



黄喉拟水龟性别调控相关 lncRNA 和 mRNA 的筛选及初步分析

朱彦雨^{1,2}, 刘晓莉¹, 王亚坤¹, 李伟¹,
洪孝友¹, 朱新平^{1,2*}, 徐红艳^{1,2*}

(1. 中国水产科学研究院珠江水产研究所, 农业农村部热带亚热带水产资源利用与
养殖重点实验室, 广东 广州 510380;
2. 上海海洋大学, 水产科学国家级实验教学示范中心, 上海 201306)

摘要: 性别异形在动物界广泛存在, 长期以来一直是生物学中的热门话题。黄喉拟水龟性别分化属温度依赖型, 其温度响应分子机制仍不清楚。为了进一步解析龟鳖动物性别分化的分子机制, 本研究初步比较分析了黄喉拟水龟转录本中性别差异表达 (different expressed, DE) 的长链非编码 RNA (long noncoding RNA, lncRNA) 及其调控的靶基因。首先, 运用 Illumina 深度测序平台, 通过转录组测序 (RNA-Seqs) 对黄喉拟水龟的精巢和卵巢进行了比较转录组学分析并筛选鉴定出雌雄差异转录本, 共筛选获得 8 237 个 DE mRNA 和 9 573 个 DE lncRNA。通过 GO 功能注释及 KEGG 通路富集分析发现, 上述差异转录本主要参与龟鳖动物性别分化和性腺发育等相关信号通路。此外, 通过顺式和反式作用分析, 筛选获得了一系列与生殖发育相关的 (性腺发育和性别分化) 受 DE lncRNAs 调控的靶基因。本研究为进一步阐明黄喉拟水龟温度依赖型性别的分子机制提供了线索, 特别是为进一步挖掘利用龟鳖动物性别决定因子, 进行龟鳖动物性控育种技术研究奠定了基础。

关键词: 黄喉拟水龟; 转录组; 长链非编码 RNA; 性别决定; 性别分化

中图分类号: Q 785; S 917.4

文献标志码: A

性别差异和性别决定的遗传基础及其机制研究一直是生命科学研究中最具吸引力和最热门的领域之一。大多数脊椎动物都是雌雄异体, 在形态、结构、细胞组成、行为和 life 史上存在显著的性别差异, 从而形成了生命科学的多样化和多彩的性别世界^[1]。遗传信息和环境是决定大多数脊椎动物性别的 2 个主要因素, 这 2 种模式分别被称为遗传型性别决定机制 (genetic sex

determination, GSD) 和环境型性别决定机制 (environmental sex determination, ESD)^[2]。在一些 GSD 物种中, 如一些哺乳动物、鸟类、多数爬行动物、两栖动物和鱼类, 它们的性别在受精时就由一些具有性别差异的遗传信息所触发。相比之下, ESD 物种 (多为爬行类、两栖类和鱼类) 则是依靠一些环境因素, 例如温度、氧气、光周期、pH 和 CO₂ 等来驱动物种的性别决定^[3]。

收稿日期: 2020-03-20 修回日期: 2020-05-09

资助项目: 国家重点研发计划 (2018YFD0900201, 2018YFD0900203); 国家自然科学基金 (31873035); 广州市科技计划项目 (201904010172); 社会公益研究 (2019HY-XKQ02); 社会公益研究 (2018SJ-YZ02); 中国—东盟海上合作基金资助 (CAMC-2018F); “一带一路”沿线国家热带农业科技创新合作 (2130114)

通信作者: 朱新平, E-mail: zhuxinping_1964@163.com; 徐红艳, E-mail: xuhyzqh@163.com

黄喉拟水龟 (*Mauremys mutica*) 是我国重要的大宗淡水养殖龟类, 广泛分布在我国华东、华南、越南北部和日本^[4]。前期研究证实, 黄喉拟水龟的性别不是由遗传因素决定的, 主要受温度影响, 较高的孵化温度会增加后代雌性的比例^[5], 具有温度依赖型性别决定 (temperature-dependent sex determination, TSD) 模式。然而, 目前关于黄喉拟水龟性别决定、性别分化及配子发生的生殖过程的调控机制仍不清楚。因此, 对黄喉拟水龟生殖过程的分子机制, 尤其是与性别决定相关基因和调控途径的研究, 有助于了解其他龟鳖类乃至 ESD 物种的性别决定机制。

近年来, 随着分子遗传生物学及其技术的发展, 已经发现了一些 GSD 物种中的性别决定基因, 例如 *sry*、*dmy*、*dmw*、*dmrt1*、*amhy*、*gsdfy*、*amhr II*、*sdv* 和 *sox3* 具有临界潜在性, 可以指示性腺向精巢或卵巢发育^[6]。除此之外, 在 ESD 物种中, 性别决定机制的相关研究也取得了一些重要进展。例如, 在红耳龟 (*Trachemys scripta elegans*) 中, *dmrt1* 在性别分化过程中表现出温度依赖性和性别二态性表达模式, 揭示了 *dmrt1* 是该物种中雄性性别决定基因的候选基因^[7]。在 TSD 物种蛇龟 (*Chelydra serpentina*) 中, *cirbp* (cold inducible RNA binding protein) 能够调节温度对发育中的性腺的影响, 这就表明 *cirbp* 可以参与调节性腺的分化^[8]。但是, 目前尚不清楚这些基因是否在黄喉拟水龟两性中差异表达以及性别决定中起重要作用。

长链非编码 RNA (long noncoding RNA, lncRNA) 是长度超过 200 个核苷酸的非编码转录本, 自从 *xist* lncRNA 被发现是功能性 RNA 后, 越来越多的 lncRNA 被发现在各种生物中发挥着重要作用^[9]。特别是最近的一些研究表明, lncRNA 参与小鼠 (*Mus musculus*) 的胚胎发育和配子发生^[10], 长牡蛎 (*Crassostrea gigas*) 的性别决定和性腺发育^[11] 及拟穴青蟹 (*Scylla paramamosain*) 的性别分化^[12]。尽管近年来 lncRNA 一直是动物界的热门话题, 但目前人们对龟鳖动物中性腺发育相关 lncRNAs 的研究分析还极为罕见。

在本研究中, 我们首先利用 Illumina HiSeq™ 4 000 对黄喉拟水龟雌雄性腺 lncRNAs 和 mRNAs 的表达谱进行了比较分析, 通过基因的雌雄差异表达分析, 筛选获得黄喉拟水龟性别分化相

关的关键 lncRNAs 分子, 并进行了相关靶标基因的预测分析。本研究为探寻黄喉拟水龟性别分化和性腺发育的分子机制提供遗传基础, 也为深入理解龟鳖类性别决定机制的演化提供基础数据。

1 材料与方法

1.1 样品的采集和制备

实验所用健康成熟的黄喉拟水龟采自中国水产科学研究院珠江水产研究所龟类繁育场, 麻醉放血后, 通过解剖性腺确定其性别, 取样收集卵巢和精巢各 9 个, 置于液氮速冻, 然后保存于 -80 °C, 用于 RNA 的提取。特此声明, 本研究中所有动物实验均按照中国水产科学研究院珠江水产研究所制定的许可指南进行, 所有取样用的黄喉拟水龟都受到人道对待。

1.2 RNA 提取和 RNA-Seq 文库制备

使用 SV Total RNA Isolation System (Promega, 美国) 从性腺组织中提取总 RNA, 随后以剔除核糖体 RNA (rRNA) 的 RNA 作为模板和随机六聚体引物通过反转录合成第一链 cDNA, 再使用 RNase H 和 DNA 聚合酶 I 进行第二链 cDNA 合成。其余的黏性末端通过核酸外切酶/聚合酶活性转化为平末端。使用 Phusion 高保真 DNA 聚合酶, Universal PCR 引物和 Index(X) 引物进行 PCR 后, 在 Agilent Bioanalyzer 2100 上通过实时荧光定量 PCR (qRT-PCR) 评估最终文库质量, 然后在 Illumina HiSeq 4000™ 系统中进行测序。

1.3 质量控制和 lncRNA 分析

利用 String Tie 程序对比对上的序列 (reads) 进行组装, 以黄喉拟水龟基因组作为参考基因组进行序列比对和后续分析, 利用 gffcompare 程序组装注释新的转录本, 使用未知的转录本筛选挖掘出新的 lncRNAs。结合了 4 种计算方法, 包括编码潜能计算器 (coding potential calculator, CPC)^[13]、编码非编码索引 (coded non-coded index, CNCI)^[14]、编码潜能评估工具 (coding potential assessment tool, CPAT)^[15] 和蛋白质家族数据库 (protein family database, PFAM)^[16], 从未知转录本中推定出来的编码蛋白 RNA 中筛选出非编码蛋白 RNA 候选对象。再用 String Tie 程序计算每个

文库中的 lncRNAs 和编码基因的 FPKM 值 (每千个碱基的转录每百万映射读取的片段), 其可作为衡量转录本或基因表达水平的指标, 将每个基因转录本的 FPKM 加起来计算基因 FPKM^[17]。用 lncTar 靶基因预测工具对 lncRNA 进行靶基因预测, 利用 lncRNA 和 mRNA 之间的互补序列计算配对位点自由能和标准化自由能。标准化自由能阈值低于 -0.1 被认为是 lncRNA 反式作用的靶基因^[18]。通过 Cytoscape 软件 (Cytoscape, 美国) 对差异表达的 lncRNAs 和 mRNAs 进行相关分析, 以鉴定二者之间的相互作用^[19]。

1.4 差异表达分析和基因功能注释

基因表达具有时间和空间特异性, 外部刺激和内部环境都会影响基因的表达。此实验设置 3 个生物学重复, 使得寻找差异表达基因和差异表达转录本更具可靠性。对有生物学重复的样本选择使用 DE Seq R 软件来分析雌雄性腺之间的差异表达基因 (differentially expressed genes, DEG) 和 DE lncRNA。DE Seq 软件基于负二项分布模型来确定 2 个生物学条件之间的差异表达基因。结合整体情况, 我们将差异倍数 | Fold Change | ≥ 2 且错误发现率 (false discovery rate, FDR) < 0.01 作为筛选 mRNA 和 lncRNA 差异表达的标准^[20]。此外, 为了注释 DEG 功能, 基于 COG、GO、KEGG、KOG、Swiss-Prot 和 NR 数据库对基因进行了比对。利用 top GO R 软件包对 DEG 进行 GO 富集分析, 用 KOBAS 软件来进行 KEGG

通路富集分析^[21]。

1.5 DE mRNAs 的 qRT-PCR 分析

使用 10 \times SYBR Green Master Mix (Promega) 通过 qRT-PCR 对随机选择 10 个 mRNA 进行分析 (表 1)。使用 *gapdh* 基因作为内参基因来计算转录物的相对表达, 使用 $2^{-\Delta\Delta C_t}$ 方法分析转录物表达水平^[22], 每组实验设置 3 个平行重复。表 1 为所选 mRNAs 和内参的引物序列。

2 结果

2.1 RNA-seq 测序和序列组装

对取样得到的精巢和卵巢提取总 RNA, 通过 Illumina HiSeqTM4000 平台进行全转录组分析, 设置 3 个生物学重复。总共获得 956 208 280 个高质量可用序列 (clear reads), 卵巢中获得 465 914 500 个 reads, 精巢中获得 490 293 780 个 reads, 除去一些低质量的 reads 之后, 总共筛选到 76 545 个 mRNAs (图 1-a), 包括 17 892 (23.378%) 个已知转录本和 58 653 (76.63%) 个新的转录本。此外, RNA-seqs 筛选到了 148 072 个 lncRNA (图 1-a), 其中包括 71% 的长链基因间非编码 RNA (long intergenic ncRNAs, lincRNA), 7.4% 的反义长链非编码 RNA (antisense-lncRNA), 20% 的内含子长链非编码 RNA (intronic-lncRNA) 和 1.5% 的正义长链非编码 RNA (sense-lncRNA) (图 2)。新鉴定的 lncRNAs 比率远高于已知的 lncRNAs。mRNAs 的长度为 75~

表 1 实验所用引物及序列

Tab. 1 Primers and sequences used in this experiment

类别	feature	基因	gene name	正向引物	forward primer	反向引物	reverse primer
mRNA		<i>ddx25</i>		TAACCATTTGGCATTCTACGAGT		CAGATAATAATTGGAACCCCCG	
		<i>tdrd9</i>		GCCACCTCCTCTACCCCTACCG		CACTATTGATGCCTCTTCTGACCA	
		<i>nr5a2</i>		CTTGAAGTGCAGCCAGATGA		TGCCCAGTAACCAGAAGGATAGA	
		<i>vasa</i>		TATCACCTGTAGGACCTCCAAACC		GGAAGAGCAAATGATGACCCCT	
		<i>piwil1</i>		TGACTGATAAGATGCGTAGCGATT		TTGCCTTGTTTCATTTGGATGC	
		<i>piwil2</i>		AGAAATAACAAGAGAAGCCGCCA		TCCCCAGAGTTCCCCACCTA	
		<i>rad51a</i>		CTGCTGCTGCTAAAGCCAAAG		CAAACATTCAAATCCCTCAAACA	
		<i>tdrd1</i>		AACAAGGCGAATGGAACAAAGA		CTCCCACGGATACGGAAACA	
		<i>tdrd7</i>		GTTTCGGTAAGGGTAGAGGAGGT		GATGCTGTATTGTGGCTGAGAGAT	
		<i>stra6</i>		GCTCCCTTGCTTACGATGACTT		GATTCCAAACACTGCTGCCAC	
	内参基因	internal reference	<i>gapdh</i>		CCGTGTTCCAACCTCCAATG		GGCAGCCTTCATCACCTTCTT

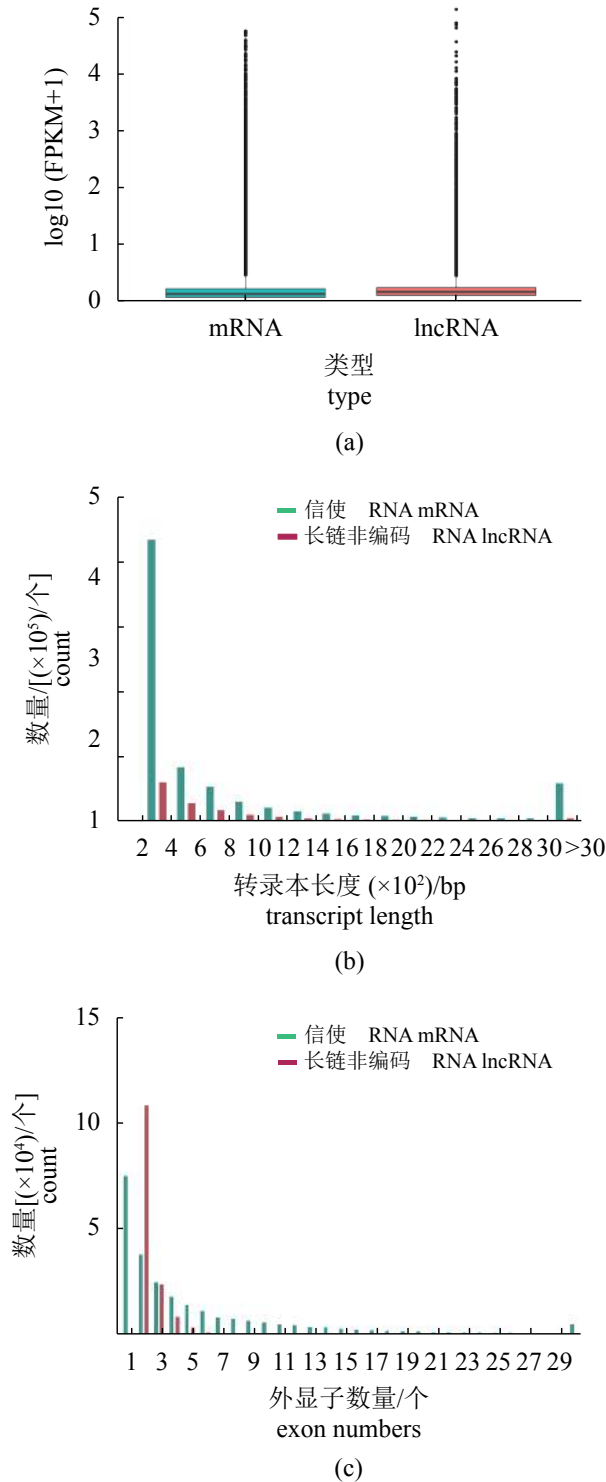


图 1 预测 lncRNA 和 mRNA 的特征比较

(a) lncRNA 和 mRNA 的表达, (b) 预测的 lncRNAs 和 mRNAs 的长度分布, (c) lncRNAs 和 mRNAs 的外显子数目分布

Fig. 1 Comparison of features of predicted lncRNAs and mRNAs

(a) expression of lncRNA and mRNA, (b) length distribution of predicted lncRNAs and mRNAs, (c) exon number distribution of lncRNAs and mRNAs

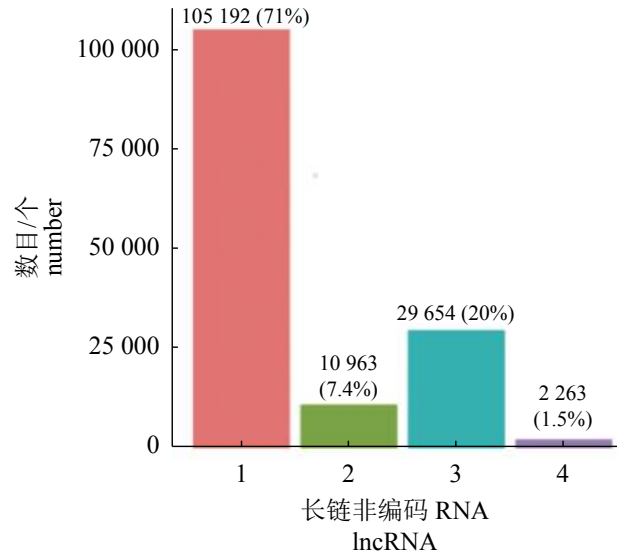


图 2 预测的 lncRNAs 的统计图

1. 长链基因间非编码 RNAs, 2. 反义长链非编码 RNAs, 3. 内含子长链非编码 RNAs, 4. 正义长链非编码 RNAs; 括号内的数字代表其所占总 lncRNA 的比例

Fig. 2 Statistics map of predicted long non-coding RNAs

1. long intergenic ncRNAs(lincRNAs), 2. antisense-lncRNAs, 3. intronic-lncRNAs, 4. sense-lncRNAs; the number in brackets indicates the proportion

94 038 bp, 平均长度为 959 bp; lncRNAs 的长度为 201~40 653 bp, 平均为 800 bp(图 1-b)。mRNAs 的平均外显子数量为 2.74, lncRNAs 的平均外显子数量为 2.47(图 1-c)。

2.2 雌雄之间差异表达基因的鉴定

根据 Illumina HiSeq™4000 平台所得的转录组测序结果, 通过倍数变化过滤 (Fold Change ≥ 2, FDR < 0.01) 和层次聚类来鉴定差异表达的 RNA, 最后共发现了 8 237 个 DE mRNA 为性别偏向基因, 包括 5 249 个雄性偏向基因和 2 988 个雌性偏向基因。其中, 有 4 237 个 mRNA 雌雄差异表达超过 10 倍。同时, 还发现 9 573 个 lncRNA 性别偏向表达, 包括 7 597 个雄性偏向 lncRNA 和 1 976 个雌性偏向 lncRNA, 其中 9 444 个 lncRNA 的雌雄差异表达超过 10 倍。此外, 相比于精巢中的差异转录本, 卵巢中分别有 5 249 和 2 988 个 mRNA 上调和下调, 有 7 597 和 1 976 个 lncRNA 转录本上调和下调(图 3-a)。火山图展示了 lncRNAs 和 mRNAs 的相似差异表达情况(图 3-b)。

为了验证测序结果的可靠性, 本实验随机选取了 10 个差异表达基因进行 qRT-PCR 验证。

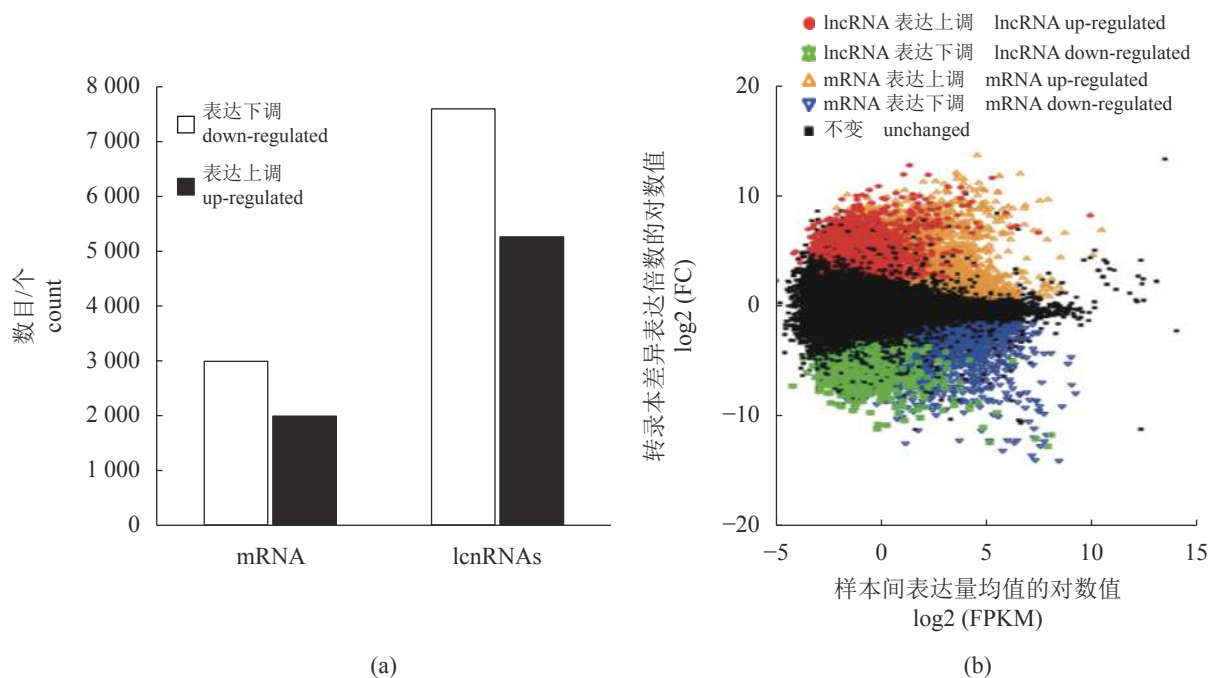


图3 卵巢和精巢中差异表达 lncRNAs 和 mRNAs 的分析

(a) 卵巢相对于精巢鉴定出的上调和下调差异表达 lncRNAs 和 mRNAs 的数量, (b) 差异表达的 lncRNAs 和 mRNAs 的火山图

Fig. 3 Analyses of differentially expressed lncRNAs and mRNAs in ovaries and testes

(a) numbers of up- and down-regulated differentially expressed lncRNAs and coding transcripts in ovary compared with testis, (b) volcano plots of differentially expressed lncRNAs and mRNAs

结果显示, 这些性别二态性基因的表达情况(图 4-a)与转录组测序分析结果(图 4-b)基本一致。*ddx25*、*tdrd9*、*nr5a2*、*vasa*、*piwil1*、*piwil2*、*rad51a* 和 *tdrd1* 在精巢中的表达水平较高, 而 *stra6* 在卵巢中表达较高。

2.3 雌雄 DEGs 的 GO 注释

对于 mRNA 转录本, 总共注释了 75.59% (8 237 个中的 6 226 个)的 DEG(表 2)。从 NR 数据库得到的注释基因数量最多, 达到 6 200 (9.58%) 个, 除此之外, 在 COG、GO、KEGG、KOG 及 Swiss-Prot 数据库分别注释得到 1 307 (20.99%)、4 106 (65.95%)、2 949 (47.37%)、3 288 (52.81%)、3 039 (48.81%) 个 DEG。这 4 106 个 DEG 在 GO 注释中分为 3 个大类别, 60 个子类别。3 大类中的优势子类别分别为生物学过程中的细胞过程 (GO: 0009987), 细胞组成中的细胞 (GO: 0005623) 和分子功能中的结合 (GO: 0005488)(图 5-a)。

随后, 我们利用 GO 聚类分析进一步探究了上述 DEGs 的生物学功能, 雌雄差异表达的 4 106 个 DEG 中, 有 35 个 DEG 与髓鞘有关, 其中 18 个在卵巢中高表达 (female-highly expressed gene,

FEG), 17 个在精巢中高表达 (male-highly expressed gene, MEG)。有 25 个 DEG 与甲基化有关, 其中 12 个在卵巢中高表达, 13 个在精巢中高表达。还有 4 个 FEG 和 18 个 MEG 与有机环状化合物合成有关(图 5-b)。

2.4 雌雄 DEGs 的 KEGG 通路富集分析

2 949 个 DEG 经过 KEGG 富集分析, 表现为 241 个差异显著的代谢途径, 其中一些与性腺发育有关(图 6-a)。显著富集的 KEGG 通路依次为“胞吞作用”(65 个 DEG), “肌动蛋白细胞骨架的调节”(60 个 DEG), “神经活性配体受体相互作用”(58 个 DEG) 和“吞噬体”(57 个 DEG)(图 6-a)。实验选择了富集显著性最可靠 (即 Q -value 最小) 的前 20 个通路进行分析, 结果表明, 前 3 个途径与“卵母细胞减数分裂”、“p53 信号传导途径”和“孕激素介导的卵母细胞”相关。同样, 还发现 4 种与性别调控有关的途径, 例如“肌动蛋白细胞骨架的调节”、“TGF-beta 信号通路”、“紧密连接”和“细胞周期”(图 6-b)。总共 316 个 DEG (168 个 FEG 和 148 个 MEG) 参与了这 20 个最主要的 KEGG 通路途径(图 6-b)。

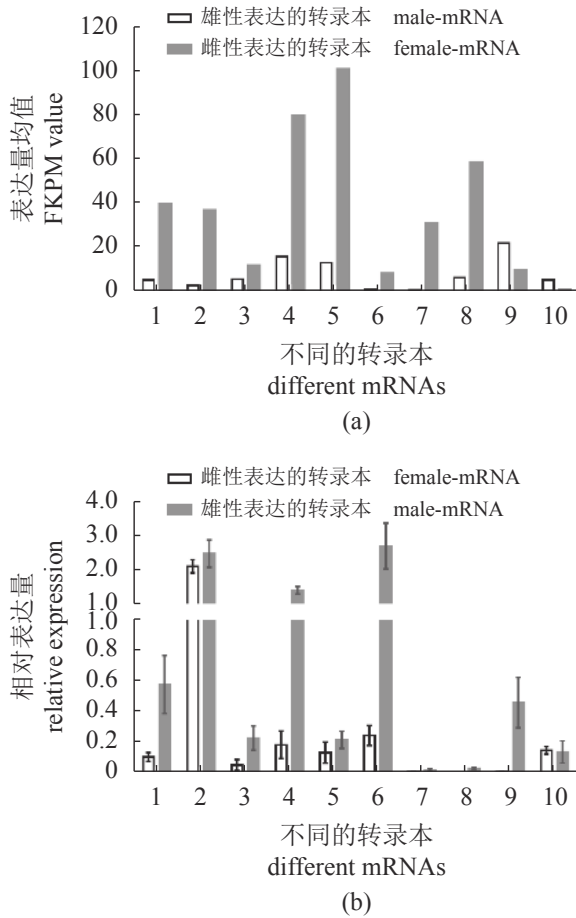


图 4 通过 qRT-PCR 验证 mRNAs 的表达

(a) 随机挑选的 mRNAs 表达量均值 (FPKM 值), (b) 通过 qRT-PCR 的随机挑选的 mRNAs 的相对表达, 表达量值表示为平均值±标准误差 (SE), 1. *ddx25*, 2. *tdrd9*, 3. *nr5a2*, 4. *vasa*, 5. *piwil1*, 6. *piwil2*, 7. *rad51a*, 8. *tdrd1*, 9. *tdrd7*, 10. *stra6*

Fig. 4 Verification of the mRNAs by qRT-PCR

(a) mean value of expression (FPKM value) of randomly selected mRNAs, (b) relative expression of randomly selected mRNAs by qRT-PCR. Values are presented as means±standard error (SE), 1. *ddx25*, 2. *tdrd9*, 3. *nr5a2*, 4. *vasa*, 5. *piwil1*, 6. *piwil2*, 7. *rad51a*, 8. *tdrd1*, 9. *tdrd7*, 10. *stra6*

2.5 DE lncRNAs 的靶基因预测

为了研究 DE lncRNAs 在黄喉拟水龟性别分化中的作用, 我们筛选了 DE lncRNAs 的直接或间接靶基因, 鉴定出 4 630 个顺式靶基因 (4 609

个被注释); 2 986 个反式靶基因 (2 965 个被注释)(表 3)。预测了参与性腺发育、配子形成、性别决定与分化等生殖过程相关的 lncRNAs 及其对应的 114 个靶基因, 在这些 DE mRNAs 中, 由顺式和反式共同调控的靶基因达 77.19%, 而仅通过顺式或反式作用调控的靶基因分别为 9.65% 和 13.16%。有趣的是, 一个 lncRNA 可以与许多 mRNA 相互作用, 这表明一个 lncRNA 可以参与调控多种 mRNA 的表达。GO 功能注释结果显示, lncRNA 顺式和反式靶基因的前 40 条 GO 术语 (terms) 分类分析结果与靶 mRNA 之间的分布模式相似, 主要分为 3 大类, 分别为“生物学过程”、“细胞成分”和“分子功能”。各部分富集程度最高的通路也分别为细胞过程, 细胞和附着 (图 7-b, d)。KEGG 注释结果显示, 4 609 个 lncRNA 顺式靶基因被注释到 245 条代谢途径且划分为 49 个 terms, 其中最主要的是“核糖体”(60 个 DEG)、“肌动蛋白细胞骨架的调控”(59 个 DEG) 和“嘌呤代谢”(57 个 DEG)(图 7-c)。2 965 个 lncRNA 反式靶基因被注释到 1 911 条 KEGG 划分为 49 个 terms, 其中富集程度最高的是“黏附斑”(53 个 DEG)、“肌动蛋白细胞骨架的调节”(50 个 DEG) 和“内吞作用”(50 个 DEG)(图 7-e)。

3 讨论

自从哺乳动物 Y 染色体上雄性决定基因 *sry* 被发现以来^[23], 研究者们随后又在鸟类、鱼类、两栖类和爬行类中鉴定了一系列性别决定基因或性别决定候选基因^[1]。如青鳉 (*Oryzias latipes*) 中的 *dmy/dmrt1bY* 与精巢发育有关^[24], 红原鸡 (*Gallus gallus*)^[25] 和半滑舌鳎 (*Cynoglossus semilaevis*)^[26] 中与 Z 染色体连锁的雄性决定基因 *dmrt1*, 非洲爪蟾 (*Xenopus laevis*) W 染色体连锁的雌性决定基因 *dm-w*^[27] 以及在吕宋青鳉 (*O. luzonensis*)^[28] 和恒河青鳉 (*O. dancena*) 中的 *gsdf* 和 *sox3*^[29]。本研究通过转录组比较分析, 共发现有 8 237 个差异表达的 mRNA 是性别偏向基因 (图 3)。GO 和 KEGG

表 2 差异表达的编码转录本注释概述

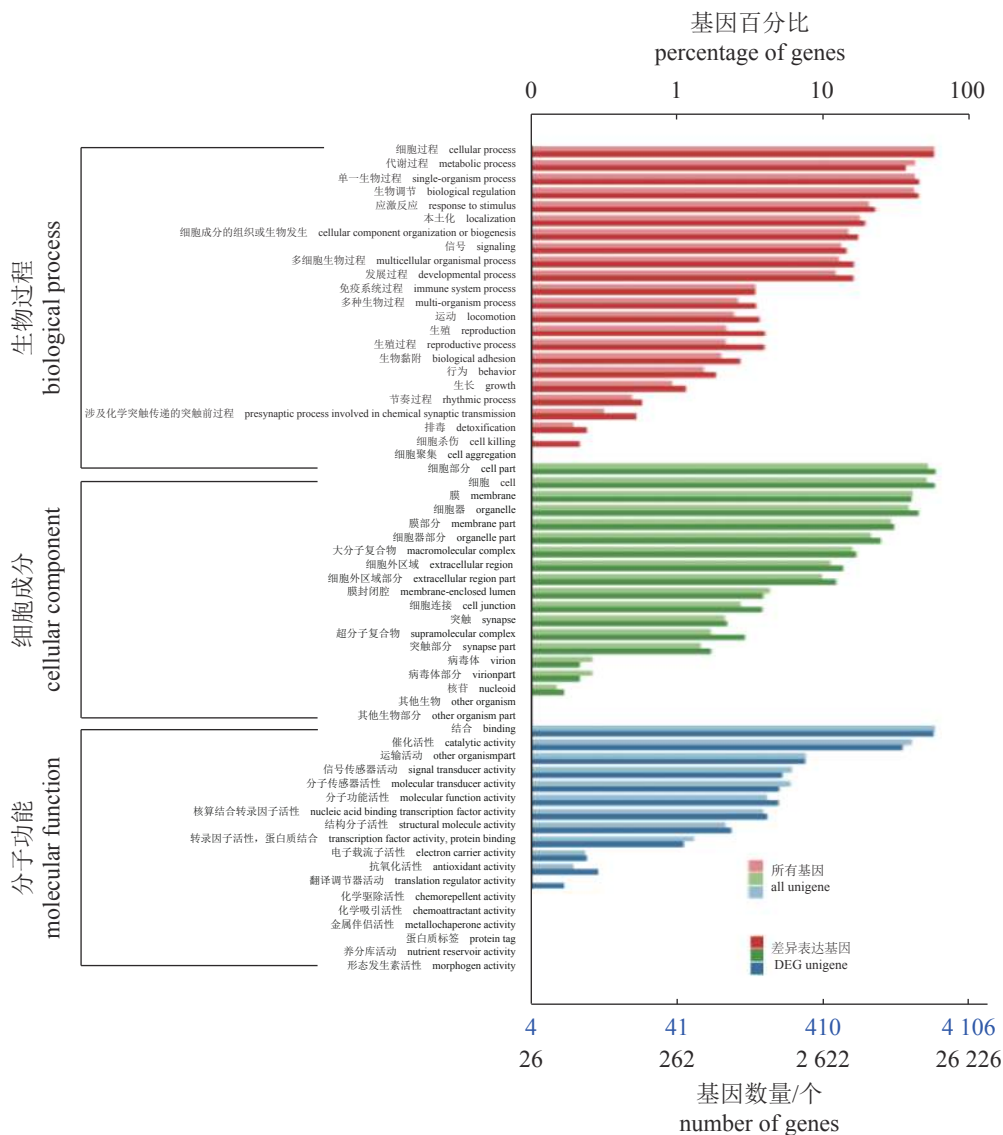
Tab. 2 Overview of differentially expressed coding transcripts annotation.

差异表达基因 DEGs	COG 注释 COG-annotation	GO 注释 GO-annotation	KEGG 注释 KEGG-annotation	KOG 注释 KOG-annotation	Swiss-prot 注释 Swissprot-annotation	NR 注释 NR-annotation	所有注释 all-annotation
数量/个 number	1 269	3 810	2 930	3 115	2 963	5 490	5 493
百分比/% percentage	23.10	69.36	53.34	56.71	53.94	99.96	

分析表明一些 DEGs, 如 *ddx25*、*tdrd9*、*nr5a2*、*vasa*、*piwil1*、*piwil2*、*rad51a*、*tdrd1*、*tdrd7* 和 *stra6* 在“卵母细胞减数分裂”、“孕酮介导的卵母细胞成熟”、“精子发育”和“性腺发育”等与性别调控相关的通路中显著富集(图 5, 7)。因此, lncRNAs 可能通过一个复杂的 lncRNA -mRNA 网络, 通过顺式或/和反式相关途径参与性别调控。此外, 基因注释的结果显示, *nr5a2* 参与了类固醇激素介导的信号通路及雌性性腺发育过程, *tdrd7* 参与了精子形成的减数分裂过程, *stra6* 则参与了卵巢发育, 这与先前的报道中这些基因主要参与炎症反应或与糖尿病的功能不同^[30-31], 后续我们将进一步对这些基因在黄喉拟水龟性

腺发育及性别决定中的生物学功能进行探究与验证。

lncRNA 属于非编码转录本, 与 mRNA 不同, 曾一度被认为是无用的信息。但是近些年的研究发现, lncRNA 在激活发育和分化关键调控因子中发挥着出乎意料的作用^[32]。在大黄鱼 (*Larimichthys crocea*) 中, lncRNA MSTRG.24346 与大黄鱼的性别决定候选基因 *dmrt1* 距离相近, 且相关性极显著, 表明 lncRNA 可能在其性别分化中起重要作用^[33]。尽管近年来发现 lncRNA 广泛参与了动植物的性别决定、性别差异和配子发生等多种生殖过程, 但关于 ESD 爬行类中 lncRNA 的生物学功能研究报道较少。在本研究中, 我



(a)

(图 5 Fig. 5)

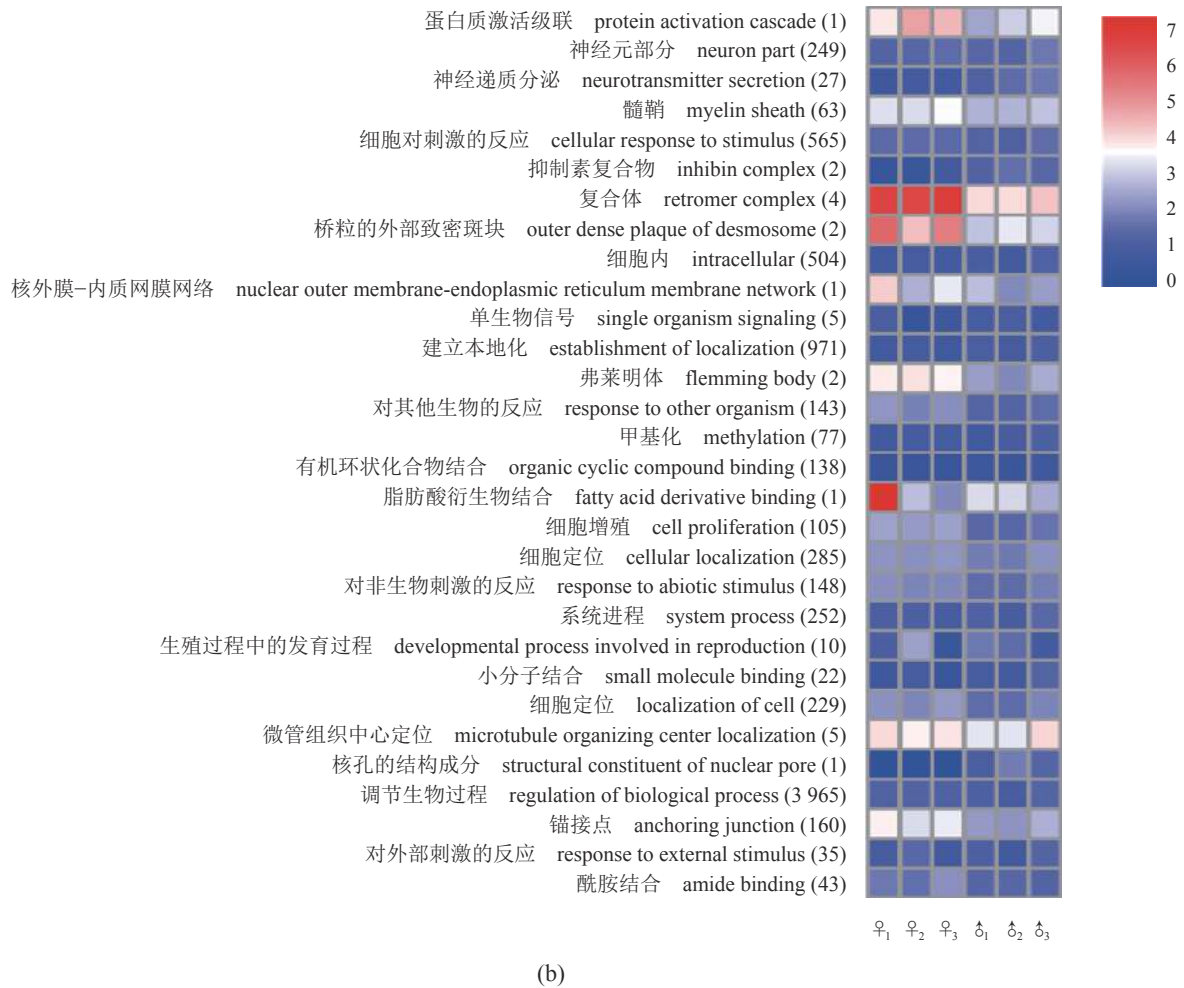


图 5 差异表达编码转录本的注释摘要和分类

(a) 差异表达编码转录本 GO 分类的直方图, (b) 差异表达的编码转录本的 GO 富集图; 红色表示高表达功能分类, 蓝色表示相对低表达的术语, 每个术语标签后的括号内是其中包含显著差异的基因数量

Fig. 5 Annotation summary and classification of differentially expressed coding transcripts of ovaries and testes

(a) histogram of GO classification of differentially expressed coding transcripts, (b) GO enrichment cluster diagram of differentially expressed coding transcripts; red denotes high expression functional classification, while blue denotes relatively low expression term; within the parentheses after each term tag is the number of genes with significant differences contained in the term

们预测了具有生物学意义的 lncRNAs, 获得了 9 573 个 DE lncRNA, 这是首次观察到这些 lncRNAs 在黄喉拟水龟精巢和卵巢中差异表达。有趣的是, 差异表达 lncRNAs 在精巢中的数量比卵巢中多, 这点与人的相似, 但与中华鳖 (*Pelodiscus sinensis*) 不同, 在中华鳖卵巢中存在较多的 lncRNAs。这些发现暗示 lncRNAs 在触发黄喉拟水龟性腺分化为精巢或卵巢时可能发挥特定功能。

lncRNA 主要通过顺式或反式 2 种方式调节靶基因的表达, 顺式作用的 lncRNA 在转录水平或转录后水平上调节相邻基因的表达, 而反式

作用的 lncRNA 则在独立的基因座上影响基因的表达^[34]。在黑腹果蝇 (*Drosophila melanogaster*) 中, lncRNA 可能会影响与雌性个体性别激活有关的主效基因启动子的表达。在小鼠中, lncRNA 通过顺式作用调控相邻编码基因的表达, 从而影响生殖干细胞分化生成配子。本研究中, 我们筛选了参与性腺发育、配子形成、性别决定与分化等生殖相关的 lncRNA-mRNA 对, 通过顺式和反式模型分析预测了参与黄喉拟水龟雌、雄性腺发育相关的 lncRNAs 及其对应的 114 个目标基因, 为进一步解析 lncRNAs 在黄喉拟水龟温度依赖型性别决定分子机制奠定了基础。

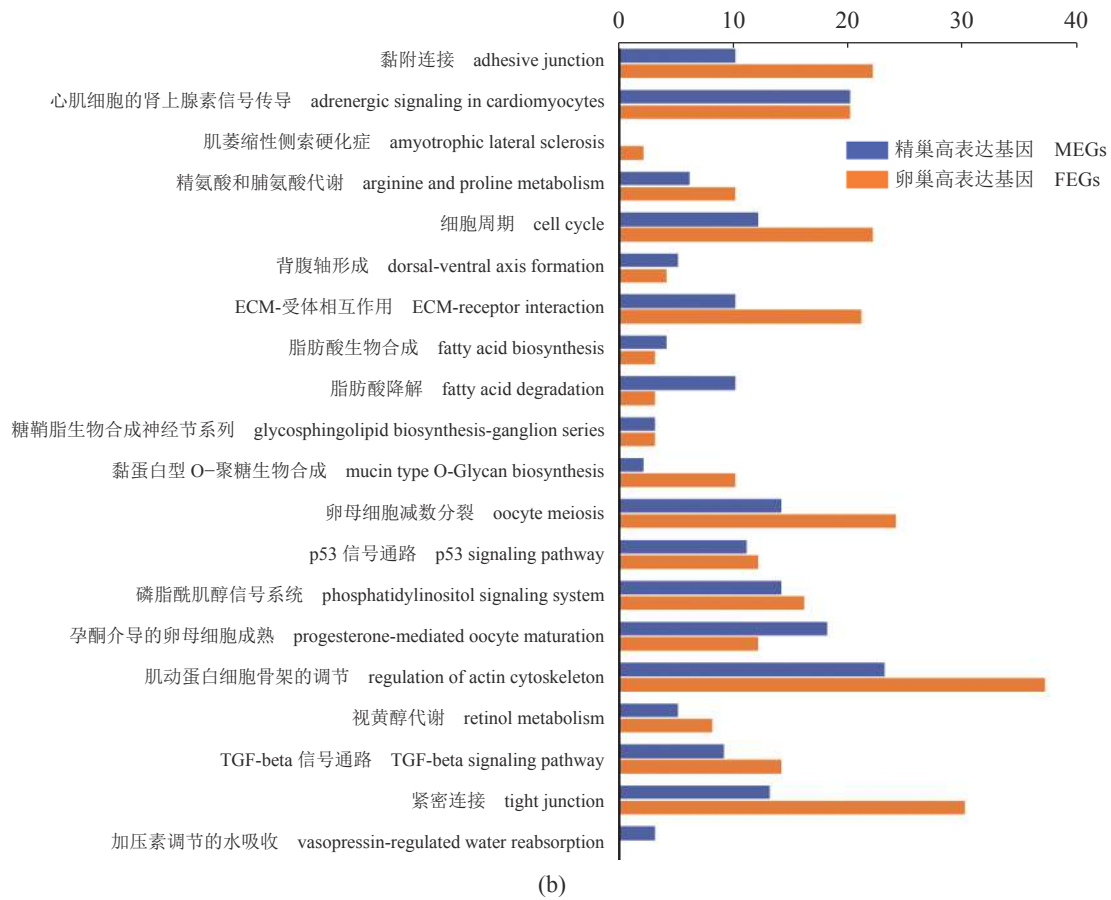
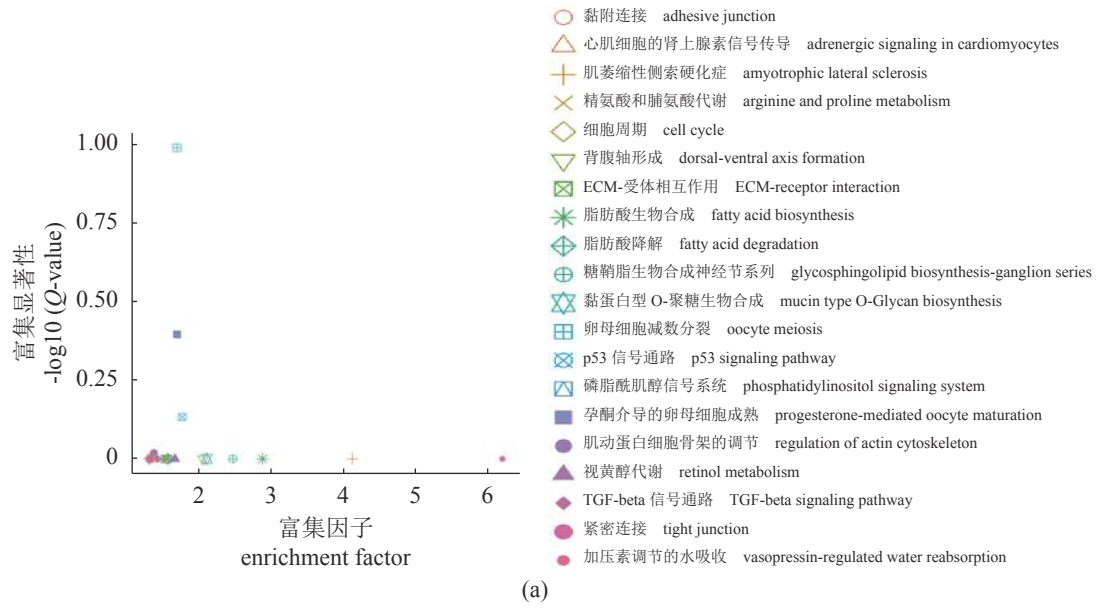


图 6 差异表达的编码转录本的 KEGG 分类

(a) 具有差异表达的编码转录本的前 20 个富集的 KEGG 信号途径, X 轴代表 KEGG 富集因子, Y 轴代表 KEGG 途径中差异表达的编码转录本的富集重要性, (b) 参与前 20 个富集 KEGG 信号通路的雌性高表达基因 (FEGs) 和雄性高表达基因 (MEGs)

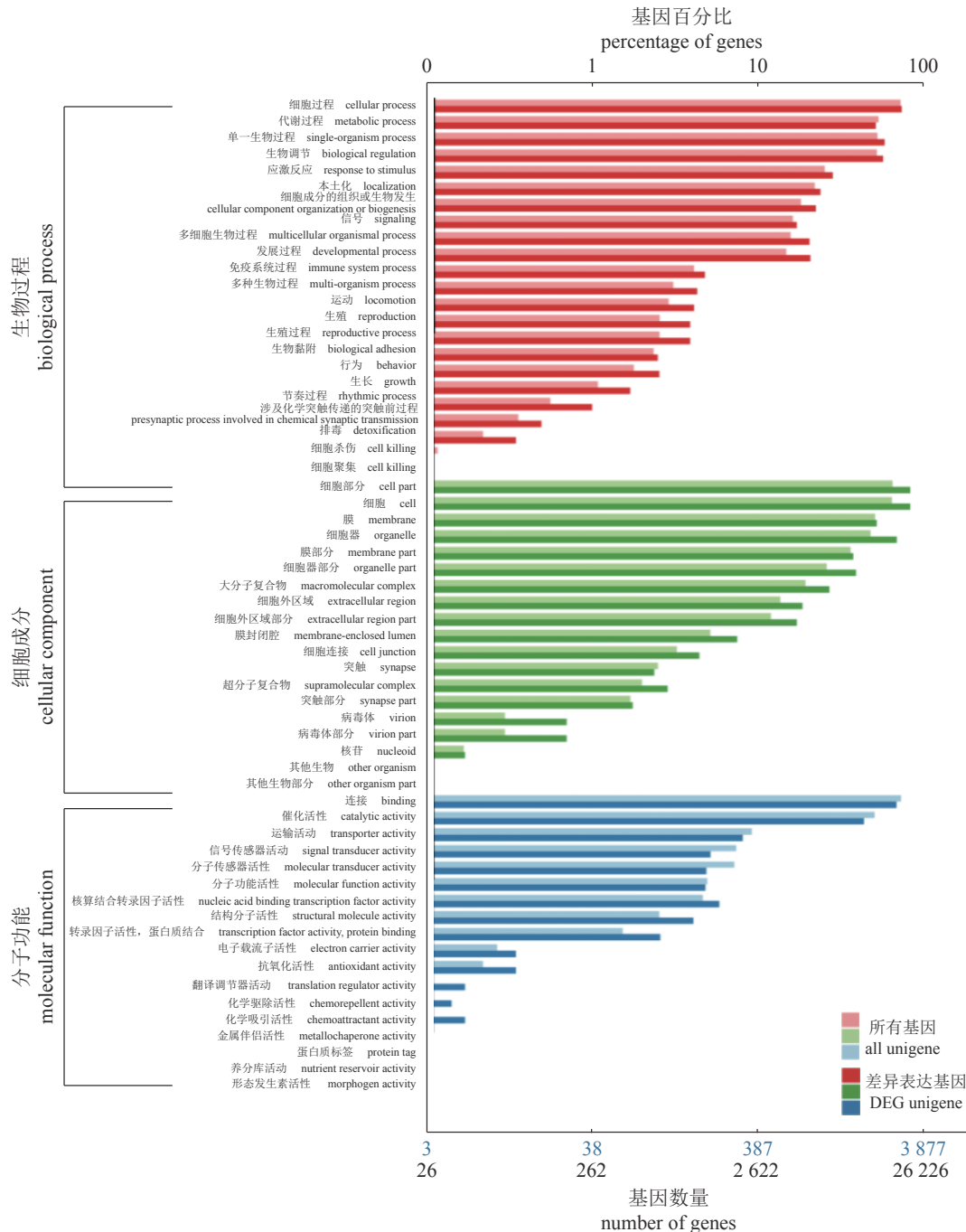
Fig. 6 KEGG classification of differentially expressed coding transcripts annotation

(a) top 20 enriched KEGG signal pathways with differentially expressed coding transcripts, X-axis represents KEGG enrichment factor, Y-axis represents enrichment significance of differentially expressed coding transcripts in the KEGG pathway, (b) female-highly expressed genes (FEGs) and male-highly expressed genes (MEGs) involved in the top 20 enriched KEGG signal pathways

表 3 lncRNA 顺式和反式差异表达靶基因注释的概述

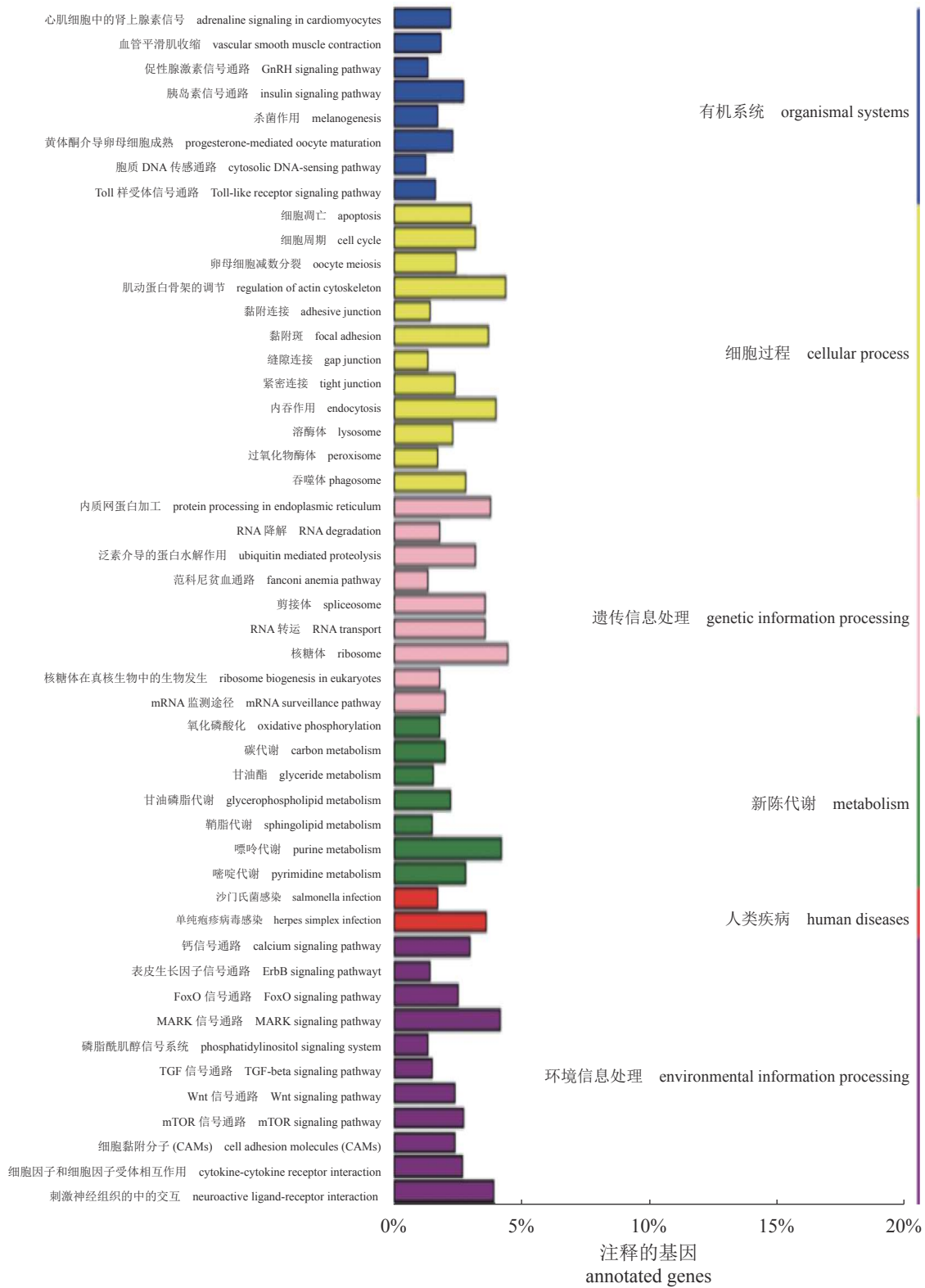
Tab. 3 Overview of differentially expressed coding transcripts annotation which targeted cis- and trans-acting of lncRNAs

		COG注释 COG- annotation	GO注释 GO- annotation	KEGG注释 KEGG- annotation	KOG注释 KOG- annotation	Swiss-prot注释 Swissprot- annotation	NR注释 NR- annotation	所有注释 all- annotation
长链非编码RNA 顺式靶基因 lncRNA-cis	数量 number 百分比/% percentage	1 442	3 877	3 022	3 315	3 173	4 609	4 609
长链非编码RNA 反式靶基因 lncRNA-trans	数量 number 百分比/% percentage	970	2 369	1 964	2 116	2 116	2 965	2 965

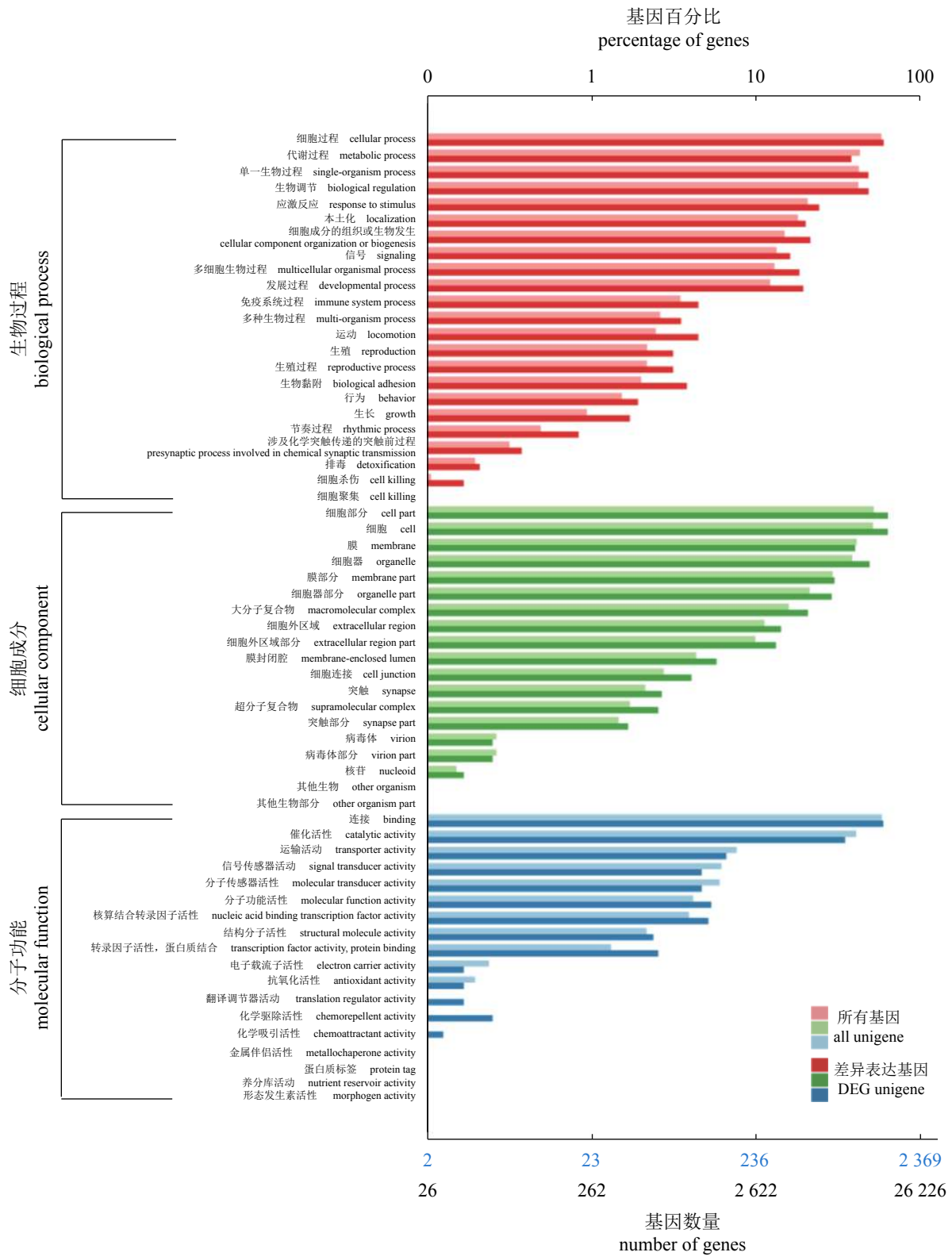


(a)

(图 7 Fig. 7)

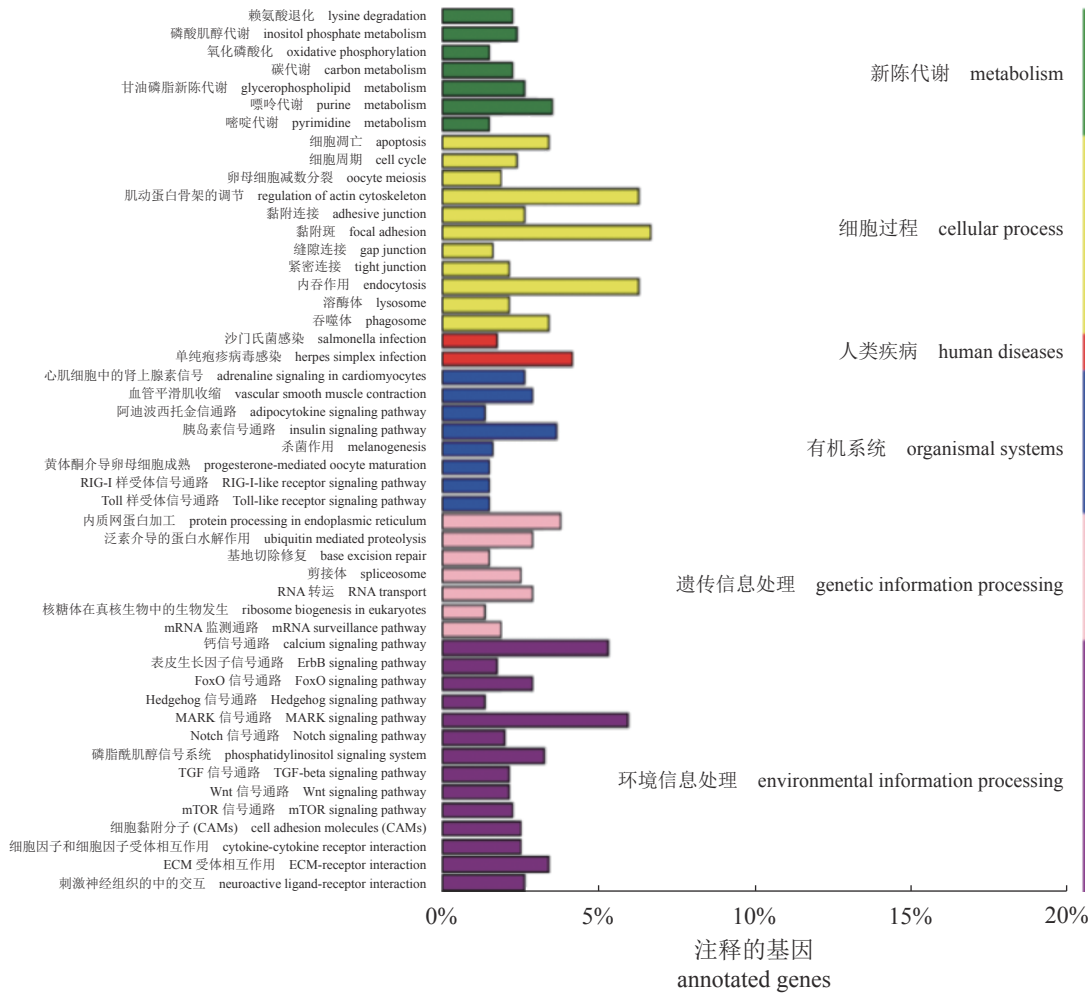


(图 7 Fig. 7)



(c)

(图 7 Fig. 7)



(d)

图 7 对 lncRNA 顺式和反式靶基因进行注释总结和分类

(a) (c) lncRNAs 的顺式和反式差异表达靶基因注释的 GO 分类直方图, (b) (d) lncRNAs 的顺式和反式差异表达靶基因注释的 KEGG 分类的直方图

Fig. 7 Annotation summary and classification of target genes by cis- and trans-acting of lncRNAs

(a) and (c) histogram of GO classification of differentially expressed coding transcripts annotation which targeted cis- and trans-acting of lncRNAs, (b) and (d) histogram of KEGG classification of differentially expressed coding transcripts annotation which targeted cis- and trans-acting of lncRNAs

4 结论

本研究利用 Illumina HiSeq™4000 测序平台首次对黄喉拟水龟雌雄性腺进行了比较转录组分析, 筛选出了一些与黄喉拟水龟性别分化及性腺发育密切相关的 lncRNAs 及其相互作用的靶标 mRNAs。通过对差异表达转录本进行的 GO 功能注释和 KEGG 功能富集分析, 发现雌雄差异分子主要在“卵母细胞减数分裂”、“孕酮介导的卵母细胞成熟”、“精子发育”和“性腺发育”等与性别调控相关的通路中显著富集。本研究结果为进一步阐明黄喉拟水龟性别分化的分子机

制提供了基础数据。然而, 本研究仅初步分析了黄喉拟水龟雌雄性腺差异分子, 从中鉴定到关键基因或 lncRNAs, 并解析龟鳖动物温度依赖性性别分化基因调控网络, 日后还需要开展大量的更为细致的研究工作。

参考文献 (References):

[1] Li X Y, Gui J F. Diverse and variable sex determination mechanisms in vertebrates[J]. *Science China Life Sciences*, 2018, 61(12): 1503-1514.

[2] Ge C T, Ye J, Weber C, *et al.* The histone demethylase
中国水产学会主办 sponsored by China Society of Fisheries

- KDM6B* regulates temperature-dependent sex determination in a turtle species[J]. *Science*, 2018, 360(6389): 645-648.
- [3] Baroiller J F, D'Cotta H, Saillant E. Environmental effects on fish sex determination and differentiation[J]. *Sexual Development*, 2009, 3(2-3): 118-135.
- [4] Cheng Y Y, Chen T Y, Yu P H, *et al.* Observations on the female reproductive cycles of captive Asian yellow pond turtles (*Mauremys mutica*) with radiography and ultrasonography[J]. *Zoo Biology*, 2010, 29(1): 50-58.
- [5] 朱新平, 陈永乐, 魏成清, 等. 温度对黄喉拟水龟性别决定的影响[J]. *生态学报*, 2006, 26(2): 620-625.
Zhu X P, Chen Y L, Wei C Q, *et al.* Temperature effects on sex determination in yellow pond turtle (*Mauremys mutica* Cantor)[J]. *Acta Ecologica Sinica*, 2006, 26(2): 620-625.
- [6] Capel B. Vertebrate sex determination: evolutionary plasticity of a fundamental switch[J]. *Nature Reviews Genetics*, 2017, 18(11): 675-689.
- [7] Ge C T, Ye J, Zhang H Y, *et al.* *Dmrt1* induces the male pathway in a turtle species with temperature-dependent sex determination[J]. *Development*, 2017, 144(12): 2222-2233.
- [8] Schroeder A L, Metzger K J, Miller A, *et al.* A novel candidate gene for temperature-dependent sex determination in the common snapping turtle[J]. *Genetics*, 2016, 203(1): 557-571.
- [9] Li T T, Wang S Y, Wu R M, *et al.* Identification of long non-protein coding RNAs in chicken skeletal muscle using next generation sequencing[J]. *Genomics*, 2012, 99(5): 292-298.
- [10] Cossetti C, Lugini L, Astrologo L, *et al.* Soma-to-germline transmission of RNA in mice xenografted with human tumour cells: possible transport by exosomes[J]. *PLoS ONE*, 2014, 9(7): e101629.
- [11] Yue C Y, Li Q, Yu H. Gonad transcriptome analysis of the pacific oyster *Crassostrea gigas* identifies potential genes regulating the sex determination and differentiation process[J]. *Marine Biotechnology*, 2018, 20(2): 206-219.
- [12] Yang X L, Ikhwanuddin M, Li X C, *et al.* Comparative transcriptome analysis provides insights into differentially expressed genes and long non-coding RNAs between ovary and testis of the Mud crab (*Scylla paramamosain*)[J]. *Marine Biotechnology*, 2018, 20(1): 20-34.
- [13] 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 Research*, 2007, 35: W345-W349.
- [14] Sun L, Luo H T, Bu D C, *et al.* Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts[J]. *Nucleic Acids Research*, 2013, 41(17): e166.
- [15] Wang L G, Park H J, Dasari S, *et al.* CPAT: coding-potential assessment tool using an alignment-free logistic regression model[J]. *Nucleic Acids Research*, 2013, 41(6): e74.
- [16] Finn R D, Bateman A, Clements J, *et al.* Pfam: the protein families database[J]. *Nucleic Acids Research*, 2014, 42: D222-D230.
- [17] Trapnell C, Williams B A, Pertea G, *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation[J]. *Nature Biotechnology*, 2010, 28(5): 511-515.
- [18] Li J W, Ma W, Zeng P, *et al.* LncTar: a tool for predicting the RNA targets of long noncoding RNAs[J]. *Briefings in Bioinformatics*, 2015, 16(5): 806-812.
- [19] Shannon P, Markiel A, Ozier O, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks[J]. *Genome Research*, 2003, 13(11): 2498-2504.
- [20] Anders S, Huber W. Differential expression analysis for sequence count data[J]. *Genome Biology*, 2010, 11(10): R106.
- [21] Xie C, Mao X Z, Huang J J, *et al.* KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases[J]. *Nucleic Acids Research*, 2011, 39(S2): W316-W322.
- [22] Li X Y, Li Z, Zhang X J, *et al.* Expression characterization of testicular *dmrt1* in both sertoli cells and spermatogenic cells of polyploid gibel carp[J]. *Gene*, 2014, 548(1): 119-125.
- [23] Koopman P, Gubbay J, Vivian N, *et al.* Male development of chromosomally female mice transgenic for

- Sry*[J]. *Nature*, 1991, 351(6322): 117-121.
- [24] Matsuda M, Nagahama Y, Shinomiya A, *et al.* *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish[J]. *Nature*, 2002, 417(6888): 559-563.
- [25] Smith C A, Roeszler K N, Ohnesorg T, *et al.* The avian Z-linked gene *DMRT1* is required for male sex determination in the chicken[J]. *Nature*, 2009, 461(7261): 267-271.
- [26] Shao C W, Li Q Y, Chen S L, *et al.* Epigenetic modification and inheritance in sexual reversal of fish[J]. *Genome Research*, 2014, 24(4): 604-615.
- [27] Yoshimoto S, Okada E, Umemoto H, *et al.* A W-linked DM-domain gene, *DM-W*, participates in primary ovary development in *Xenopus laevis*[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(7): 2469-2474.
- [28] Myosho T, Otake H, Masuyama H, *et al.* Tracing the emergence of a novel sex-determining gene in Medaka, *Oryzias luzonensis*[J]. *Genetics*, 2012, 191(1): 163-170.
- [29] Takehana Y, Matsuda M, Myosho T, *et al.* Co-option of *Sox3* as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*[J]. *Nature Communications*, 2014, 5: 4157.
- [30] Nair A K, Sugunan D, Kumar H, *et al.* Case-control analysis of SNPs in *GLUT4*, *RBP4* and *STRA6*: association of SNPs in *STRA6* with type 2 diabetes in a South Indian population[J]. *PLoS ONE*, 2010, 5(7): e11444.
- [31] Melia T, Waxman D J. Sex-biased lncRNAs inversely correlate with sex-opposite gene coexpression networks in diversity outbred mouse liver[J]. *Endocrinology*, 2019, 160(5): 989-1007.
- [32] Ørom U A, Derrien T, Beringer M, *et al.* Long noncoding RNAs with enhancer-like function in human cells[J]. *Cell*, 2010, 143(1): 46-58.
- [33] 王伟佳, 韩兆方, 李完波, 等. 大黄鱼雌雄性腺长链非编码RNA的挖掘与差异分析[J]. *中国水产科学*, 2019, 26(5): 852-860.
- Wang W J, Han Z F, Li W B, *et al.* The identification and analysis of long noncoding RNA in testes and ovaries of the large yellow croaker (*Larimichthys crocea*)[J]. *Journal of Fishery Sciences of China*, 2019, 26(5): 852-860.
- [34] Jin G X, Wen H S, He F, *et al.* Molecular cloning, characterization expression of P450c17-I and P450c17-II and their functions analysis during the reproductive cycle in males of barfin flounder (*Verasper moseri*)[J]. *Fish Physiology and Biochemistry*, 2012, 38(3): 807-817.

Identification and analysis of lncRNAs and mRNAs involved in sex regulation in Asian yellow pond turtle (*Mauremys mutica*)

ZHU Yanyu^{1,2}, LIU Xiaoli¹, WANG Yakun¹, LI Wei¹,
HONG Xiaoyou¹, ZHU Xinping^{1,2*}, XU Hongyan^{1,2*}

(1. Key Laboratory of Tropical & Subtropical Fishery Resource Application & Cultivation of Ministry of Agriculture and Rural Affairs, Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou 510380, China;

2. National Demonstration Center for Experimental Fisheries Science Education, Shanghai Ocean University, Shanghai 201306, China)

Abstract: Sexual dimorphism is widespread in animal kingdom and has been a hot topic in biology for a long time. In particular, the sex of the *Mauremys mutica* is determined by incubation temperature, however, the molecular mechanism of temperature response is still unclear. In order to understand the molecular mechanism of the sex differentiation of turtles, this study made a preliminary comparative analysis of the sex-differentiated lncRNAs and their target genes in the transcripts of *M. mutica*. First, the Illumina deep sequencing platform was used to perform comparative transcriptomics analysis of the testis and ovary of *M. mutica* by RNA-Seqs and screened to identify male and female differential transcripts. A total of 8 237 differentially expressed mRNAs and 9 573 differentially expressed lncRNAs were screened. Through GO functional annotation and KEGG pathway enrichment analysis, we found that these differential transcripts were mainly involved in the sex differentiation and gonadal development of turtles. In addition, a series of target genes related to reproductive development (gonad development and sex differentiation) regulated by DE lncRNA were obtained through the analysis of *cis*- and *trans*-effects. This study provides clues for further elucidating the temperature-dependent sex differentiation mechanism of *M. mutica*, and especially, it lays the foundation for further exploring the use of the sex determinants of turtles to conduct research on animal sexuality control breeding techniques for turtles.

Key words: *Mauremys mutica*; transcriptome; long noncoding RNA (lncRNA); sex determination; sex differentiation

Corresponding authors: ZHU Xinping. E-mail: zhuxinping_1964@163.com;
XU Hongyan. E-mail: xuhyzqh@163.com

Funding projects: National Key Research and Development Program of China (2018YFD0900201, 2018YFD0900203); National Natural Science Foundation of China (31873035); Science and Technology Project of Guangzhou (201904010172); Social Public Welfare Research (2018SJ-YZ02; 2019HY-XKQ02); China-ASEAN Maritime Cooperation Fund (AMC-2018F); International Collaborations on Agriculture Technologies between China and the Tropical Countries along the Belt and Road (2130114)