGIG9940\_27563 基因均参与胚胎发育、生殖生理等相关过程的调控。RT-qPCR 结果表明 LOC105612389、BGIG9940\_27655、RPS28、BGIG9940\_27548 和 BGIG9940\_27563 基因在未孕母羊血液和卵巢中表达量很低,在妊娠母羊的血液和卵巢中表达量显著升高,通过测序数据和荧光定量数据分析 LOC105612389、BGIG9940\_27655、RPS28、BGIG9940\_27548 和 BGIG9940\_27563 在血液中的表达量达某一临界值时,预示着动物的妊娠。5 个 lncRNA 的靶基因均参与早期胚胎各组织器官发育过程,结合在妊娠和未妊娠母羊表达模式的不同可推测该 5 个 lncRNA (LOC105612389、BGIG9940\_27655、RPS28、BGIG9940\_27548 和 BGIG9940\_27563) 有作为母羊早期妊娠诊断生物标志物的可行性,但该试验妊娠血清样品、卵巢样品仅有 30、10 个,样品数量及测定范围还有待进一步增加及细化以验证其作为生物标志物的精确度和准确度。microRNA 测序分析发现,妊娠早期牛血浆中 miR-26a 显著上调,可作为早期

妊娠的潜在标志物<sup>[27-28]</sup>,但其在妊娠诊断中的应用有待进一步研究; bta-miR-221 和 bta-miR-223 等 6 个 miRNAs 在妊娠早期鲜奶样品中持续上调表达,但是因不能有效区分妊娠牛和发情牛,导致其不能用于奶牛的早期妊娠检测<sup>[29]</sup>。理想的早期妊娠诊断生物标志物应能及时准确反映母畜妊娠状态,在特定妊娠阶段呈现特异性上调或下调,受日粮、环境以及药物等非动物性因素的影响小,能够反应孕体的年龄和生存能力,可以在易获得的体液如血液、血清、尿液中检测,而且要有持续的表达周期以提供足够的样品采集时间<sup>[30]</sup>。血液/血清作为动物共有的体液样本比较容易采集,在临床生物标志物检测中应用广泛<sup>[31]</sup>。5 个 lncRNA (LOC105612389、BGIG9940\_27655、RPS28、BGIG9940\_27548 和 BGIG9940\_27563) 是通过高通量测序,结合生物信息学并通过临床血液定量表达模式的验证分析获得的具有早期妊娠诊断潜力的重要小分子生物,其中 LOC105612389、BGIG9940\_27655、BGIG9940\_27548 和 BGIG9940\_27563 在数据库中均未查到相关文献,属首次报道;但是,由于动物在品种、胎次、样本类型以及采集、处理与检测方法、人员操作等方面的差异,5 个 lncRNA 作为候选早期妊娠检测生物标志物的特异性和准确性还有待进一步研究。

## 4 结论

试验通过高通量测序技术成功构建了绵羊 RNA 文库,结合生物信息学分析共鉴定出 418 个差异 mRNA 和 49 个差异 lncRNA,对 49 个差异 lncRNA 通过 cis 和 trans 关系,分析获得 341 个靶基因;5 个 lncRNA (LOC105612389、BGIG9940\_27655、RPS28、BGIG9940\_27548 和 BGIG9940\_27563) 在妊娠期表达量显著升高;通过靶基因、GO 和 KEGG 通路分析发现,5 个 lncRNA 均可能参与早期胎儿发育及调控过程;通过临床表达模式研究发现,该 5 个基因表达量达到一定临界值时有望作为早期妊娠检查的候选生物标志物。综上,本研究对绵羊早期妊娠诊断具有十分重要的意义。

## 参考文献:

- [1] 蒋汶君,唐曲,顾锡娟,等.基于微流控技术的疾病标志物即时检测研究进展[J].分析化学,2023,51(9):1403-1413.
- [2] 靳和玥,张怡敏,陶芳标,等.早产预测的血液生物标志物研究进展[J].中国生育健康杂志,2023,34(4):373-376.
- [3] MARKKANDAN K, AHN K, LEE D J, et al. Profiling and identification of pregnancy-associated circulating microRNAs in dairy cattle [J]. Genes Genomics, 2018, 40 (10): 1111-1117.

- [4] SAED G M, NAWAZ A, ALVERO A A, et al. Monomeric myeloperoxidase is a specific biomarker for early-stage ovarian cancer [J]. *Biomarkers*, 2023, 28 (7): 663-671.
- [5] 吴垚群, 陈少康, 盛熙晖, 等. 用高通量测序技术研究松辽黑猪与长白猪背最长肌 mRNA 和 lncRNA 的差异表达 [J]. *中国农业科学*, 2020, 53 (4): 836-847.
- [6] CHANG W H, WANG J H, CUI Z L. Integrative transcriptomic and proteomic analysis of ovaries at different physiological periods in Dolang Sheep [J]. *Indian Journal of Animal Research*, 2022, 56 (12): 1492-1498.
- [7] 崔子龙, 常倩倩, 杨学亮, 等. 基于 iTRAQ 蛋白定量技术对多浪羊不同生理时期卵巢候选蛋白生物标志物的筛选 [J]. *基因组学与应用生物学*, 2021, 40 (4): 1522-1530.
- [8] 李耀坤, 许香萍, 陆婷婷, 等. 影响山羊产羔数 lncRNA 的筛选及功能分析 [J]. *畜牧兽医学报*, 2020, 51 (8): 1853-1865.
- [9] WANG C, ZHAO Y, YUAN Z, et al. Genome-wide identification of mRNAs, lncRNAs, and proteins, and their relationship with sheep fecundity [J]. *Front Genet*, 2022, 12: 750947.
- [10] 彭星, 杨孟欢, 马海明, 等. 反义长链非编码 RNA 调控机制及其在畜禽生理过程中的作用研究进展 [J]. *中国畜牧杂志*, 2023, 59 (11): 53-59.
- [11] 赵念念, 冯榕, 张丽, 等. lncRNA/hnRNP 在调控基因表达中的机制及功能 [J]. *生命的化学*, 2020, 40 (10): 1815-1824.
- [12] 赵小萱, 冯晓玲, 陈璐, 等. 长链非编码 RNA 与复发性流产及反复着床失败关系的研究进展 [J]. *中国计划生育学杂志*, 2020, 28 (11): 1929-1934.
- [13] SMOLARZ B, ZADROZNA-NOWAK A, ROMANOWICZ H. The role of lncRNA in the development of tumors, including breast cancer [J]. *International Journal of Molecular Sciences*, 2021, 22 (16): 8427.
- [14] GORDON-RODRIGUEZ E, QUINN T P, CUNNINGHAM J P. Learning sparse log-ratios for high-throughput sequencing data [J]. *Bioinformatics*, 2021, 38 (1): 157-163.
- [15] MANN M, KUMAR C, ZENG W F, et al. Artificial intelligence for proteomics and biomarker discovery [J]. *Cell Systems*, 2021, 12 (8): 759-770.
- [16] ZHONG Y, XU F, WU J, et al. Application of next generation sequencing in laboratory medicine [J]. *Annals of Laboratory Medicine*, 2021, 41 (1): 25-43.
- [17] 李志超, 王磊, 薛景才, 等. 转录组分析鉴定腰椎间盘退行性变中潜在免疫相关生物标志物 [J]. *脊柱外科杂志*, 2023, 21 (5): 316-325.
- [18] ZUL I H, NEELESH S, SARABPREET K, et al. Putative biomarkers for early detection of mastitis in cattle [J]. *Animal Production Science*, 2020, 60 (14): 1721-1736.
- [19] FORDE N, BAZER F W, SPENCER T E, et al. Conceptualizing the endometrium: identification of conceptus-derived proteins during early pregnancy in cattle [J]. *Biology of Reproduction*, 2015, 92 (6): 156.
- [20] 黄芳. IGF1F-1、Wnt4 和 Bmp2 在早孕妊娠蜕膜中的表达及其与高龄、胚胎停育的关系 [D]. 杭州: 浙江大学, 2018.
- [21] 蒋思涵, 赵娜, 宋余, 等. Ywhae 与 Notch1 在意大利蝗成虫不同组织和不同发育阶段蝗卵的表达分析 [J]. *应用昆虫学报*, 2023, 60 (5): 1358-1365.
- [22] 王雪凇, 周林, 李林材, 等. Notch 信号通路调控间充质干细胞的增殖与分化 [J]. *中国组织工程研究*, 2024, 28 (19): 3076-3083.
- [23] 刘玉芳, 陈玉林, 周祖阳, 等. miR-221-3p 靶向 BCL2L11 调控小尾寒羊卵泡颗粒细胞凋亡 [J]. *中国农业科学*, 2022, 55 (9): 1868-1876.
- [24] 荣轩, 邵顺成, 梁鹏, 等. miRNA 调控绵羊排卵过程的研究进展 [J]. *家畜生态学报*, 2023 (4): 1-6.
- [25] 吴馨媛, 袁源智. Wnt/ $\beta$ -catenin 信号通路在眼部病理性血管新生过程中作用的研究进展 [J]. *中国临床医学*, 2023, 30 (6): 1067-1073.
- [26] 杨帆, 姚晓磊, 李康, 等. 不同绵羊品种发情期卵巢组织 lncRNA 和 mRNA 的差异表达分析 [J]. *南京农业大学学报*, 2023, 46 (6): 1142-1152.
- [27] IOANNIDIS J, DONADEU F X. Circulating miRNA signatures of early pregnancy in cattle [J]. *BMC Genomics*, 2016, 17: 184.
- [28] IOANNIDIS J, DONADEU F X. Changes in circulating microRNA levels can be identified as early as day 8 of pregnancy in cattle [J]. *PLoS One*, 2017, 12 (4): e0174892.
- [29] SCHANZENBACH C I, KIRCHNER B, UIBRICH S E, et al. MicroRNA of whole milk samples are not suitable for pregnancy detection in cattle [J]. *Gene*, 2019, 692: 17-21.
- [30] BAIHARA A K, GUPTA M, SINGH S, et al. Early pregnancy diagnosis in bovines: current status and future directions [J]. *Scientific World Journal*, 2013, 2013: 958540.
- [31] 鸣易, 王晨辉. 疾病早期诊断的新技术和新标志物 [J]. *实用医院临床杂志*, 2023, 20 (5): 1-3.