研究论文

饥饿胁迫下黑腹果蝇长链非编码 RNA 的差异表达和调控网络分析*

高丽斌** 李美珍 贺 康 李 飞***

（浙江大学昆虫科学研究所，杭州 310058）

摘 要 **【目的】** 饥饿是生命体普遍经历的一种胁迫，本文旨在探究饥饿胁迫下黑腹果蝇 *Drosophila melanogaster* 长链非编码 RNA（lncRNA）的表达变化及其参与的调控网络，揭示 lncRNA 在饥饿应激反应中的潜在功能。**【方法】** 下载 NCBI SRA 数据库中黑腹果蝇非生物胁迫相关的 RNA-seq 文库，进行转录组组装，预测 lncRNA 基因。利用 DEseq2 软件分析不同时间饥饿胁迫下 lncRNA 的差异表达，利用 WGCNA 共表达分析发现与饥饿胁迫相关的基因模块，构建差异表达 lncRNA 与蛋白编码基因的调控网路，用 clusterProfiler 进行 GO 功能注释和 KEGG 通路富集分析。**【结果】** 从 34 个果蝇非生物胁迫转录组中共鉴定得到 3 612 个 lncRNA。与正常喂食组相比，在 4、12、16 和 24 h 饥饿胁迫下，分别有 260、28、200 和 26 个 lncRNA 表达上调，133、28、148 和 25 个 lncRNA 表达下调（以下简称为饥饿相关 lncRNA）。这些 lncRNA 分别与 790、2 950、725、2 961 个蛋白编码基因表达相关。GO 和 KEGG 富集分析显示，这些蛋白编码基因分别富集于吞噬体和溶酶体通路、糖代谢通路、长寿调节途径、化学刺激和味觉感受及细胞增殖和凋亡等通路中。在饥饿胁迫 4、12、16 和 24 h 下，分别有 66、1、23 和 3 个 lncRNA 基因处于调控网络中的核心（hub gene）节点位置（|基因显著性（GS）|>0.4，|模块成员（MM）|>0.8）。这些核心 RNA 可能具有更为关键的调控作用。**【结论】** LncRNA 参与了果蝇的饥饿应激反应，并与饥饿处理的时间有关。在较短时间饥饿处理下，饥饿相关 lncRNA 主要参与调控自噬及能量代谢等过程，以保证能量供应；在较长时间饥饿处理下，饥饿相关 lncRNA 则主要参与觅食、细胞增殖及凋亡以及长寿调节途径等。

关键词 黑腹果蝇；饥饿胁迫；lncRNA；WGCNA

The differential expression of long non-coding RNAs and their associated regulation networks in *Drosophila melanogaster* subject to starvation stress

GAO Li-Bin** LI Mei-Zhen HE Kang LI Fei***

(Institute of Insect Sciences, Zhejiang University, Hangzhou 310058, China)

Abstract **[Objectives]** To investigate changes in the expression of long non-coding RNA (lncRNA) and their associated regulatory networks in *D. melanogaster* under starvation stress, and thereby reveal the roles of lncRNA in the starvation stress response. **[Methods]** Abiotic stress-related RNA-Seq libraries of *D. melanogaster* were downloaded from the NCBI SRA database and transcriptomes assembled from these data. LncRNAs were identified from the assembled transcriptomes. The DEseq2 package was used to identify lncRNAs that were differentially expressed under starvation stress. Using the WGCNA tool, we obtained co-expression modules related to starvation stress and constructed regulatory networks of lncRNAs and protein coding genes. Gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed using clusterProfiler. **[Results]** A total of 3 612 lncRNAs were identified from the assembled

*资助项目 Supported projects：浙江省自然科学基金重点项目（LZ18C060001）

**第一作者 First author，E-mail：gaolb@zju.edu.cn

***通讯作者 Corresponding author，E-mail：lifei18@zju.edu.cn

收稿日期 Received：2021-03-15；接受日期 Accepted：2021-05-31

transcriptome. After 4, 12, 16, and 24 h of starvation, 260, 28, 200 and 26 lncRNAs were up-regulated and 133, 28, 148, and 25 were down-regulated, respectively, relative to the control groups. These lncRNAs are thereafter referred to starvation-related lncRNAs. Expression of starvation-related lncRNAs was correlated with 790, 2 950, 725 and 2 961 protein-coding genes, respectively. GO and KEGG analysis showed that the enriched pathways changed with the duration of starvation. Pathways associated with phagosome and lysosome production, sugar metabolism, longevity regulation, detection of chemical stimulus, sensory perception of taste, cells signals related to proliferation and apoptosis, were sequentially enriched as the duration of starvation increased. After 4, 12, 16, and 24 h of starvation, 66, 1, 23, and 3 hub lncRNAs were found (|gene significance (GS)|>0.4, |module members (MM) |>0.8). These hub lncRNAs may have key regulatory roles in responding to starvation stress. **[Conclusion]** LncRNAs play essential roles in the starvation stress response in *D. melanogaster* and the mechanism regulating their expression is closely related to starvation treatment time. Under short-term starvation stress, starvation-related lncRNAs are mainly involved in the regulation of autophagy and energy metabolism and ensuring energy supply, however, under long-term starvation stress, starvation-related lncRNAs are mainly involved in foraging, cell proliferation, apoptosis and longevity regulation.

Key words *Drosophila melanogaster*; starvation stress; lncRNA; weighted gene co-expression network analysis

所有生命过程都需要消耗能量，充足的能量供应对生物体的生存和繁殖至关重要。在复杂多变的自然条件下，饥饿是动物普遍会经历的一种胁迫，能够有效应对饥饿的生物体具有更强的环境适应性和竞争力。一直以来，生物体在饥饿胁迫下的应激反应过程受到了广泛的关注，对这一问题的研究有助于加深对能量代谢过程的理解。研究表明，动物进化出多种机制来应对营养物质的严重短缺，如促进觅食行为、减少睡眠、降低代谢率、细胞自噬等（Sawin *et al.*，2000；Kuma *et al.*，2004；Masek *et al.*，2014；Schmidt，2014）。

动物的营养应激分子机制具有一定的保守性，因此利用模式动物研究发现的调控机制具有较好的推广性和普适性（Rion and Kawecki，2007）。针对黑腹果蝇 *Drosophila melanogaster* 的研究发现，在饥饿胁迫下，调控昼夜节律的时钟基因（*clock* 和 *cycle*）、高度保守的 RNA / DNA 结合蛋白 Translin（Tsn）等参与调控果蝇睡眠抑制（Keene *et al.*，2010；Murakami *et al.*，2016）。短神经肽 F（sNPF）及其受体、γ-氨基丁酸 GABA 及其受体等参与调控饥饿胁迫下果蝇的嗅觉感受（Root *et al.*，2011；Slankster *et al.*，2020），而脂动激素受体（AKHR）激活及章鱼胺（Octopamine）表达则能促进饥饿诱导下的机能亢进（Yang *et al.*，2015；Yu *et al.*，2016），诱导果蝇觅食行为的提高。在果蝇细胞内，PI3K 和 TOR 等信号通路参与调控自噬，维持饥饿下的能量供应（Scott *et al.*，2004）。

长链非编码 RNA 基因（lncRNA）是普遍存在于基因组中、长度大于 200 nt 且不编码蛋白的 RNA 转录本，在果蝇生长发育、睡眠、运动等多个生物学过程中发挥重要作用（Soshnev *et al.*，2011；Li *et al.*，2012；Pease *et al.*，2013），但其在饥饿胁迫下的调控作用研究不多。为此，本研究通过黑腹果蝇在非生物胁迫下的转录组数据组装和注释，发现饥饿胁迫诱导的差异表达 lncRNA 基因，通过基因共表达网络构建，明确果蝇饥饿胁迫下 lncRNA 应激调控的动态网络变化，为解析果蝇饥饿应激反应的网络调控机制提供了理论参考和依据。

1 材料与方法

1.1 数据来源

转录组数据来源于 NCBI 的 SRA 数据库（https://www.ncbi.nlm.nih.gov/sra/）。下载该数据库中由 Illumina 平台测序完成的与饥饿胁迫相关的 34 个黑腹果蝇转录组测序数据（截止到 2019 年 5 月 29 日）。根据实验条件分组如下：（1）正常喂食样本 14 个：SRR5145160，SRR5145162，SRR5633695，SRR5633696，SRR5633697，SRR5633698，SRR6315674，SRR6315675，SRR6315676，SRR6315677，SRR6315678，SRR6315679，SRR6315680，SRR6315681；（2）4 h 饥饿胁迫样本 4 个：SRR5633699，SRR5633700，SRR5633701，SRR5633702；（3）12 h 饥饿胁迫

样本 2 个：SRR5145161，SRR5145163；（4）16 h 饥饿胁迫样本 6 个：SRR5633703，SRR5633704，SRR5633707，SRR5633708，SRR5633709，SRR5633710；（5）24 h 饥饿胁迫样本 8 个：SRR6315682，SRR6315683，SRR6315684，SRR6315685，SRR6315686，SRR6315687，SRR6315688，SRR6315689。

1.2　转录组组装及 lncRNA 鉴定

黑腹果蝇参考基因组和基因注释均来自 NCBI RefSeq 数据库（accession：GCF_000001215.4），参考 Hisat2+StringTie 的组合进行有参转录组组装（Pertea *et al.*，2016）（图 1）。为获得更完整的果蝇非生物胁迫相关转录组组装数据，下载 910 个果蝇非生物胁迫的转录组。首先，使用 SRAtools 和 Trimmomatic v0.36（Bolger *et al.*，2014）对这些转录组原始数据进行解压、去接头和质量过滤。然后，利用 Hisat2 v2.1.0（Pertea *et al.*，2016）将有效读段（clean reads）比对到参考基因组上，去除比对率低于 50%的 9 个文库。最后，使用 StringTie v1.3.4（Pertea *et al.*，2016）分别组装剩余的 901 个转录组，并将结果合并，得到较为完整的黑腹果蝇非生物胁迫转录组。

使用 GffCompare v0.10.4（Pertea and Pertea，2020）将组装后的转录组与黑腹果蝇 RefSeq 注释中的 2 480 个 lncRNA（标注为“antisense_RNA”及“lnc_RNA”）比较，保留分类码（class_code）为=（剪接位点完全相同的转录本）的转录本为已知 lncRNA；与 RefSeq 注释中的 13 955 个蛋白编码基因（标注为“mRNA”）和 2 480 个 lncRNA 比较，将分类码（class_code）为 u（基因间区转录本）、i（内含子区转录本）、x（负链转录本）的转录本用于预测新的 lncRNA。lncRNA 预测流程参考 Sarropoulos 等（2019），所有可变剪切需满足以下条件：1）转录本长度大于等于 200 nt，且外显子数大于 1；2）经 CPAT v1.2.4（Wang *et al.*，2013）软件预测编码可能性低于阈值（coding_prob<0.39）；3）转录本不含 Pfam32.0（Mistry *et al.*，2021）数据库中已有的蛋白结构域（HMMER v3.2.1（Finn *et al.*，2011），e-value≤1×10^{-5}）；4）转录本不包含 SwissProt 蛋白同源序列（diamond v0.8.25.87 blastx（Buchfink *et al.*，2015），e-value≤1×10^{-10}，alignment≥60 bp）；5）转录本不包含保守的蛋白编码区（https://data.broadinstitute.org/compbio1/PhyloCSFtracks/dm6/20170327/）；6）转录本不与 Rfam14.1（Kalvari *et al.*，2021）中已知的小 RNA 重合（infernal v1.1.2（Nawrocki，2014），e-value≤1×10^{-5}）。LncRNA 基因特征统计由 R-3.6.0 完成。

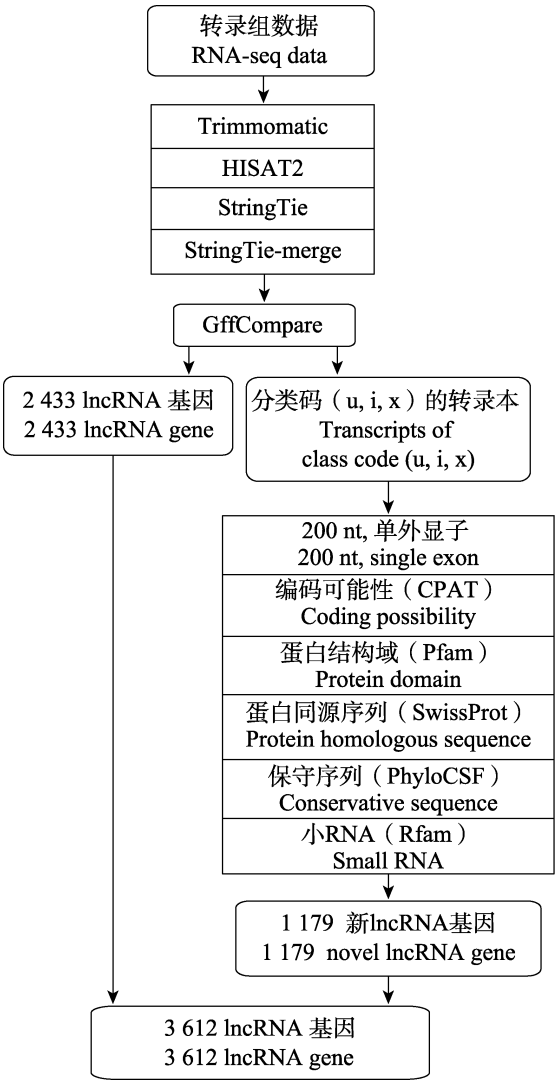


图 1　转录组组装及 lncRNA 预测流程

Fig. 1　Transcriptome assembly and lncRNA prediction pipeline

1.3　饥饿处理下 lncRNA 差异表达分析

使用 StringTie 分别计算 34 个饥饿胁迫有关文库中 lncRNA 的 FPKM，并通过脚本 prepDE.py 转换为读段数（read counts）矩阵用于后续差异

表达分析。使用 R 包 DEseq2 v1.24.0（Love *et al.*，2014）对不同样本的读段数（read counts）进行归一化，比较 lncRNA 在不同饥饿胁迫下与对照组的表达差异。筛选标准设置为：校正 *P* 值 adjusted.pvalue < 0.05，差异倍数|log2fold change（处理组/对照组）| > 1。

1.4 差异表达 lncRNA 的实时荧光定量 PCR 验证

供试黑腹果蝇 w^{1118} 由浙江大学黄健华研究员馈赠。将黑腹果蝇成虫分组进行 4、12、16、24 h 饥饿处理（2%的琼脂培养基），以同时间正常喂食的果蝇作为对照组，每个处理均设置 4 个生物学重复，每组 8 头成虫。采用 Trizol 法提取样品 RNA，取 1 μg RNA 利用 HIScript Ⅲ RT SuperMix for qPCR（+gDNA wiper）（Vazyme #R323）反转录试剂盒合成 cDNA 模板。

随机分别选取 4、12、16、24 h 饥饿处理下 2 个上调及 2 个下调的差异表达 lncRNA，使用 NCBI primer 设计特异性引物，选取黑腹果蝇 *α-tubulin* 基因（CG1913）作为内参（Ponton *et al.*，2011），利用 ChamQ Universal SYBR qPCR Master Mix（Vazyme）试剂盒进行荧光定量 PCR。qPCR 反应体系（20 μL）：2×ChamQ Universal SYBR qPCR Master Mix 10 μL，上下游引物（10 μmol·L^{-1}）各 0.4 μL，cDNA 模板 2 μL，ddH_2O 7.2 μL。反应程序设置为：95 ℃程序预变性 30 s，95 ℃变性 10 s，60 ℃退火延伸 30 s，40 个循环，结束后进行溶解曲线采集。QuantStudio 3（Applied Biosystems）定量 PCR 仪上完成反应，通过系统自带软件获取临界循环数 *C*t 值。采用 $2^{-\Delta\Delta Ct}$ 法计算 lncRNA 相对表达量，并通过 Student's *t*-test 方法检验差异显著性。

1.5 WGCNA 模块构建

使用 R 包 WGCNA v1.69，利用加权基因共表达网络分析（Weighted gene co-expression network analysis，WGCNA）进行共表达分析（Zhang and Horvath，2005）。首先，过滤在饥饿处理样本中 FPKM 均为 0 的基因，利用符合条件的 17 137 个基因进行标准化表达量矩阵构建。然后，采用软阈值 β 为 6（软阈值 β=6 时，R^2=0.85）计算邻接矩阵，并转换为拓扑重叠矩阵以计算基因间的关联程度，再转换为相异度矩阵，根据相异度矩阵对基因构建层次聚类树形图。采用动态剪枝法划分基因模块，设定模块最小基因数为 30，根据模块特征值的相似度 0.75 对表达模式相近的模块进行合并。最后，关联基因模块与不同时间饥饿胁迫，绘制模块与处理热图。模块与处理的相关性绝对值越大，则该模块中的基因与该处理相关性更强，取相关性大于 0.6 的模块进行下一步分析。

1.6 GO 功能注释和 KEGG 通路富集分析

取与饥饿胁迫相关性大于 0.6 的模块中的 lncRNA 与差异表达 lncRNA 的交集作为感兴趣的 lncRNA 用于后续分析。WGCNA 中连通性表征基因间的相关性，连通性越高表示基因间的相关性越强。选择与 lncRNA 连通性（weight）大于 0.3 的蛋白编码基因，通过 R 包 clusterProfiler v3.12.0（Yu *et al.*，2012）对其进行 GO 功能注释和 KEGG 通路富集分析，阈值设置为：*P*-value<0.05，FDR<0.05。

1.7 Hub lncRNA 筛选及网络构建

基因显著性（Gene significance，GS）表征基因与样本性状之间的相关性，绝对值越接近 1，代表这个基因与处理之间的关系越强。模块成员（Module membership，MM）表征基因与模块之间的相关性，绝对值越接近 1，代表这个基因与这个模块相关程度越高。计算 GS 和 MM，以如下标准筛选 hub lncRNA：|GS|>0.4，|MM|>0.8，p.GS<0.05，p.MM<0.05。利用 Cytoscape v3.7.2（Otasek *et al.*，2019）对 hub lncRNA 及与其连通性大于 0.3 的蛋白编码基因的共表达数据进行可视化，构建共表达网络。网络中每个节点代表一个基因，边代表基因间存在相互调控关系。

2 结果与分析

2.1 非生物胁迫下黑腹果蝇转录组中 lncRNA 鉴定及特征分析

将与黑腹果蝇非生物胁迫相关的 901 个 RNA-seq 数据进行比对拼接，共得到 17 392 个

基因，其中包括 13 780 个蛋白编码基因，3 612 个 lncRNA 基因，其中 1 179 个为新发现的 lncRNA 基因，对应 1 898 个转录本(包含基因间区 lncRNA 841 条，反义 lncRNA 594 条，内含子 lncRNA 463 条)。

分析发现，82.32%的 lncRNA 只含有两个及以下外显子，而在蛋白编码基因中只有 24.09%的基因只含有两个及以下的外显子（图 2：A）。lncRNA 的平均外显子长度（497 bp）显著长于蛋白编码基因（482 bp）(P<0.05)（图 2：B）。LncRNA 基因的可变剪切数目较少，15.09%的 lncRNA 发生了可变剪切，而有 38.63%的蛋白编码基因发生了选择性可变剪切（图 2：C）。LncRNA 的平均转录本长度（960 bp）显著短于蛋白编码基因（2 942 bp）(P<0.05)（图 2：D）。

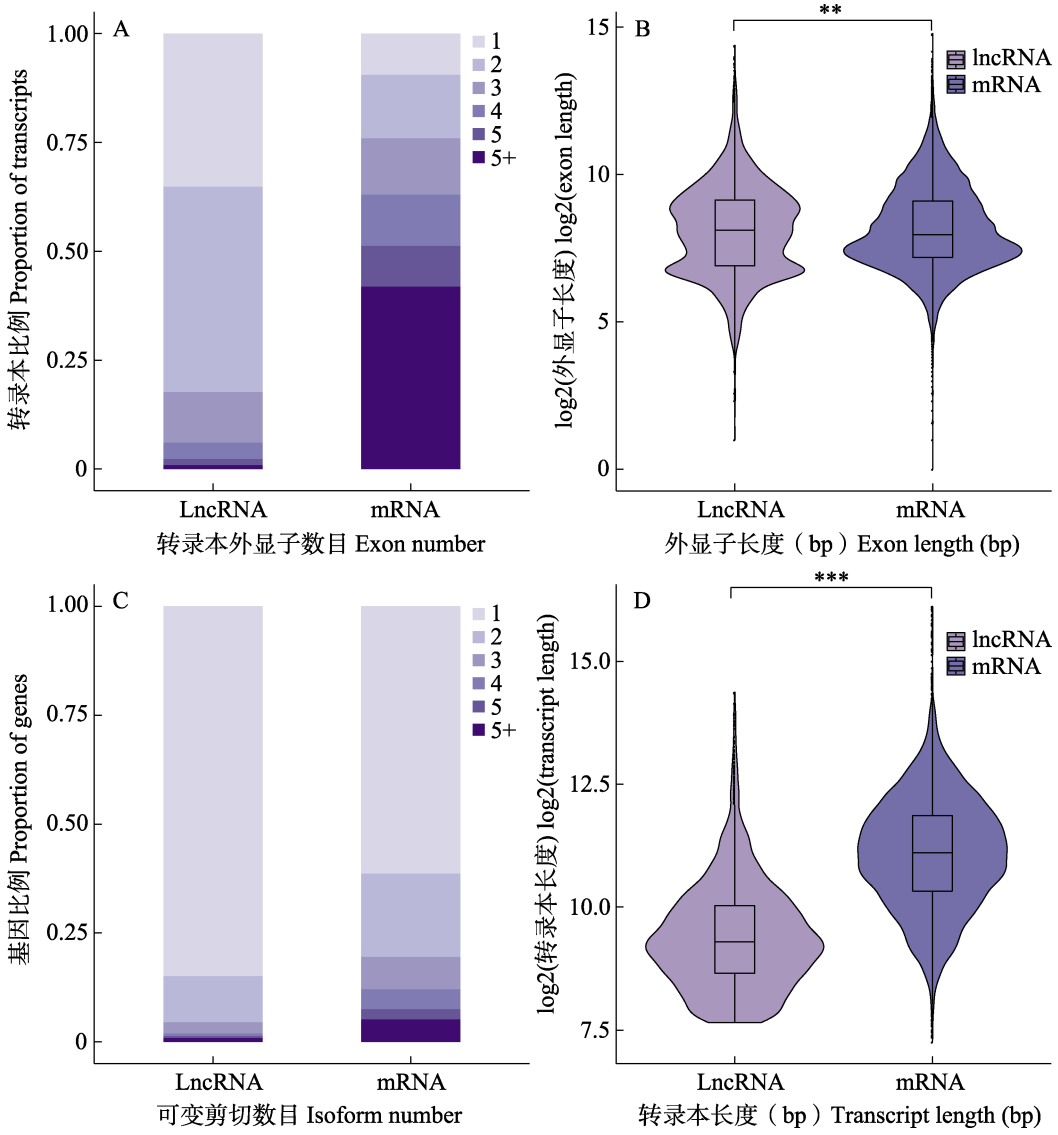


图 2　LncRNA 的特征统计

Fig. 2　The characters of lncRNAs

A. 每个 lncRNA、蛋白编码基因的转录本对应的外显子数目；B. LncRNA、蛋白编码基因的外显子的长度；C. 每个 lncRNA、蛋白编码基因对应的可变剪切数目；D. LncRNA、蛋白编码基因的转录本长度。**代表 0.001≤P<0.01，***代表 P<0.001。

A. The exon number per transcript of lncRNA and protein coding genes; B. The exon length of lncRNA and protein coding genes; C. The isoform counts per lncRNA and protein coding gene; D. The transcript length of lncRNA and protein coding genes. ** stands for 0.001≤P<0.01, *** stands for P<0.001.

2.2 饥饿应激下 lncRNA 基因差异表达情况

比较果蝇在不同时间饥饿胁迫下的 lncRNA 基因表达情况，发现与正常喂食相比，在 4、12、16 和 24 h 饥饿胁迫下，各有 260、28、200 和 26 个 lncRNA 基因表达显著上调，133、28、148 和 25 个 lncRNA 基因表达显著下调（adjusted *P* value < 0.05，|log2 (fold change)（处理组/对照组）|> 1），以下简称为饥饿相关 lncRNA。这些 lncRNA 可能分别参与了不同时间饥饿下果蝇应激反应的调控。

2.3 差异表达 lncRNA 的定量 PCR 验证

随机分别选取 4、12、16、24 h 饥饿处理下的 2 个上调及 2 个下调的差异表达 lncRNA（共 16 个），利用 NCBI primer 设计特异性引物（表 1），进行定量 PCR 验证。结果表明，15 个 lncRNA 基因表达变化趋势与预测结果一致（93.75%），其中 5 个 lncRNA 的差异表达结果具有显著性（图 4），说明 lncRNA 基因差异表达分析结果具有一定的可靠性。

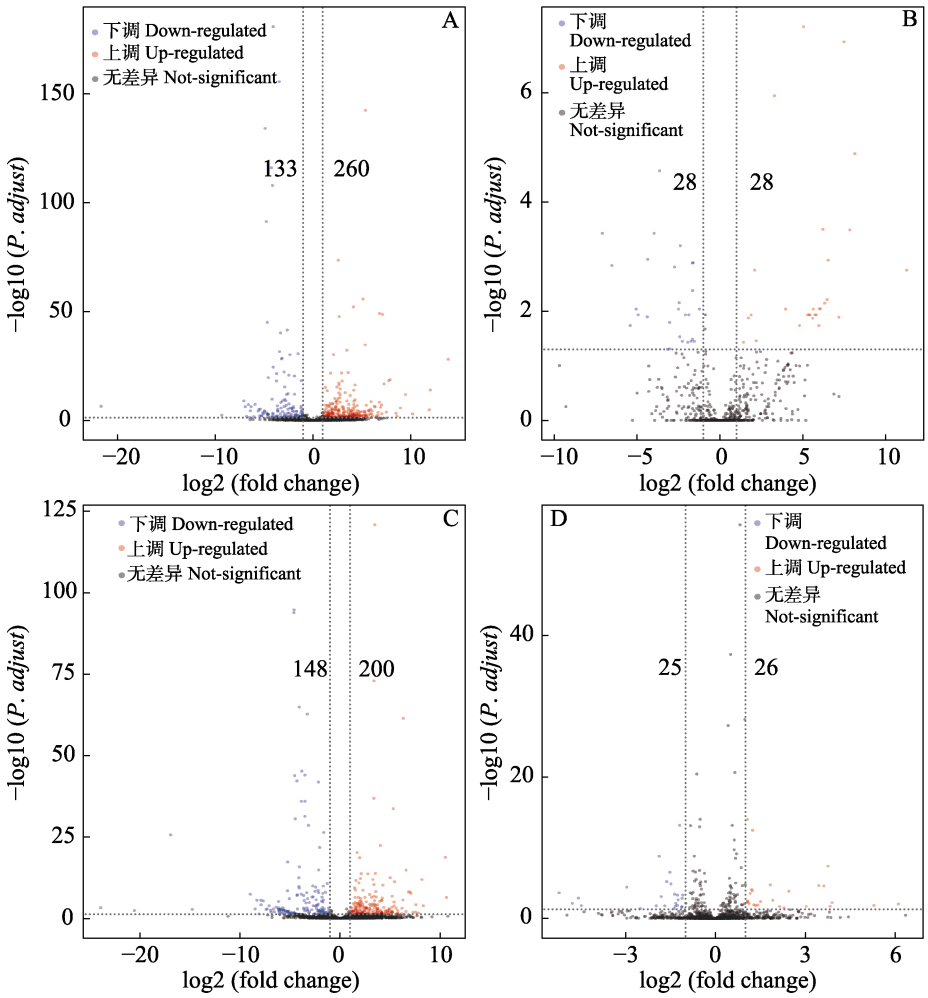


图 3 饥饿胁迫下 lncRNA 的差异表达火山图

Fig. 3 Volcano maps of differentially expressed lncRNA under starvation stress

A. 4 h 饥饿胁迫下 lncRNA 的差异表达火山图；B. 12 h 饥饿胁迫下 lncRNA 的差异表达火山图；C. 16 h 饥饿胁迫下 lncRNA 的差异表达火山图；D. 24 h 饥饿胁迫下 lncRNA 的差异表达火山图。图中标注的数字为差异表达的 lncRNA 数目。

A. Volcano map of differentially expressed lncRNA under 4 h starvation stress; B. Volcano map of differentially expressed lncRNA under 12 h starvation stress; C. Volcano map of differentially expressed lncRNA under 16 h starvation stress; D. Volcano map of differentially expressed lncRNA under 24 h starvation stress. The numbers marked in the figures are the numbers of differentially expressed lncRNAs.

表 1　LncRNA 基因荧光定量 PCR 引物序列信息

Table 1　Information of lncRNA primers used for qPCR

差异表达情况 Results predicted by DEseq2		引物名称 Primer name	引物序列（5′-3′） Primer sequence
内参基因 Reference gene		Tubulin-F	TGTCGCGTGTGAAACACTTC
		Tubulin-R	AGCAGGCGTTTCCAATCTG
4 h 饥饿 4 h starvation	上调 Up-regulated	FBgn0264602-F	CGAATGGGTGGGTTTTGCAG
		FBgn0264602-R	TCTGGAGGCCGAGATACACA
		FBgn0266974-F	CACCTGGGAAGCACAAGTGA
		FBgn0266974-R	TGGGTTGCTGGGTGATTCAA
	下调 Down-regulated	FBgn0259936-F	TGTTGCTAGCCTACGTGCTT
		FBgn0259936-R	CAAGGGAGCCAAGTGTGGAA
		FBgn0045800-F	CCACACGTGTCTCAAAGTAAGA
		FBgn0045800-R	CTTAAGCAGACGGCCAGGAT
12 h 饥饿 12 h starvation	上调 Up-regulated	FBgn0052660-F	TCGAACAAACGGAGCTGTCA
		FBgn0052660-R	GCAATCCGTGCCATTGGTTT
		FBgn0266645-F	TTACCATGAAGTGCCCCTGAC
		FBgn0266645-R	GACGACAAATGGCGGAGAAC
	下调 Down-regulated	FBgn0267613-F	CAACCGCCTGGTAGGAAGAG
		FBgn0267613-R	CATATCGGGGAGTGTCTGCC
		FBgn0264874-F	CATCGACGCTGTGGGATTCT
		FBgn0264874-R	TCCCGCACTTGATTTGTTGAC
16 h 饥饿 16 h starvation	上调 Up-regulated	MSTRG.2861-F	GTCTGCCTCCTCCGATTTCC
		MSTRG.2861-R	AATTCGTCCGAAGTGCCGAT
		FBgn0265074-F	TGAGACCAGGCGATTTGCAT
		FBgn0265074-R	TTATTTTCGCAACCGCAGCC
	下调 Down-regulated	FBgn0259993-F	AACCGTCTGTGAACAGCGTC
		FBgn0259993-R	GTGACGTAGGCGACACTTGA
		FBgn0050055-F	TTACCGGTGAGCCAAAATGC
		FBgn0050055-R	TGGTCGGAACGAAGTTGAAGT
24 h 饥饿 24 h starvation	上调 Up-regulated	MSTRG.16731-F	TCCCATGGCCCCCATAATTG
		MSTRG.16731-R	CCGATAGTCGTCGGAGTGTG
		FBgn0266258-F	AGGGCGTATGAGTAATGCGG
		FBgn0266258-R	TCACGTCACGTTTTGCGAAC
	下调 Down-regulated	FBgn0267987-F	AGCATACCATTTTGCGGTGC
		FBgn0267987-R	ACCTTGACGAGATTGCCGTT
		FBgn0051044-F	GAACTTGGAACTGCAACGCC
		FBgn0051044-R	GTTGGGTGCCCTGAATGTTG

编号以 MSTRG 开头的基因是通过图 1 所示流程鉴定的新 lncRNA。表 3 同。

The genes whose numbers start with MSTRG are novel lncRNAs predicted by the pipeline shown in Fig. 1. The same as table 3.

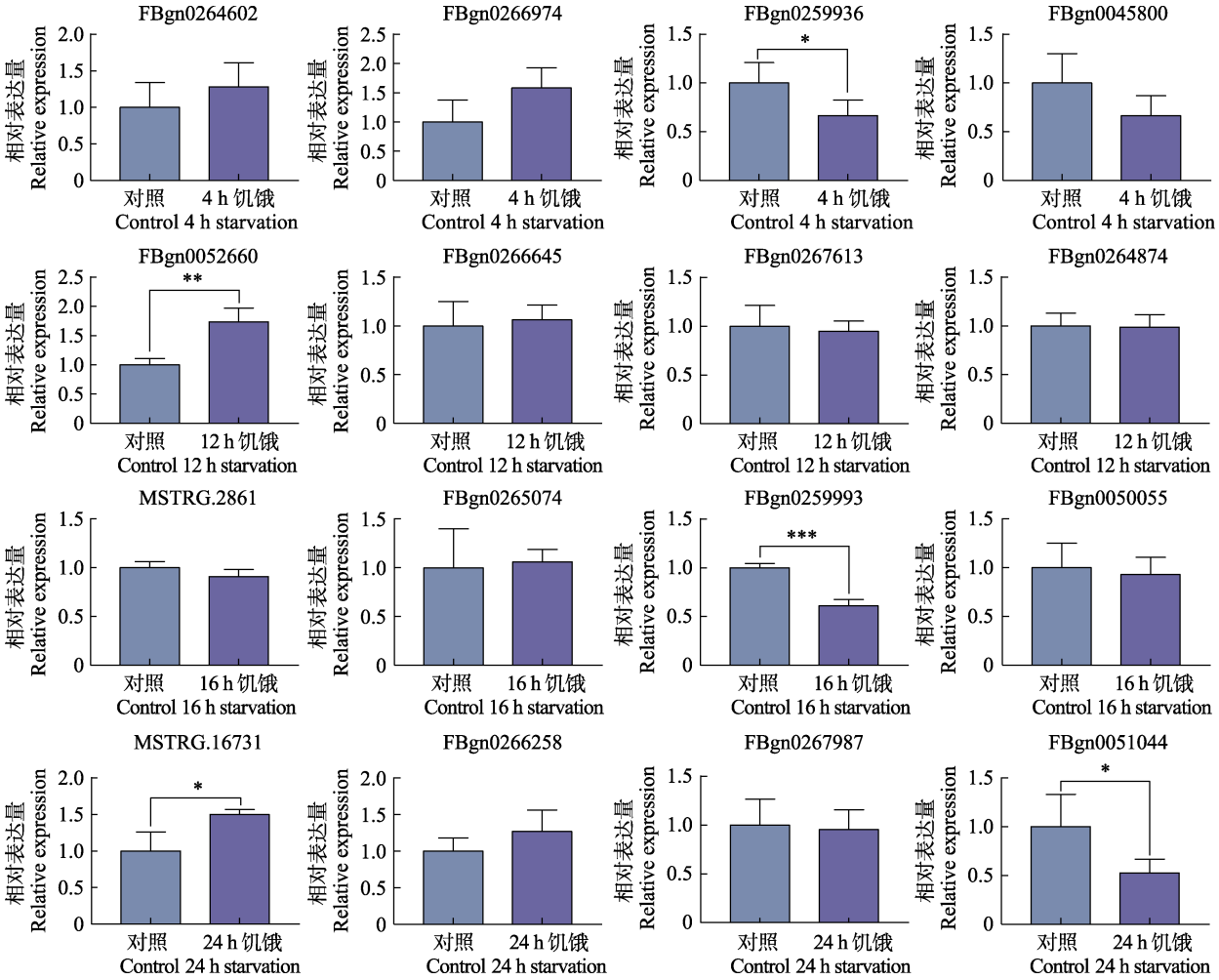


图 4　饥饿处理下黑腹果蝇差异表达 lncRNA 基因的荧光定量 PCR 验证结果

Fig. 4　The quantitative real-time PCR results of differentially expressed lncRNAs in *Drosophila melanogaster* under starvation

数据差异显著性分析采用 Student's *t*-test 检验。*代表 $0.01 \leqslant P<0.05$，**代表 $0.001 \leqslant P<0.01$，***代表 $P<0.001$。

Gene expression is analyzed by Student's *t*-test. * stands for $0.01 \leqslant P<0.05$, ** stands for $0.001 \leqslant P<0.01$, *** stands for $P<0.001$.

2.4　WGCNA 模块分析

为进一步探究饥饿相关 lncRNA 参与调控的基因网络，利用 WGCNA 共表达分析基因间的关联性，以 β=6（R^2=0.85）进行网络构建，聚类形成 22 个模块（图 5：A）。

以相关系数大于 0.6 筛选与饥饿胁迫高度相关的模块，发现与 4 h 饥饿处理高度相关基因模块分别为 royalblue 模块（R=0.74，$P=6\times10^{-7}$）和 pink 模块（R=0.85，$P=3\times10^{-10}$），分别包含了 1 个和 86 个饥饿相关 lncRNA 基因；与 12 h 饥饿处理高度相关基因模块为 darkturquoise 模块（R=0.7，$P=4\times10^{-6}$）和 blue 模块（R=0.72，$P=2\times10^{-6}$），分别包含了 1 个和 5 个饥饿相关 lncRNA 基因；与 16 h 饥饿处理高度相关基因模块为 green（R=0.78，$P=5\times10^{-8}$）模块，包含了 26 个饥饿相关 lncRNA 基因；与 24 h 饥饿处理高度相关基因模块为 turquoise 模块（R=0.61，$P=1\times10^{-4}$），包含了 20 个饥饿相关 lncRNA 基因（图 5：B，表 2），高度相关模块中的基因很有可能直接参与响应饥饿胁迫下的应激调控。

2.5　GO 功能注释和 KEGG 通路富集分析

LncRNA 通常被认为作用于与其表达密切

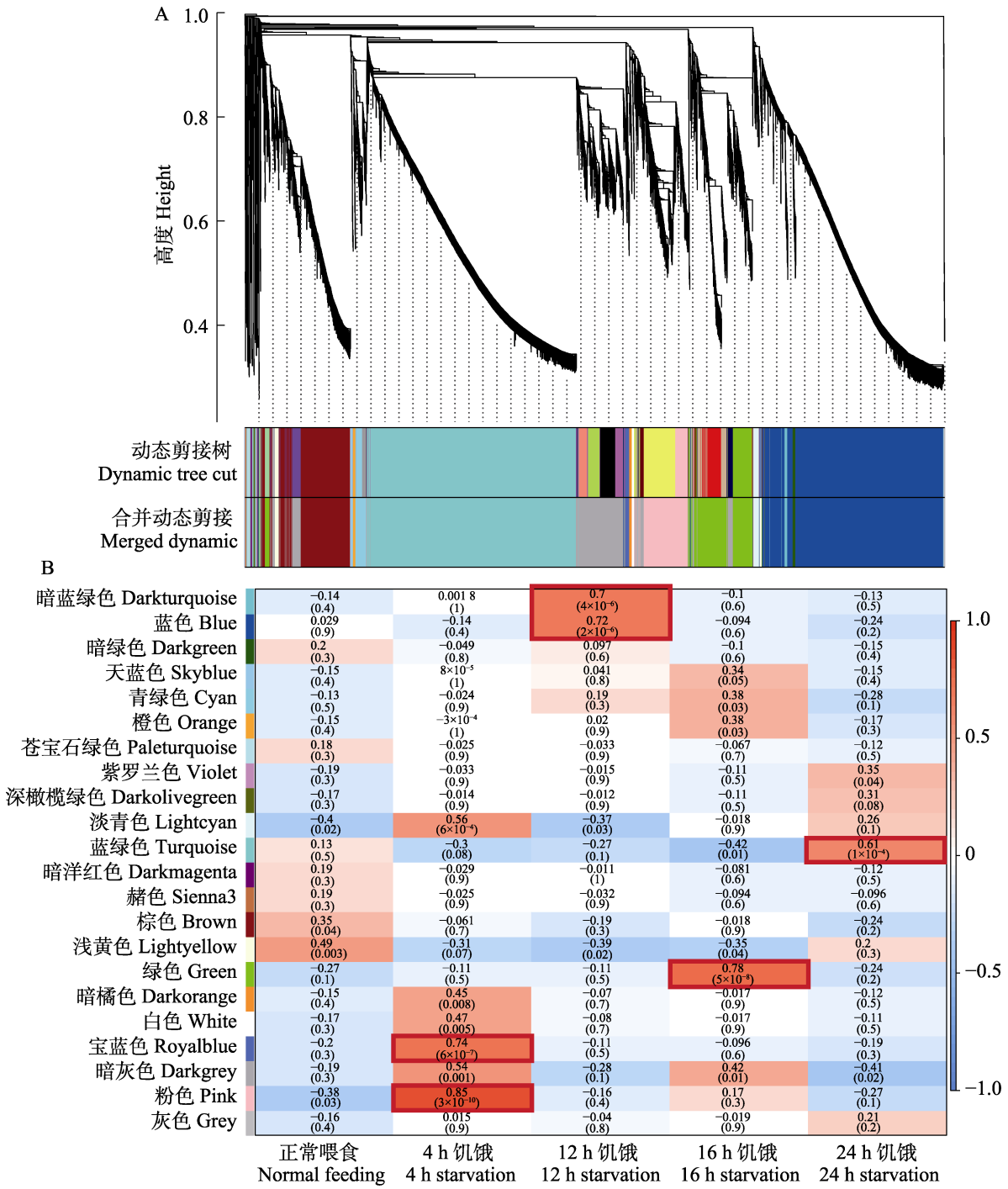


图 5　WGCNA 基因聚类及模块特征图

Fig. 5　WGCNA gene clustering and module feature map

A. 基因聚类图。基因的树状图聚类，以及原始模块和合并的模块，每个模块用不同颜色表示，灰色为未分类到具体有意义模块的基因；B. 模块特征关联热图。每行对应一个模块，每列对应一个处理，每个方形包含相应的相关性及 *P* 值。(红框选中的为 6 个与饥饿胁迫高度相关的模块)。

A. Clustering dendrogram of genes, with dissimilarity based on topological overlap, together with assigned merged module colors and the original module colors. Each module is represented by a color, and gray module includes genes that had not been classified into a specific meaningful module. B. Module-trait associations. Each row corresponds to a module eigengene, column to a trait. Each cell contains the corresponding correlation and *P*-value. The table is color-coded by correlation according to the color legend. (The modules selected by red frame are highly related to starvation stress.)

表 2 与饥饿胁迫高度相关的 6 个模块中的 lncRNA 情况

Table 2 The lncRNA in the six modules highly related to starvation stress

模块 Module	模块中基因数量 Number of genes in module	模块中 lncRNA 数目 Number of lncRNA in module	饥饿相关 lncRNA 数目* Number of starvation-related lncRNA
宝蓝色 Royalblue	91	31	1
粉色 Pink	1 162	239	86
暗蓝绿色 Darkturquoise	77	51	1
蓝色 Blue	4 375	666	5
绿色 Green	1 509	431	26
蓝绿色 Turquoise	5 213	1 062	20

*：与蛋白编码基因连通性大于 0.3 的饥饿相关 lncRNA。

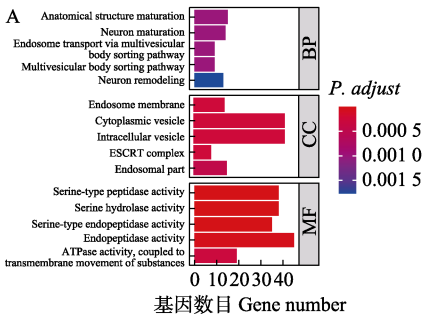
*: Starvation-related lncRNAs with a protein-coding gene weight greater than 0.3.

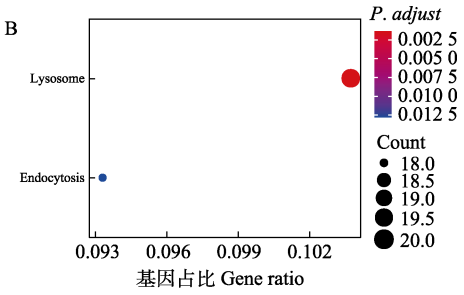
相关的蛋白编码基因。为探究目标模块中饥饿相关 lncRNA 的功能，筛选与饥饿相关 lncRNA 具有表达相关性的蛋白编码基因（连通性大于 0.3），并通过 GO 功能注释和 KEGG 通路富集对这些基因进行分析。

研究发现，饥饿胁迫 4 h 后，与饥饿相关 lncRNA 具有表达相关性的蛋白编码基因显著富集在溶酶体及胞吞作用等通路，参与了内体转运、多囊体分选、单磷酸核苷代谢、嘌呤单磷酸核苷合成、ATP 合成耦合质子传输等生物学过程（*P.adjust* < 0.05）（图 6：A-B）；饥饿胁迫 12 h 后，显著富集在吞噬体、丙酮酸代谢、糖酵解、柠檬酸循环等通路，参与嘌呤二磷酸核苷代谢、核苷酸磷酸化、ADP 代谢、脂肪酸延伸、长链脂肪酸代谢等生物学过程（*P.adjust* < 0.05）（图 6：C，D）；饥饿胁迫 16 h 后，显著富集在溶酶体、药物代谢-细胞色素 P450、戊糖和葡萄糖醛酸酯的相互转化、长寿调节途径等通路，参与了氨基糖代谢、含氨基葡萄糖的化合物代谢、化学刺激检测、味觉的感官知觉等生物学过程（*P*.adjust < 0.05）（图 6：E，F）；饥饿胁迫 24 h 后，显著富集在氧化磷酸化、磷酸肌醇代谢、长寿调节途径，以及 Wnt、Hippo、FoxO、mTOR、MAPK 等信号通路，参与了细胞间信号、突触信号、钾离子转运、昼夜节律、内吞、胞吐作用等生物学过程（*P.adjust* < 0.05）（图 6：G，H）。以上结果表明，饥饿相关 lncRNA 可能参与果蝇的能量代谢、信号传导以及细胞增殖凋亡生物学过程，并在不同时间饥饿胁迫下呈现动态变化，参与调控不同的生命过程。

2.6 Hub lncRNA 筛选及其调控网络

Hub lncRNA 是与性状关联度及与其他基因关联度高的基因，是 WGCNA 网络中的关键节点，更可能是功能网络的核心元件，对于整个共表达关系网络的稳定具有重要意义。为挖掘饥饿应激反应调控中的核心（hub）lncRNA，对不同时间饥饿胁迫下果蝇体内参与饥饿胁迫下应激调控的饥饿相关 lncRNA 进行筛选。发现，饥饿 4 h 后，有 66 个 hub lncRNA 在 royalblue 模块和 pink 模块发挥关键作用；饥饿 12 h 后，只有 1 个 hub lncRNA 在 blue 模块发挥关键作用，





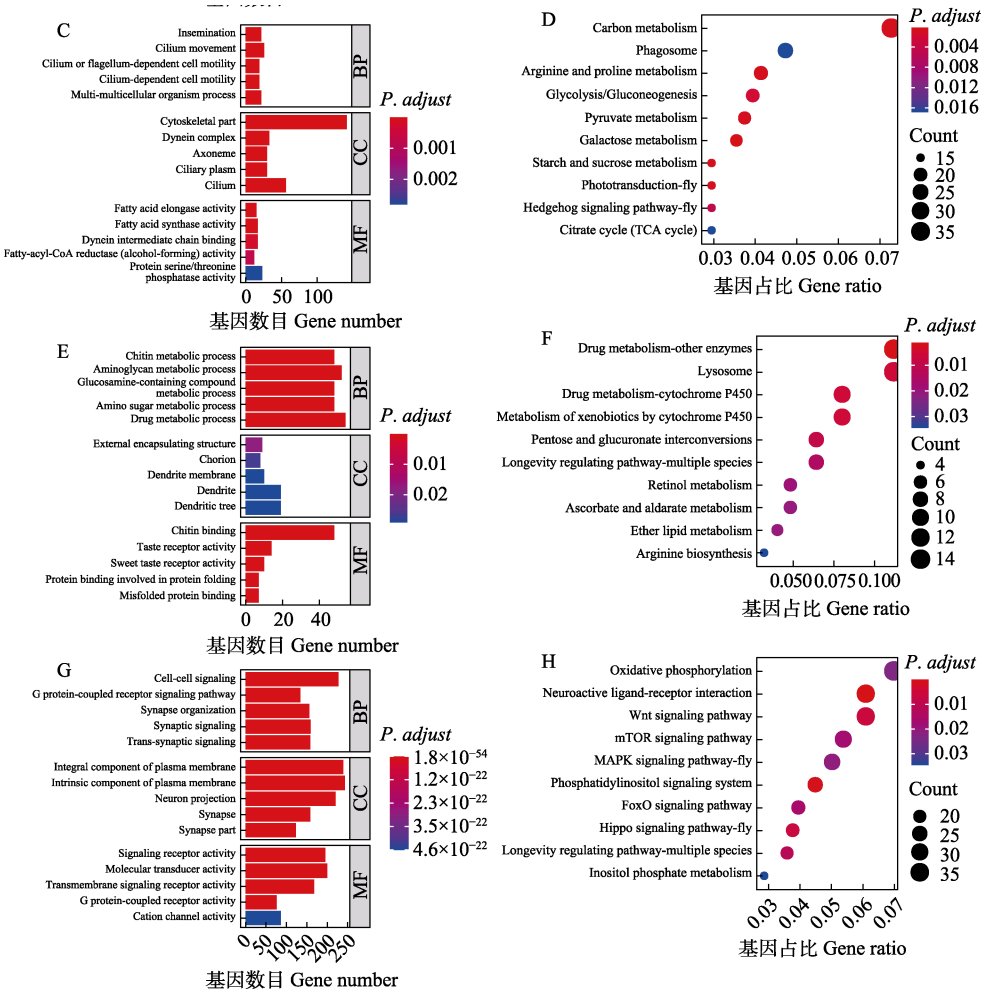


图 6　模块中与饥饿相关 lncRNA 具有表达相关性的蛋白编码基因 GO 和 KEGG 富集分析结果

Fig. 6　GO and KEGG enrichment results of protein coding genes related to starvation-related lncRNA in the module

A. 与饥饿 4 h 处理高度相关的 royalblue、pink 模块中与饥饿相关 lncRNA 具有表达相关性的蛋白编码基因的 GO 功能注释结果；B. 与饥饿 4 h 处理高度相关的 royalblue、pink 模块中与饥饿相关 lncRNA 具有表达相关性的蛋白编码基因的 KEGG 富集结果；C. 与饥饿 12 h 处理高度相关的 darkturquoise、blue 模块中与饥饿相关 lncRNA 具有表达相关性的蛋白编码基因的 GO 功能注释结果；D. 与饥饿 12 h 处理高度相关的 darkturquoise、blue 模块中与饥饿相关 lncRNA 具有表达相关性的蛋白编码基因的 KEGG 富集结果；E. 与饥饿 16 h 处理高度相关的 green 模块中与饥饿相关 lncRNA 具有表达相关性的蛋白编码基因的 GO 功能注释结果；F. 与饥饿 16 h 处理高度相关的 green 模块中与饥饿相关 lncRNA 具有表达相关性的蛋白编码基因的 KEGG 富集结果；G. 与饥饿 24 h 处理高度相关的 turquoise 模块中与饥饿相关 lncRNA 具有表达相关性的蛋白编码基因的 GO 功能注释结果；H. 与饥饿 24 h 处理高度相关的 turquoise 模块中与饥饿相关 lncRNA 具有表达相关性的蛋白编码基因的 KEGG 富集结果。

A. The GO function annotation result of protein coding genes related to starvation-related lncRNA in the royalblue and pink modules that are highly related to 4 h starvation; B. The KEGG enrichment result of protein coding genes related to starvation-related lncRNA in the royalblue and pink modules that are highly related to 4 h starvation; C. The GO function annotation result of protein coding genes related to starvation-related lncRNA in the darkturquoise and blue modules that are highly related to 12 h starvation; D. The KEGG enrichment result of protein coding genes related to starvation-related lncRNA in the darkturquoise and blue modules that are highly related to 12 h starvation; E. The GO function annotation result of protein coding genes related to starvation-related lncRNA in the green module that is highly related to 16 h starvation; F. The KEGG enrichment result of protein coding genes related to starvation-related lncRNA in the green module that is highly related to 16 h starvation; G. The GO function annotation result of protein coding genes related to starvation-related lncRNA in the turquoise module that is highly related to 24 h starvation; H. The KEGG enrichment result of protein coding genes related to starvation-related lncRNA in the turquoise module that is highly related to 24 h starvation.

dark turquoise 无满足条件的 hub lncRNA；在饥饿 16 h 后，则有 23 个 hub lncRNA 在 green 模块发挥关键作用；饥饿 24 h 后，只有 3 个 hub lncRNA 在 turquoise 模块发挥关键作用（表 3）。

为明确 hub lncRNA 及与其具有表达相关性的蛋白编码基因间的互作关系，利用 Cytoscape 软件构建饥饿胁迫下关键 lncRNA 参与的复杂调控网络。结果表明，随着饥饿应激时间的增加，hub lncRNA 调控的基因网络也发生了剧烈改变，lncRNA 在饥饿不同时间阶段可能在不同的生物学过程中发挥关键作用（图 7）。

3 讨论

对于动物而言，食物短缺造成的饥饿胁迫是最为普遍的一种压力。动物必须应对饥饿造成的能量短缺以及进一步的生命威胁（Rion and Kawecki，2007）。大量的研究已发现黑腹果蝇在饥饿条件下会出现睡眠抑制等应激行为，引起基因和信号通路变化，如 *clock*、*foxo* 等表达的改变（Kramer *et al.*，2008；Keene *et al.*，2010；Masek *et al.*，2014；Slankster *et al.*，2020），但对 lncRNA 的相关研究较少。本研究将饥饿胁迫相关的应激反应研究扩展到了 lncRNA 领域，通过对公共数据库中黑腹果蝇非生物胁迫下转录组进行分析，对饥饿胁迫下的 lncRNA 的表达水平变化及调控过程进行了较为全面的分析。

差异表达分析发现，在 4、12、16 和 24 h 饥饿胁迫下分别有 393、56、348 和 51 个 lncRNA 相比对照组出现差异表达。12 h 饥饿胁迫条件下得到的差异表达 lncRNA 数目较少，可能是由于该样本分组包含的转录组文库较少所致。对饥饿胁迫处理相关模块中差异表达 lncRNA 密切相关的蛋白编码基因分析表明，在相对较短时间的饥饿胁迫（小于 16 h）下，对应的基因主要富集在吞噬体、溶酶体通路以及能量代谢相关的生物学过程中。吞噬体、溶酶体是自噬的重要组成，而自噬被认为是饥饿胁迫下重要的营养及能量的

表 3 hub lncRNA 列表

Table 3 hub lncRNA list

胁迫 Srtess	模块 Module	hub lncRNA	数目 Number
饥饿 4 h 4 h starvation	宝蓝色 Royalblue	FBgn0267666	1
	粉色 Pink	FBgn0052194, FBgn0058469, FBgn0085582, FBgn0262142, FBgn0262148, FBgn0262722, FBgn0262898, FBgn0263019, FBgn0263049, FBgn0263617, FBgn0263766, FBgn0264713, FBgn0264822, FBgn0264874, FBgn0264879, FBgn0264947, FBgn0264982, FBgn0265074, FBgn0265075, FBgn0265254, FBgn0265311, FBgn0265338, FBgn0265542, FBgn0265786, FBgn0265798, FBgn0265810, FBgn0265843, FBgn0265961, FBgn0266041, FBgn0266143, FBgn0266252, FBgn0266349, FBgn0266408, FBgn0266631, FBgn0266680, FBgn0266735, FBgn0266740, FBgn0266807, FBgn0266829, FBgn0266880, FBgn0266912, FBgn0266973, FBgn0267152, FBgn0267153, FBgn0267208, FBgn0267459, FBgn0267581, FBgn0267587, FBgn0267591, FBgn0267621, FBgn0267639, FBgn0267769, MSTRG.10802, MSTRG.16102, MSTRG.17004, MSTRG.17007, MSTRG.17463, MSTRG.20261, MSTRG.20722, MSTRG.21450, MSTRG.3993, MSTRG.4693, MSTRG.5900, MSTRG.5943, MSTRG.6108	65
饥饿 12h 12 h starvation	蓝色 Blue	FBgn0267534	1
饥饿 16 h 16 h starvation	绿色 Green	FBgn0259841, FBgn0259846, FBgn0259853, FBgn0263660, FBgn0266218, FBgn0266275, FBgn0266681, FBgn0266929, FBgn0267321, FBgn0267445, FBgn0267809, FBgn0267924, FBgn0286827, MSTRG.13138, MSTRG.20870, MSTRG.21643, MSTRG.21959, MSTRG.2861, MSTRG.4622, MSTRG.5743, MSTRG.6021, MSTRG.7078, MSTRG.7772	23
饥饿 24 h 24 h starvation	蓝绿色 Turquoise	FBgn0264724, FBgn0266963, MSTRG.17611	3

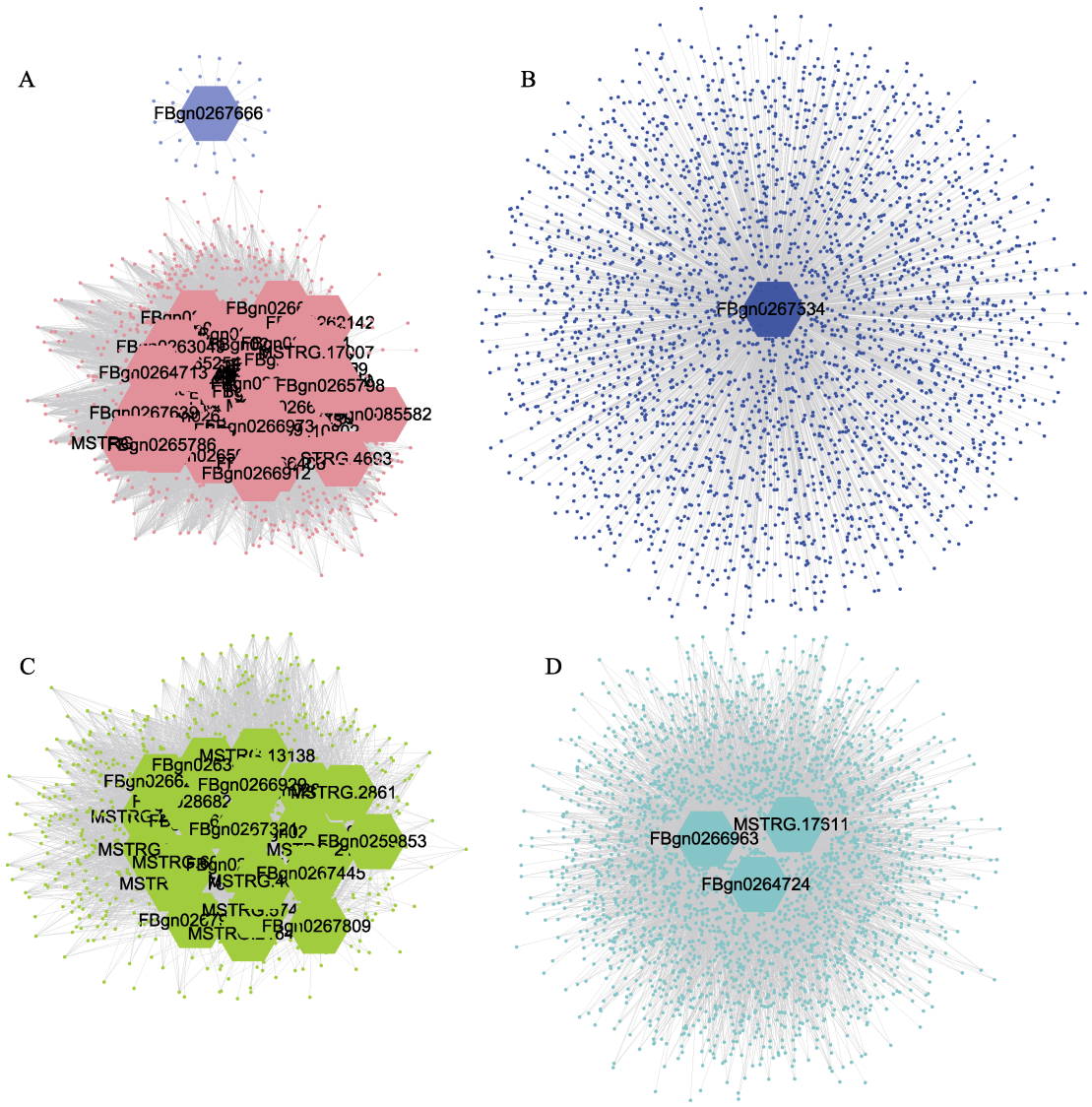


图 7　hub lncRNA 调控网络图

Fig. 7　Diagram of hub lncRNA regulatory network

A. 与 4 h 饥饿处理高度相关的 royalblue、pink 模块中的 hub lncRNA 调控网络；B. 与 12 h 饥饿处理高度相关的 blue 模块中的 hub lncRNA 调控网络；C. 与 16 h 饥饿处理高度相关的 green 模块中的 hub lncRNA 调控网络；D. 与 24 h 饥饿处理高度相关的 turquoise 模块中的 hub lncRNA 调控网络。网络中每个节点代表一个基因，一个小圆点代表一个蛋白编码基因，一个六边形代表一个 hub lncRNA，节点的大小在此图中无具体含义，节点的颜色为所在模块名字的颜色。

A. The hub lncRNA regulatory network in the royalblue and pink modules that are highly related to 4 h starvation stress; B. The hub lncRNA regulatory network in the blue module is highly related to 12 h starvation stress; C. The hub lncRNA regulatory network in the green module that is highly related to 16 h starvation stress; D. The hub lncRNA regulatory network in the turquoise module that is highly related to 24 h starvation stress. Each node in the network represents a gene, a small dot represents a protein coding gene, and a hexagon represents a hub lncRNA. The size of the dot has no specific meaning in the figure, and the color of the dot is the color of the module name.

来源（Scott *et al.*，2004；Rabinowitz and White，2010）。推测在此期间 lncRNA 可能参与到果蝇的自噬中，与能量代谢紧密相关。而在相对长时间的饥饿胁迫（16 h 和 24 h）下，对应的基因主要富集在长寿调节途径以及信号转导通路，同时也包含响应外界刺激及昼夜节律等相关的生物学过程，表明 lncRNA 通过调控细胞增殖及凋亡相关的途径，并刺激果蝇的觅食行为。由此可推

测，在饥饿胁迫下，黑腹果蝇中饥饿相关 lncRNA 参与的调控可能从短时间内的通过自噬等生命过程影响果蝇体内的能量代谢，保证能量供应，到相对长时间的饥饿胁迫下的提高对外界刺激的感受以及信号传导来促进觅食行为、维持生命存活。

本研究利用 lncRNA 与蛋白编码基因的表达相关性，借助与 lncRNA 表达相关的蛋白编码基因来推测 lncRNA 的调控功能。但由于 lncRNA 真实作用方式及调控过程十分复杂，涉及染色质修饰、转录调控、转录后调控等多种调控水平（Mercer *et al.*，2009；Pelechano and Steinmetz，2013；Engreitz *et al.*，2016；Joung *et al.*，2017），lncRNA 在细胞中的具体调控方式难以从转录组层面进行完全解析。此外，与 lncRNA 表达相关性不高的蛋白编码基因也有可能直接或间接地受到 lncRNA 的调控，而本研究中较为严格的筛选方式可能导致最终得到的调控网络存在一定的局限性，需要后续更深入分析以全面解析 lncRNA 在饥饿应激反应中的功能。

综上所述，本研究发现了在饥饿胁迫下起关键作用的 hub lncRNA，表明 lncRNA 可能参与黑腹果蝇的细胞内自噬、信号传导、细胞增殖及凋亡等应激生理过程，与已知的饥饿应激反应结论相似，为深入解析 lncRNA 在饥饿应激反应中的作用机理提供了重要的基础和数据支撑。

参考文献 (References)

Bolger AM, Lohse M, Usadel B, 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15): 2114–2120.

Buchfink B, Xie C, Huson DH, 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1): 59–60.

Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, Lander ES, 2016. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, 539(7629): 452–455.

Finn RD, Clements J, Eddy SR, 2011. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.*, 39(Web server issue): W29–W37.

Joung J, Engreitz JM, Konermann S, Abudayyeh OO, Verdine VK, Aguet F, Gootenberg JS, Sanjana NE, Wright JB, Fulco CP, Tseng YY, Yoon CH, Boehm JS, Lander ES, Zhang F, 2017. Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature*, 548(7667): 343–346.

Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, Griffiths-Jones S, Toffano-Nioche C, Gautheret D, Weinberg Z, Rivas E, Eddy SR, Finn RD, Bateman A, Petrov AI, 2021. Rfam 14: Expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.*, 49(D1): D192–D200.

Keene AC, Duboue ER, McDonald DM, Dus M, Suh GS, Waddell S, Blau J, 2010. Clock and cycle limit starvation-induced sleep loss in *Drosophila*. *Curr. Biol.*, 20(13): 1209–1215.

Kramer JM, Slade JD, Staveley BE, 2008. *foxo* is required for resistance to amino acid starvation in *Drosophila*. *Genome*, 51(8): 668–672.

Kuma A, Hatano M, Matsui M, Yamamoto A, Nakaya H, Yoshimori T, Ohsumi Y, Tokuhisa T, Mizushima N, 2004. The role of autophagy during the early neonatal starvation period. *Nature*, 432(7020): 1032–1036.

Li M, Wen S, Guo X, Bai B, Gong Z, Liu X, Wang Y, Zhou Y, Chen X, Liu L, Chen R, 2012. The novel long non-coding RNA CRG regulates *Drosophila locomotor* behavior. *Nucleic Acids Res.*, 40(22): 11714–11727.

Love MI, Huber W, Anders S, 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, 15(12): 550.

Masek P, Reynolds LA, Bollinger WL, Moody C, Mehta A, Murakami K, Yoshizawa M, Gibbs AG, Keene AC, 2014. Altered regulation of sleep and feeding contributes to starvation resistance in *Drosophila melanogaster*. *J. Exp. Biol.*, 217(17): 3122–3132.

Mercer TR, Dinger ME, Mattick JS, 2009. Long non-coding RNAs: Insights into functions. *Nature Reviews Genetics*, 10(3): 155–159.

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A, 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res.*, 49(D1): D412–D419.

Murakami K, Yurgel ME, Stahl BA, Masek P, Mehta A, Heidker R, Bollinger W, Gingras RM, Kim YJ, Ja WW, Suter B, DiAngelo JR, Keene AC, 2016. translin is required for metabolic regulation of sleep. *Curr. Biol.*, 26(7): 972–980.

Nawrocki EP, 2014. Annotating functional RNAs in genomes using Infernal. *Methods Mol. Biol.*, 1097: 163–197.

Otasek D, Morris JH, Boucas J, Pico AR, Demchak B, 2019. Cytoscape Automation: Empowering workflow-based network analysis. *Genome Biol.*, 20(1): 185.

Pease B, Borges AC, Bender W, 2013. Noncoding RNAs of the

Ultrabithorax domain of the *Drosophila* bithorax complex. *Genetics*, 195(4): 1253–1264.

Pelechano V, Steinmetz LM, 2013. Gene regulation by antisense transcription. *Nat. Rev. Genet.*, 14(12): 880–893.

Pertea G, Pertea M, 2020. GFF Utilities: GffRead and GffCompare. *F1000Research*, 9:304.

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL, 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.*, 11(9): 1650–1667.

Ponton F, Chapuis MP, Pernice M, Sword GA, Simpson SJ, 2011. Evaluation of potential reference genes for reverse transcription-qPCR studies of physiological responses in *Drosophila melanogaster*. *J. Insect Physiol.*, 57(6): 840–850.

Rabinowitz JD, White E, 2010. Autophagy and metabolism. *Science*, 330(6009): 1344–1348.

Rion S, Kawecki TJ, 2007. Evolutionary biology of starvation resistance: What we have learned from *Drosophila*. *J. Evol. Biol.*, 20(5): 1655–1664.

Root CM, Ko KI, Jafari A, Wang JW, 2011. Presynaptic facilitation by neuropeptide signaling mediates odor-driven food search. *Cell*, 145(1): 133–144.

Sarropoulos I, Marin R, Cardoso-Moreira M, Kaessmann H, 2019. Developmental dynamics of lncRNAs across mammalian organs and species. *Nature*, 571(7766): 510–514.

Sawin ER, Ranganathan R, Horvitz HR, 2000. *C. elegans* locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. *Neuron*, 26(3): 619–631.

Schmidt MH, 2014. The energy allocation function of sleep: A unifying theory of sleep, torpor, and continuous wakefulness. *Neurosci. Biobehav. Rev.*, 47: 122–153.

Scott RC, Schuldiner O, Neufeld TP, 2004. Role and regulation of starvation-induced autophagy in the *Drosophila* fat body. *Dev. Cell*, 7(2): 167–178.

Slankster E, Kollala S, Baria D, Dailey-Krempel B, Jain R, Odell SR, Mathew D, 2020. Mechanism underlying starvation-dependent modulation of olfactory behavior in *Drosophila* larva. *Sci. Rep.*, 10(1): 3119.

Soshnev AA, Ishimoto H, McAllister BF, Li X, Wehling MD, Kitamoto T, Geyer PK, 2011. A conserved long noncoding RNA affects sleep behavior in Drosophila. *Genetics*, 189(2): 455–468.

Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W, 2013. CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res.*, 41(6): e74.

Yang Z, Yu Y, Zhang V, Tian Y, Qi W, Wang L, 2015. Octopamine mediates starvation-induced hyperactivity in adult Drosophila. *Proc. Natl. Acad. Sci. USA*, 112(16): 5219–5224.

Yu G, Wang LG, Han Y, He QY, 2012. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS*, 16(5): 284–287.

Yu Y, Huang R, Ye J, Zhang V, Wu C, Cheng G, Jia J, Wang L, 2016. Regulation of starvation-induced hyperactivity by insulin and glucagon signaling in adult *Drosophila*. *Elife*, 5: e15693.

Zhang B, Horvath S, 2005. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.*, 4(1): 17.