



## 大黄鱼生长速率差异个体肌肉组织的转录组比较分析

张波, 姜丹, 张东玲, 王志勇, 方铭\*

(集美大学水产学院, 农业农村部东海海水健康养殖重点实验室, 福建 厦门 361021)

**摘要:** 为了探究大黄鱼生长性状的分子调控机制, 实验从同一个网箱内养殖的大黄鱼中选取生长速率较快的体重均值为  $(527.1 \pm 83.6)$  g 的个体 182 尾、生长速率较慢的体重均值为  $(238.4 \pm 52.3)$  g 的个体 230 尾, 共 412 尾进行研究。对其肌肉组织进行总 RNA 提取和转录组测序, 并对 2 组进行基因差异表达分析。以  $P_{adj} < 0.05$ 、 $abs[\log_2(\text{FoldChange})] > 1$  为标准, 共筛选到 227 个差异表达基因, 包括 125 个上调基因、102 个下调基因。差异表达基因在 NCBI-nr 和 Uniprot (Swiss-Prot) 数据库的注释率分别为 99.12%、81.94%。通过差异表达基因的功能注释结果, 初步筛选出了可能与肌肉生长相关的候选功能基因, 如 *gdf9*、*ckm*、*tnni2* 和 *des* 等。GO 和 KEGG 分析预测出部分与肌肉生长相关的信息, 如 GO term 有肌动蛋白丝组织、肌动蛋白细胞骨架组织、肌动蛋白丝基过程、微管组织中心和发育生长等, KEGG 通路有细胞和生长因子、脂肪酸生物合成、转化生长因子- $\beta$  信号通路 (TGF- $\beta$  信号通路)、胰岛素信号通路和肌动蛋白细胞骨架的调控等。加权基因共表达网络分析 (WGCNA) 分析发现, purple 模块与体重性状相关性较强, 其核心基因 *myoz1*、*tpm1* 和 *tnni2* 可能与肌肉生长相关。本研究结果可以为后续相关基因功能的深度挖掘验证以及大黄鱼生长性状的分子调控机制解析提供参考。

**关键词:** 大黄鱼; 肌肉生长; 转录组; 差异表达基因; 加权基因共表达网络分析 (WGCNA)

中图分类号: Q 786; S 965.322

文献标志码: A

大黄鱼 (*Larimichthys crocea*) 属鲈形目 (Perciformes) 石首鱼科 (Sciaenidae) 黄鱼属 (*Larimichthys*)<sup>[1-2]</sup>, 俗称黄鱼、大黄花鱼、黄瓜鱼等, 与小黄鱼、带鱼、乌贼并称为我国传统的“海洋四大经济鱼类”<sup>[3]</sup>。其主要分布于南海、东海和黄海南部, 是目前我国海水网箱养殖单品种产量最大的养殖鱼类, 并且是当前养殖量最大的海水鱼<sup>[4]</sup>, 2020 年国内大黄鱼的总养殖产量达 25.40 万 t<sup>[5]</sup>。大黄鱼已在主产区形成极具特色的“大黄鱼文化”, 但是从 19 世纪 70 年代起野生大黄鱼资源开始严重衰退<sup>[6]</sup>。虽然我国在 1987 年已经建立起了适用于

大黄鱼的人工养殖与育苗技术<sup>[7]</sup>, 但是野生种群和种质资源仍未得到有效的保护, 并且经过 30 多年连续多代未加选育的人工繁殖, 由于近交以及不良的人工养殖条件的负选择作用等原因, 造成养殖大黄鱼群体的遗传性状退化、抗病力降低、生长缓慢和养殖成活率下降等问题, 严重影响了整个大黄鱼产业链的健康与可持续发展, 同时对生态环境也造成了一定的破坏。随着大黄鱼人工养殖育苗技术的成熟和规模化养殖产业的发展, 养殖户为了追求更大化的利益, 越来越关注大黄鱼的养殖周期即生长速率, 因此对大黄鱼生长性

收稿日期: 2021-08-10 修回日期: 2021-11-03

资助项目: 国家自然科学基金 (31672399, U1705231, 31872560, 32172964); 福建省自然科学基金 (2020J01672, 2021J02045)

第一作者: 张波 (照片), 从事水产动物遗传与育种研究, E-mail: 1725633321@qq.com

通信作者: 方铭, 从事基因组统计遗传学研究, E-mail: fangming618@126.com



状的相关研究也变得越来越迫切, 生长性状对于大黄鱼来说也就成了极其重要的经济性性状之一。提高大黄鱼的生长速率不仅可以缩短养殖周期, 节省养殖成本, 最终增加养殖产量, 在一定程度上还可以在给养殖户带来更大化的经济效益和生态效益, 同时满足人们对大黄鱼的食用需求。

然而, 生长性状是受基因、环境以及基因与环境交互作用的一种复杂数量性状, 但是目前关于该物种的分子遗传研究较少, 尤其是在生长性状方面。Ni 等<sup>[8]</sup>对大黄鱼生长激素基因的研究表明, 其第一、二个内含子内部存在与生长性状相关的单核苷酸多态性。薛良义等<sup>[9]</sup>对大黄鱼肌肉生长抑制素基因的研究表明, 大黄鱼肌肉生长抑制素基因第一个外显子内部存在与体长和体重均显著相关的单核苷酸及氨基酸多态性。对于研究者来说, 解析大黄鱼生长相关的遗传机制, 对通过标记辅助选择、基因组选择及基因编辑等方式对其生长性状进行遗传改良具有重要意义。

转录组分析是基因组研究和功能基因鉴定的一种非常有效的方法<sup>[10-11]</sup>。转录组学是从 RNA 水平上研究特定细胞或组织中基因转录情况及其转录调控规律的一类组学方法<sup>[12]</sup>, 是对细胞功能和生物表型进行研究的非常重要的一种手段。与基因组不同的是, 转录组具有时空表达的特异性, 因此转录组是研究生物生长发育、应激生理和抗病免疫等分子机制的强有力工具。通过转录组分析可以得到很多信息, 包括基因表达信息、可变剪切、反义转录本、基因结构、新基因、差异表达基因和单核苷酸多态性等, 可进一步对相关基因进行功能注释、GO 富集分析、KEGG 通路富集分析、蛋白互作网络分析、加权基因共表达网络分析 (weighted gene co-expression network analysis, WGCNA) 和单核苷酸多态性分析等<sup>[11]</sup>。WGCNA 是利用转录组基因表达量信息来构建无尺度分布基因共表达网络的一种技术, 该技术通过软阈值筛选来鉴定表达模式相似的基因集合模块 (module), 通过与表型性状之间的关联分析以找出与性状相关的共表达模块, 进而对该共表达模块中的核心基因进行深度挖掘和功能分析<sup>[13-14]</sup>。转录组测序技术在研究鱼类基因表达调控机制及挖掘控制鱼类特异性状的关键基因等方面发挥着极其重要的作用<sup>[15]</sup>。目前关于大黄鱼的转录组分析主要集中在抗病免疫功能<sup>[16-17]</sup>和性别分化<sup>[18]</sup>等方面。有研究表明, 鱼类的骨骼肌占鱼类总体重的

40%~50%, 并且肌肉的生长对从幼鱼到成鱼的发育有着重要影响<sup>[19]</sup>。已经发现的与鱼类生长相关的基因主要包括肌肉生长抑制素基因 (*mstn*) 和生肌调节因子基因 (*mrf*), *mrf* 家族包括 *myogenin*、*myoD*、*myf-5* 和 *myf-6* 等基因, 研究表明这些基因均参与肌肉的生成过程, 并且生肌调节因子基因只在脊椎动物的骨骼肌中表达<sup>[20-22]</sup>, 另有研究表明, 小清蛋白基因也存在于鱼类肌肉中并调控肌肉纤维的松弛<sup>[23]</sup>。鱼类生长相关的调控网络较为复杂, 除了上述已经被发现的直接与生长相关的基因外, 还有其他可能通过调节能量代谢或肌肉生长的生理网络而显著影响生长的基因也逐步被揭示出来<sup>[24]</sup>。

为了探究大黄鱼生长性状的分子调控机制并挖掘相关的功能基因, 实验从同一个网箱内养殖的大黄鱼中选取生长速率存在差异的个体进行比较转录组分析, 对差异表达基因进行功能注释、GO 和 KEGG 通路富集分析, 通过 WGCNA 来鉴定与大黄鱼体重性状相关的共表达模块并进行核心基因的深度挖掘, 以期获得可能与大黄鱼肌肉生长相关的功能基因, 为后续相关基因功能的深度挖掘验证以及大黄鱼生长性状的分子调控机制解析提供参考。

## 1 材料与方法

### 1.1 实验样品采集

本实验于 2019 年 12 月在福建省宁德市三都澳大黄鱼养殖基地同一个网箱内养殖的 2 龄大黄鱼中选取了生长速率较快的体重均值为 (527.1 ± 83.6) g 的个体 182 尾和生长速率较慢的体重均值为 (238.4 ± 52.3) g 的个体 230 尾, 共 412 尾。测量其体重后, 采集背部肌肉组织置于装有 RNA 保护液的冻存管中, 存于 -80 °C 超低温冰箱中备用。动物实验操作遵循集美大学水产学院动物伦理委员会的要求进行。

### 1.2 总 RNA 提取及 cDNA 文库构建和测序

取存于 -80 °C 超低温冰箱中备用的肌肉组织样品, 使用 TRIzol Reagent Kit (Invitrogen, 美国) 法提取总 RNA, 将提取好的总 RNA 送至北京诺禾致源科技股份有限公司, 使用琼脂糖凝胶电泳分析 RNA 的纯度和完整性、Nanodrop 检测 RNA 的纯度 (OD<sub>260/280</sub> 值)、Qubit 2.0 对 RNA 浓度进行

精确定量和 Agilent 2100 生物分析仪精确检测 RNA 的完整性。RNA 样品检测合格后, 用带有 Oligo (dT) 的磁珠富集真核生物 mRNA 来进行 cDNA 文库的构建。文库构建合格后使用 Illumina 高通量测序平台 NovaSeq 6000 进行 2×150 bp 双末端测序。测序完成后从公司获取原始数据 (raw data) 和质控后数据 (clean data) 进行下游分析。

### 1.3 转录组分析

将 raw data 通过 fastp 软件 (<https://github.com/OpenGene/fastp>)<sup>[25]</sup> 进行处理: 去接头 reads; 当任一测序 read 中 N 含量超过该 read 碱基数的 10% 时, 去除此 paired reads; 当任一测序 read 中含有的低质量 ( $Q \leq 5$ ) 碱基数超过该条 read 碱基数的 50% 时, 去除此 paired reads, 由此即可获得 clean data。在本实验中又使用 fastp 软件<sup>[25]</sup> 切除了测序前端不稳定的 10 bp 碱基和去除了长度低于 75 bp 的短片段 read, 经上述质量过滤后, 检查双端测序 reads, 去掉未配对的测序 read。

质控完成后使用 STAR 软件 (<https://github.com/alexdobin/STAR/>)<sup>[26]</sup> 将测序 reads 比对到大黄鱼参考基因组上 (参考基因组和注释文件均为本实验室所组装), 比对完成后使用 StringTie 软件 (<http://ccb.jhu.edu/software/stringtie/>)<sup>[27-29]</sup> 基于参考基因组注释文件对所有转录本进行整合组装, 并使用 TransDecoder 软件 (<https://github.com/TransDecoder/TransDecoder/releases>) 对所组装转录本的候选编码区进行鉴定, 最后使用 featureCounts 软件 (<http://bioinf.wehi.edu.au/featureCounts/>)<sup>[30]</sup> 进行基因表达的计数定量, 获得表达矩阵。

### 1.4 基因差异表达分析

进行基因差异表达分析分组前, 先对所有大黄鱼个体的体重表型进行性别矫正, 矫正方式为“离均差矫正”, 即“样本个体体重-相应性别组体重均值”, 以矫正后的数据作为分组依据, 小于 0 为生长慢个体组, 大于 0 为生长快个体组。使用 DESeq2 软件 (<http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html>)<sup>[31]</sup>, 以生长慢个体组为对照组, 生长快个体组为实验组进行基因差异表达分析, 以  $P_{adj} < 0.05$ 、 $abs[\log_2(\text{FoldChange})] > 1$  为标准筛选差异表达基因。

### 1.5 差异表达基因的功能注释

使用 diamond 软件 (<http://www.crystalimpart.com/diamond>)<sup>[32]</sup> 中的 blastx 功能将筛选得到的差异表达基因的最长转录本序列比对到 NCBI-nr 和 Uniprot (Swiss-Prot) 蛋白数据库中进行功能注释。为提高注释结果的可靠性, 进行功能注释时  $E$  值阈设定为  $1 \times 10^{-5}$ , 并且使用 --more-sensitive 参数来提高序列比对的敏感度。

com/diamond)<sup>[32]</sup> 中的 blastx 功能将筛选得到的差异表达基因的最长转录本序列比对到 NCBI-nr 和 Uniprot (Swiss-Prot) 蛋白数据库中进行功能注释。为提高注释结果的可靠性, 进行功能注释时  $E$  值阈设定为  $1 \times 10^{-5}$ , 并且使用 --more-sensitive 参数来提高序列比对的敏感度。

### 1.6 差异表达基因的 GO 和 KEGG 分析

使用基于 eggNOG 数据库开发的 eggNOG-mapper 工具 (<https://github.com/jhcepas/eggnog-mapper.git>)<sup>[33-35]</sup>, 利用组装所得的所有基因的氨基酸序列构建用于 GO 和 KEGG 通路富集分析的 OrgDb 包, 构建完成后使用 clusterProfiler 软件 (<http://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html>)<sup>[36]</sup> 对差异表达基因进行 GO 和 KEGG 通路富集分析。

### 1.7 加权基因共表达网络分析 (WGCNA)

使用 WGCNA 软件 (<https://CRAN.R-project.org/package=WGCNA>) 构建无尺度加权基因共表达网络<sup>[37-39]</sup>, 以上述经过 DESeq2 软件标准化后的基因表达量矩阵作为输入数据, 筛选出中位绝对偏差 (median absolute deviation, MAD) 前 75% 且至少 MAD 值大于 0.01 的基因以构建无尺度加权基因共表达网络, 使用样本聚类的方法检测是否存在离群样本点, 如有则去除这些离群样本, 使用软件中的 pickSoftThreshold 函数计算共表达网络的软阈值, 并使用 powerEstimate 函数估计最优软阈值, 使用一步法构建无尺度加权基因共表达网络并分析体重性状与各模块间的相关性, 鉴定与体重性状相关性较强的模块, 挖掘这些模块中的核心基因并通过 NCBI-nr 和 Uniprot (Swiss-Prot) 蛋白数据库进行功能注释。

### 1.8 实时荧光定量 PCR (qRT-PCR) 验证

为了验证转录组测序分析结果的可靠性, 以  $\beta$ -actin 为内参基因, 随机挑选 10 个差异表达基因 (上调表达 5 个, 下调表达 5 个) 进行 qRT-PCR 验证。qRT-PCR 所用系统为 QuantStudio™ 6 Flex System。实验用逆转录试剂盒是北京全式金生物技术股份有限公司的 TransScript All-in-One First-Strand cDNA SuperMix for qPCR (One-Step gDNA Removal), qRT-PCR 试剂盒是北京全式金生物技术股份有限公司的 PerfectStart Green qPCR Super-Mix。反应体系为 20  $\mu$ L: 4  $\mu$ L cDNA 模板, 10  $\mu$ L



2×PerfectStart Green qPCR SuperMix, 0.4 μL Passive Reference Dye (50×), 上下游引物各 0.4 μL (10 μmol/L), 4.8 μL Nuclease-free Water。反应采取两步法: 预变性 95 °C 3 min, 变性 95 °C 10 s, 退火、延伸 60 °C 30 s (此步骤中采集荧光), 共 40 个循环。

## 2 结果

### 2.1 转录组测序和组装结果

测序所得的 raw reads 均值为 15 708 595 条, clean reads 均值为 15 087 186 条, raw data 数据量均值为 4.71 GB, clean data 数据量均值为 4.53 GB, Q20 均值为 98.46%, Q30 均值为 95.15%, GC 含量均值为 50.68%。使用 StringTie 软件对测序数据进行有参组装, 组装后共获得 113 847 条转录本序列 (contigs), 相应的基因数 (unigene) 为 51 615

个, contigs 序列长度范围为 152~82 856 bp, 序列长度均值为 2 704.9 bp, N50 为 4 704。使用 TransDecoder 软件对组装转录本进行开放阅读框 (Open Reading Frame, ORF) 区域预测后共获得 94 482 条转录本序列 (contigs), 相应的基因数 (unigene) 为 24 849 个, contigs 序列长度范围为 297~82 856 bp, 序列长度均值为 4 521.5 bp, N50 为 5 896, 预测得到的 ORF 序列长度范围为 258~30 501 bp, 序列长度均值为 1 309.3 bp。

### 2.2 基因差异表达分析

根据  $P_{adj} < 0.05$ 、 $abs[\log_2(\text{FoldChange})] > 1$  的标准筛选生长速率较快个体组与较慢个体组间的差异表达基因, 共得到 227 个差异表达基因, 其中上调表达的有 125 个, 下调表达的有 102 个 (图 1)。

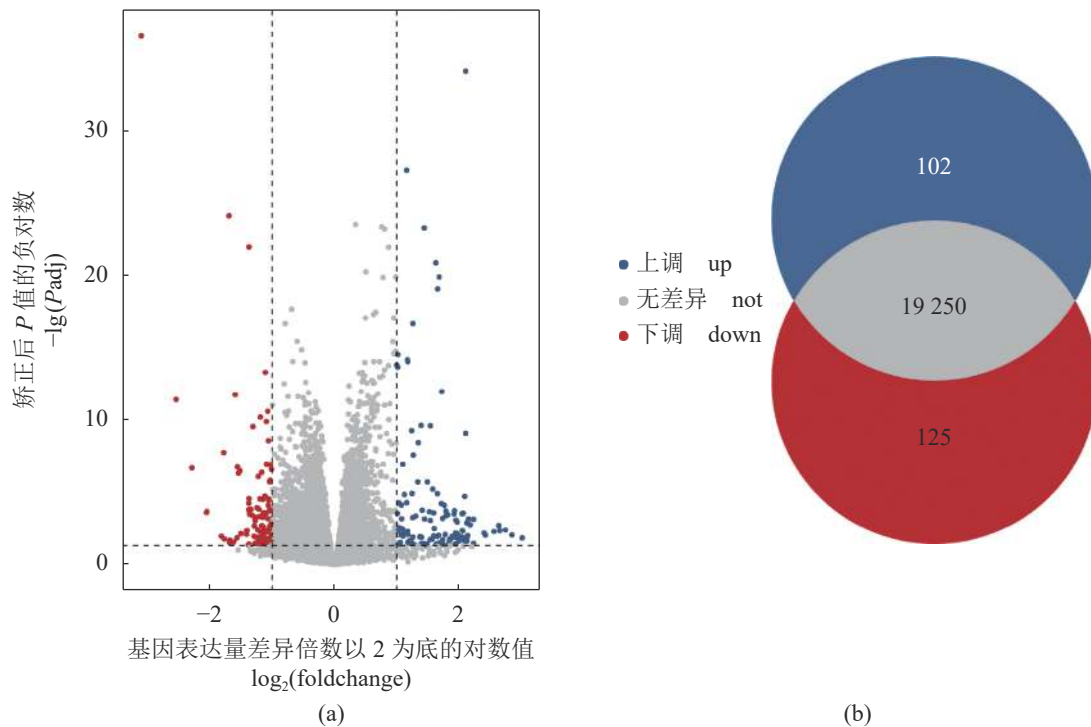


图 1 生长较快/慢大黄鱼肌肉组织基因差异表达分析

(a) 火山图, (b) 韦恩图; x 轴代表基因表达量差异倍数以 2 为底的对数值, y 轴代表矫正后 P 值的负对数值。

Fig. 1 Analysis of differential gene expression in muscle tissue of *L. crocea* with fast-growth rate and slow-growth rate

(a) volcano plot, (b) Venn plot; the x-axis represents the logarithm value of the difference multiple of gene expression with 2 as the base, the y-axis represents the negative logarithm value of the corrected P-value.

### 2.3 差异表达基因的功能注释

使用 diamond 软件中的 blastx 功能对 227 个差异表达基因进行功能注释, 在 NCBI-nr 非冗余蛋白数据库中得到功能注释的基因数为 225 个,

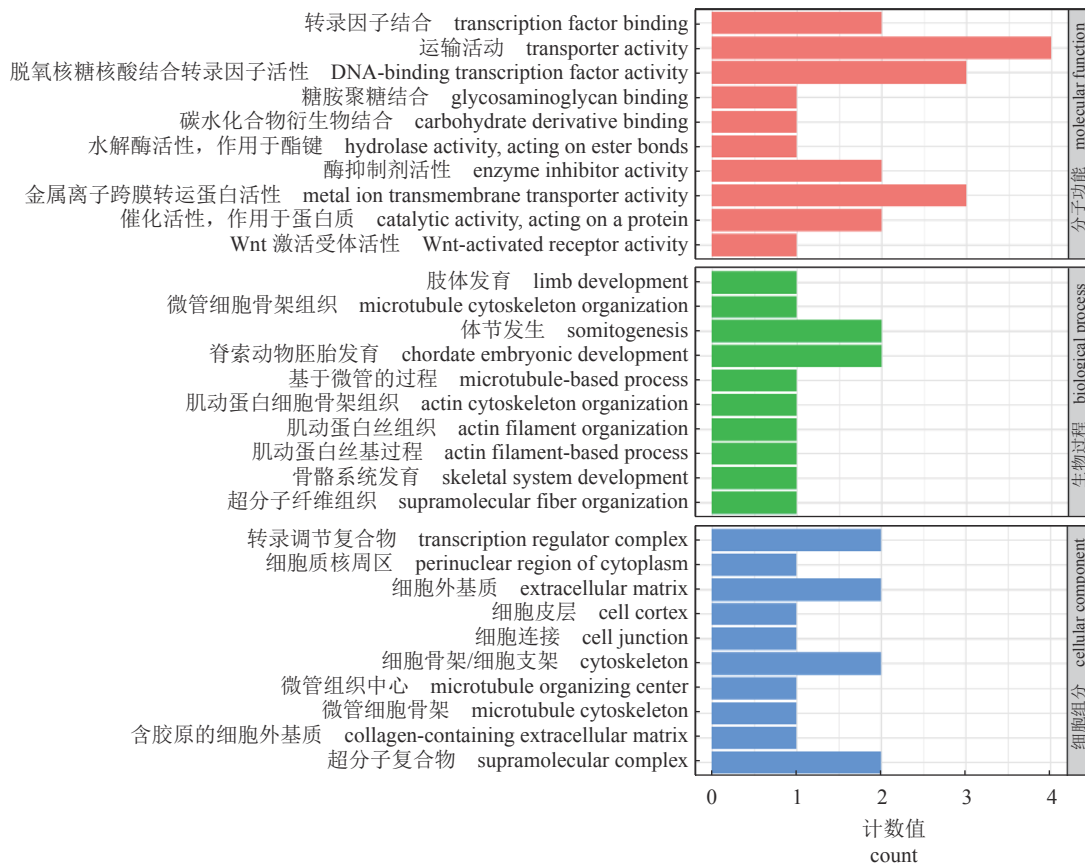
占差异表达基因的 99.12%, 在 Uniprot (Swiss-Prot) 蛋白数据库中得到功能注释的基因数为 186 个, 占差异表达基因的 81.94%。通过对差异表达基因功能注释结果的进一步分析, 发现了一些可能与肌肉生长相关的候选基因, 如生长/分化

因子9基因 (*gdf9*)、肌酸激酶M型基因 (*ckm*)、低密度脂蛋白受体相关蛋白4基因 (*lrp4*)、肌钙蛋白I/快速骨骼肌基因 (*ttni2*)、纤维蛋白原 $\gamma$ 链基因 (*fgg*)、胞质磷酸烯醇丙酮酸羧激酶 [GTP] 基因 (*pck1*)、乙醛酸还原酶 (*gyar*)、多不饱和脂肪酸脂氧合酶 ALOX8 (*alox8*)、载脂蛋白B-100基因 (*apob*) 及肌间线蛋白基因 (*des*) 等。

### 2.4 差异表达基因的 GO 和 KEGG 分析

为了进一步探究 227 个差异表达基因的生物学术功能, 本实验使用 clusterProfiler 软件通过 OrgDb 包对 227 个差异表达基因进行 GO 和 KEGG 通路富集分析。富集到的 GO term 总共有 679 个, 其中包括生物过程 (biological process, BP) 525 个, 细胞组分 (cellular component, CC) 63 个, 分子功能 (molecular function, MF) 91 个, 通过进一步分析, 发现一些差异表达基因富集到了一些可能与肌肉生长相关的 GO term 中, 如肌动蛋白丝组织 (actin filament organization, GO: 0007015)、肌动蛋白细胞骨架组织 (actin cytoskeleton organization, GO: 0030036)、超分子纤维组织 (supramolecular fiber

organization, GO: 0097435)、肌动蛋白丝基过程 (actin filament-based process, GO: 0030029)、微管细胞骨架组织 (microtubule cytoskeleton organization, GO: 0000226)、基于微管的过程 (microtubule-based process, GO: 0007017)、生长 (growth, GO: 0040007)、微管组织中心 (microtubule organizing center, GO: 0005815)、微管细胞骨架 (microtubule cytoskeleton, GO: 0015630)、体节发生 (somitogenesis, GO: 0001756)、肢体发育 (limb development, GO: 0060173)、脊索动物胚胎发育 (chordate embryonic development, GO: 0043009)、骨骼系统发育 (skeletal system development, GO: 0001501)、发育生长 (developmental growth, GO: 0048589)、组织形态发生 (tissue morphogenesis, GO: 0048729) 及胚胎形态发生 (embryonic morphogenesis, GO: 0048598) 等 (图 2-a), 这些 GO term 中富含了较多可能与肌肉生长相关的基因信息, 值得深入研究。富集到的 KEGG 通路共有 156 条, 通过进一步分析, 发现了一些可能与肌肉生长相关的通路, 如细胞和生长因子 (cytokines and growth factors, ko04052)、蛋白质消化吸收 (protein digestion and absorption,



(图 2 Fig. 2)

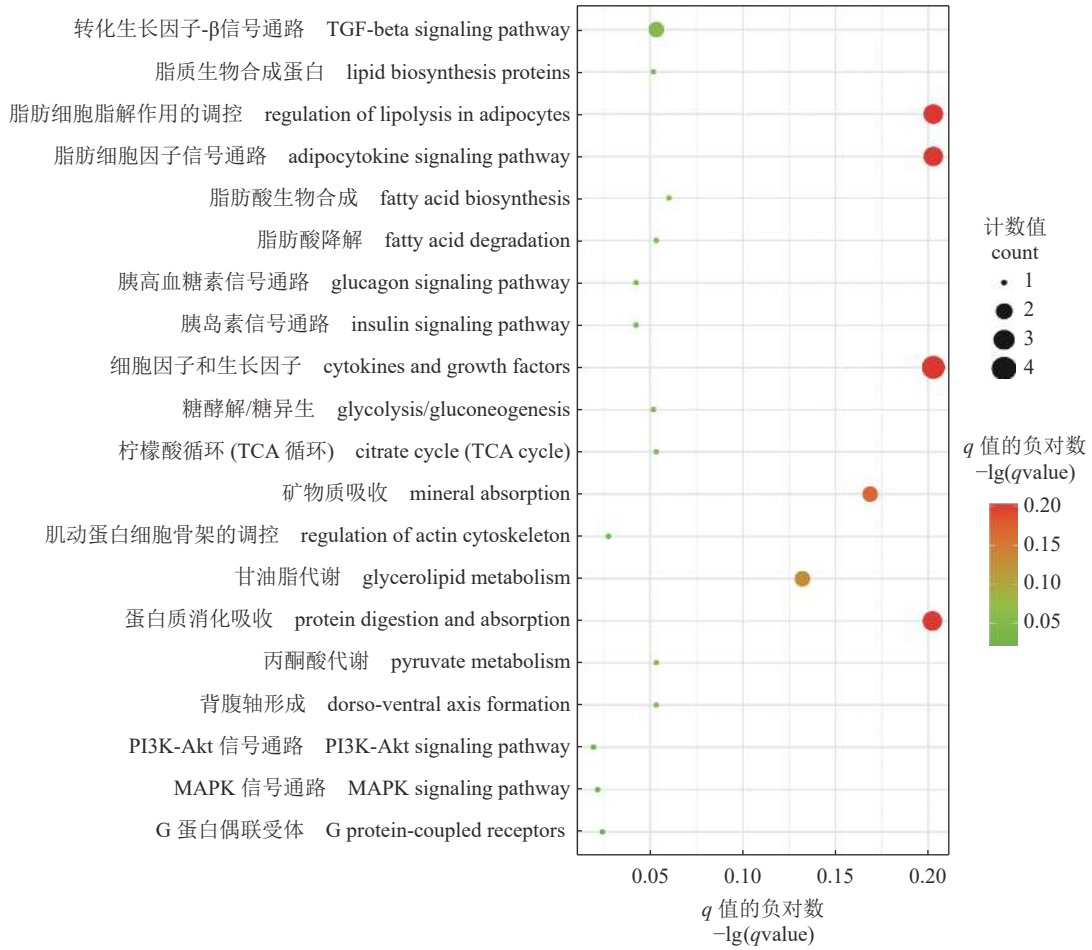


图 2 差异表达基因 GO 和 KEGG 通路富集分析

(a) GO 富集分析部分结果, x 轴代表基因数, y 轴代表 GO 条目; (b) KEGG 通路富集分析部分结果, x 轴代表富集因子, 颜色越深表示值越大、富集程度越高, 圆圈大小表示该通路中的基因数, y 轴代表通路名称。

Fig. 2 Enrichment analysis of GO and KEGG pathway of differentially expressed genes

(a) the partial results of GO enrichment analysis, the x-axis represents the number of genes, the y-axis represents GO term; (b) the partial results of KEGG pathway enrichment analysis, the x-axis represents the enrichment factor, the darker the color, the larger the value and the higher the enrichment degree, the size of the circle indicates the number of genes in the pathway, and the y-axis represents the name of the pathway.

ko04974)、矿物质吸收 (mineral absorption, ko04978)、脂肪酸生物合成 (fatty acid biosynthesis, ko00061)、转生长因子-β 信号通路 (TGF-β 信号通路) (TGF-beta signaling pathway, ko04350)、柠檬酸循环 (TCA 循环) (citrate cycle (TCA cycle), ko00020)、脂肪酸降解 (fatty acid degradation, ko00071)、丙酮酸代谢 (pyruvate metabolism, ko00620)、糖酵解/糖异生 (glycolysis/gluconeogenesis, ko00010)、脂质生物合成蛋白 (lipid biosynthesis proteins, ko01004)、胰高血糖素信号通路 (glucagon signaling pathway, ko04922)、胰岛素信号通路 (insulin signaling pathway, ko04910)、肌动蛋白细胞骨架的调控 (regulation of actin cytoskeleton,

ko04810)、G 蛋白偶联受体 (G protein-coupled receptors, ko04030)、MAPK 信号通路 (MAPK signaling pathway, ko04010) 及 PI3K-Akt 信号通路 (PI3K-Akt signaling pathway, ko04151) 等 (图 2-b), 这些通路中富含了较多可能与肌肉生长、脂肪酸和糖类物质代谢的基因信息, 值得深入研究。

2.5 WGCNA 结果

使用 WGCNA 软件进行无尺度加权基因共表达网络构建, 以标准化后的基因表达量矩阵作为输入, 共 412 个样本、19 477 个基因, 以中位绝对偏差前 75% 且至少 MAD 值大于 0.01 为标准筛选出了 14 607 个基因, 通过样本聚类的方法剔除

掉了 5 个离群样本, 使用 `powerEstimate` 函数估计的最适软阈值为 4, 此时无尺度网络拟合指数  $R^2$  为 0.9, 平均连通性趋于 0 (图 3), 该软阈值和无尺度网络拟合指数可以获得符合要求的无尺度网络。采用动态剪切法划分模块, 合并相似度大于 75% 的模块, 最终构建了 33 个共表达模块 (图 4), 不同颜色代表了不同模块, 各模块中基因数最多的为 `turquoise` 模块, 基因数为 2 408 个, 基因数最少的为 `darkolivegreen` 模块, 基因数为 33 个, 其他模块中的基因数为 40~2 145, `grey` 模块中的基因是没有划分到共表达模块中的基因, 有 2 145 个 (图 5)。通过各模块与体重性状之间的相关性分析发现, 与体重性状呈较强正相关性的共表达模块为 `midnightblue` 和 `purple` 模块, `midnightblue` 模块包含 163 个基因, `purple` 模块包含 313 个基因, 呈较强负相关性的共表达模块为 `darkgrey`、`green` 和 `magenta` 模块, `darkgrey` 模块包含 70 个基因, `green` 模块包含 640 个基因, `magenta` 模块包含 316 个基因, 正相关性最强的模块为 `midnight-`

`blue` 模块, 相关性系数为 0.47, 负相关性最强的模块为 `magenta` 模块, 相关性系数为 -0.27 (图 5, 图 6)。通过绘制 Top800 基因的模块间相关性热图, 发现模块间基因共表达相关性较高 (图 7)。对这些与体重性状相关性较强的共表达模块进行核心基因挖掘, 每个共表达模块中选择模块关系值 (module membership, MM) 前 3 的基因作为该模块的核心基因, 对这些核心基因通过 NCBI-nr 和 Uniprot (Swiss-Prot) 蛋白数据库进行功能注释, 注释结果表明, `midnightblue` 模块中的核心基因为 *L*-乳酸脱氢酶 A 链基因 (*ldha*)、糖原磷酸化酶/肌肉形式基因 (*pygm*) 和双磷酸甘油酸变位酶基因 (*bpgm*), `purple` 模块中的核心基因为肌凝素 1 基因 (*myoz1*)、原肌球蛋白  $\alpha$ -1 链基因 (*tpm1*) 和肌钙蛋白 I /快速骨骼肌基因 (*tnni2*) (图 8), `darkgrey` 模块中的核心基因为肌营养不良蛋白相关蛋白 2 基因 (*drp2*)、嗅觉调节素样 2A 基因 (*olfml2a*) 和转录因子 SOX-1 基因 (*sox1*), `green` 模块中的核心基因为非常规肌球蛋白-XVIIIb 基因 (*myo18b*)、Cul-

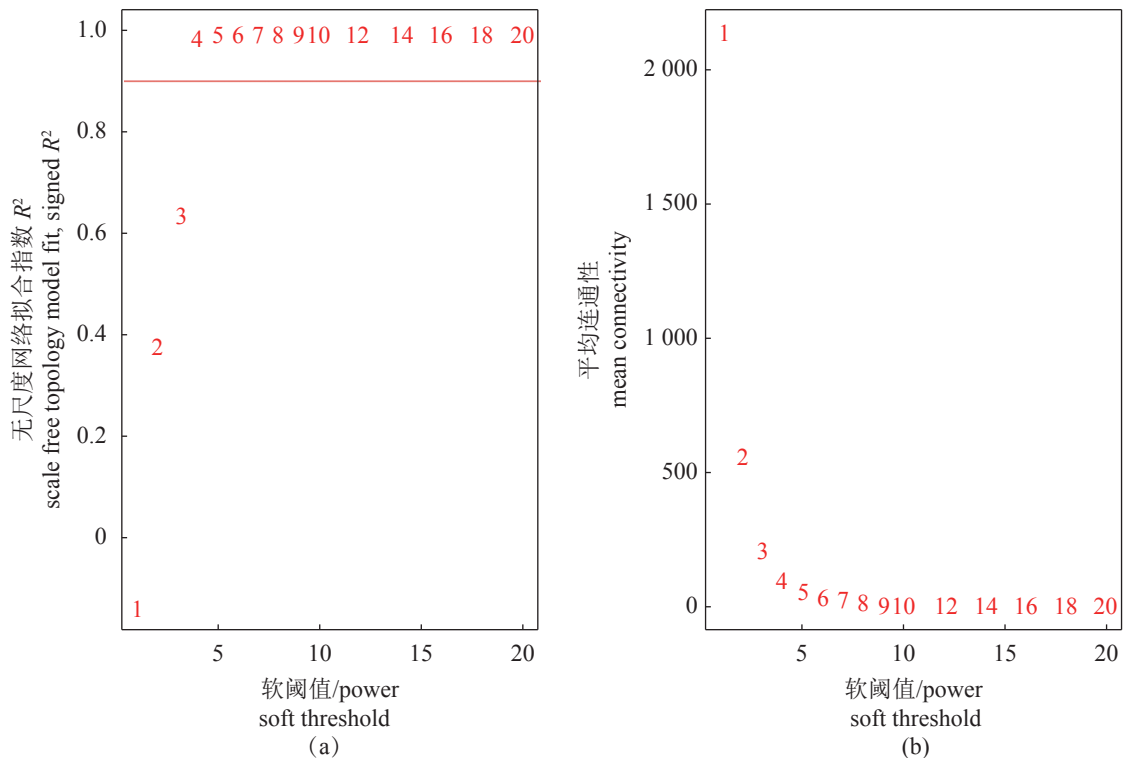


图 3 软阈值的筛选

(a) 不同软阈值对应的无尺度网络拟合指数  $R^2$ , 红色直线表示的  $R^2 = 0.9$ ; (b) 不同软阈值对应的平均连通性。

### Fig. 3 Screening of soft threshold

(a) the scale-free network fitting index  $R^2$  corresponding to different soft thresholds, the red straight line represents  $R^2 = 0.9$ ; (b) the average connectivity corresponding to different soft thresholds.



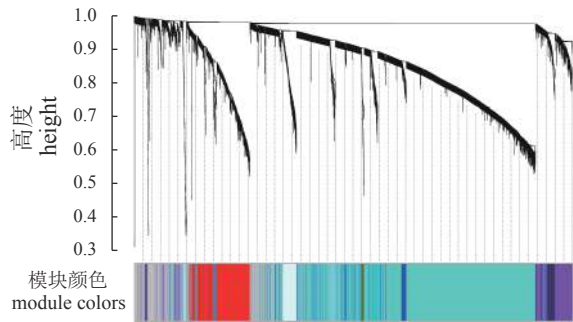


图 4 基因聚类树及模块划分

Fig. 4 Gene clustering tree and module division

lin 相关 NEDD8 解离蛋白 1 基因 (*cand1*) 和斯帕达辛基因 (*spg11*), magenta 模块中的核心基因为 26S 蛋白酶体非 ATPase 调节亚基 11B 基因 (*psmd11b*)、26S 蛋白酶体非 ATPase 调节亚基 6 基因 (*psmd6*) 和泛素样蛋白 7 基因 (*ubl7*)。

## 2.6 qRT-PCR 验证

为了验证转录组测序分析结果的可靠性, 随机挑选了 10 个差异表达基因 (上调表达 5 个, 下调表达 5 个) 进行 qRT-PCR 验证。上调表达的 5 个基因分别为 *kiaa754*、*slbp*、*gtf3a*、*ccnp*、*kiz*; 下调表达的 5 个基因分别为 *ttn*、*pnp*、*mmp13*、*npr1*、*epd*。经 qRT-PCR 实验验证, 所选差异表达基因的表达模式均与转录组测序分析结果一致 (图 9), 表明转录组测序分析结果的可信度较高。

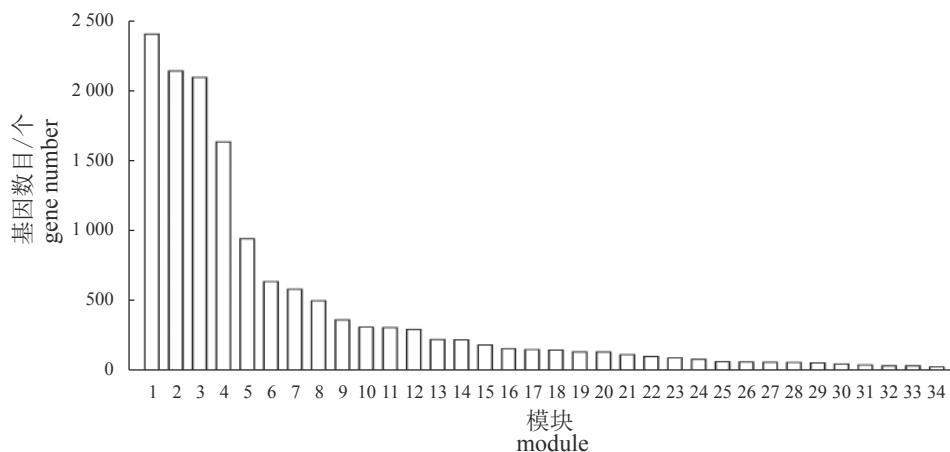


图 5 各模块中相应的基因数

1. 青绿色, 2. 灰色, 3. 蓝色, 4. 棕色, 5. 黄色, 6. 绿色, 7. 红色, 8. 黑色, 9. 粉红色, 10. 品红色, 11. 紫色, 12. 黄绿色, 13. 棕褐色, 14. 鲑肉色, 15. 蓝绿色, 16. 午夜蓝, 17. 浅青色, 18. 灰色, 19. 浅绿色, 20. 浅黄色, 21. 宝蓝色, 22. 深红色, 23. 深绿色, 24. 深青绿色, 25. 深灰色, 26. 橙色, 27. 深橙色, 28. 白色, 29. 天蓝色, 30. 马鞍棕色, 31. 钢蓝色, 32. 苍绿色, 33. 紫罗兰色, 34. 深橄榄绿色。

Fig. 5 The corresponding number of genes in each module

1. turquoise, 2. grey, 3. blue, 4. brown, 5. yellow, 6. green, 7. red, 8. black, 9. pink, 10. magenta, 11. purple, 12. greenish yellow, 13. tan, 14. salmon, 15. cyan, 16. midnightblue, 17. lightcyan, 18. grey, 19. lightgreen, 20. lightyellow, 21. royalblue, 22. darkred, 23. darkgreen, 24. darkturquoise, 25. darkgrey, 26. orange, 27. darkorange, 28. white, 29. skyblue, 30. saddlebrown, 31. steelblue, 32. pale turquoise, 33. violet, 34. darkolivegreen.

## 3 讨论

本实验使用生长较快大黄鱼 182 尾和生长较慢大黄鱼 230 尾进行比较转录组分析, 共获得差异表达基因 227 个, 其中上调表达的有 125 个, 下调表达的有 102 个, 这些差异表达基因在 NCBI-nr 非冗余蛋白数据库中的注释率为 99.12%, 在 Uniprot (Swiss-Prot) 蛋白数据库中的注释率为 81.94%, 两个数据库中注释率的差异可能是由数据库本身的丰富程度引起的。

在高通量测序技术的助力下, 加速了研究者对水产品包括生长性状在内的许多经济性状的遗传解析进程。目前大家公认的是生长性状是由多基因控制的极其复杂的一个数量性状, 尽管有大量的科研工作者对多种水产养殖品种生长性状的遗传分子机制进行了解析, 但也只是鉴定出了部分与生长性状相关的候选功能基因和调控通路, 目前并没有完全被解析。鱼类生长是一个复杂且受着精确调控的过程<sup>[40]</sup>。据报道, 鱼类骨骼肌质量的增加被认为对其身体发育有重要的贡献, 在硬骨鱼中, 其轴向肌肉由肌节组成, 鱼的游动是通过身体两侧这些连续排列的肌节的交替收缩来进行的, 鱼快肌纤维是中轴肌的重要组成部分, 这些纤维的发育受成肌祖细胞 (MPC) 的调控和维持, 骨骼肌发生, 也称为肌肉发生, 包括 MPC 来源的成肌细胞的增殖、迁移、融合和维持, 肌肉



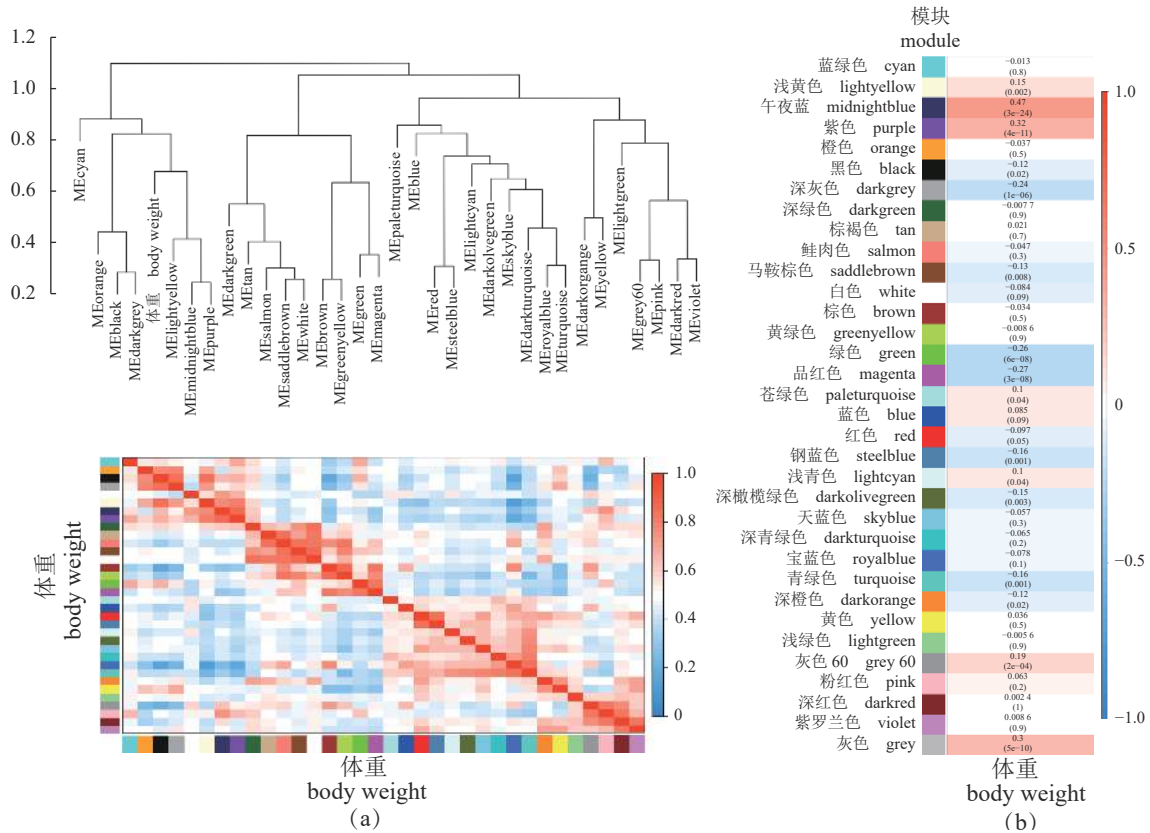


图 6 模块与体重性状相关性分析

(a) 上图为与体重性状相关的模块间树形聚类图，下图为与体重性状相关的模块间相关性热图；(b) 模块与体重性状间的相关性热图及相关性系数。

Fig. 6 Correlation analysis between module and body weight traits

(a) the picture on top shows the inter-module tree clustering map related to body weight traits, and the picture at bottom shows the heat map of inter-module correlation related to body weight traits; (b) the correlation heat map and correlation coefficient between modules and body weight traits.

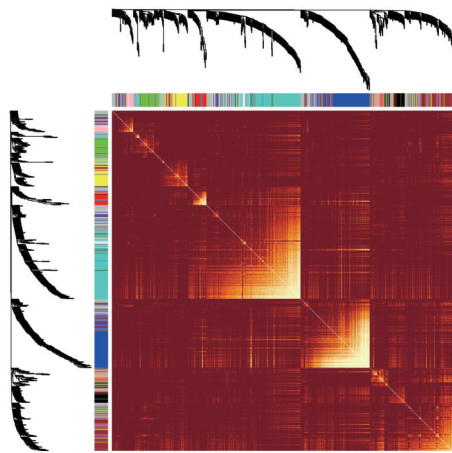


图 7 Top800 基因的模块间相关性热图

Fig. 7 Correlation heat map between modules of top 800 genes

发生是一个高度可塑性的过程<sup>[19]</sup>。本实验通过对筛选得到的差异表达基因进行功能注释，发现了一些可能与肌肉生长相关的候选功能基因，包括

*gdf9*、*ckm*、*tnni2* 和 *des* 等。有研究表明，*gdf9* 基因编码 TGF-β 超家族蛋白质的分泌配体，可促进原始卵泡发育和刺激颗粒细胞增殖，该基因还可以在体外激活人原始卵泡的发育<sup>[41]</sup>，有研究者在对鲫 (*Carassius auratus*) 体重相关的高密度遗传连锁图谱和 QTL 精细定位研究中，在定位到的 QTL 区间内发现了与鲫生长相关的潜在候选功能基因 *tgf-β*<sup>[42]</sup>，本实验中 *gdf9* 在生长较快个体组中的高表达很有可能是通过控制 TGF-β 分泌配体的表达来影响 *tgf-β* 的功能，进而影响肌肉生长；*ckm* 一般在能量需求波动较大组织 (如骨骼肌、心脏、大脑和精子等) 的能量转导中起着核心作用，本实验中该基因在生长较快个体组中呈现上调表达，推测这可能与生长较快的鱼骨骼肌中能量波动较大有关，最终导致了生长较快的鱼中该基因的高表达；*tnni2* 编码的是一种快速抽动的骨骼肌蛋白，这与本实验中该基因在生长较快个体组中呈现上调表达结果相一致，该基因是肌钙蛋白 I

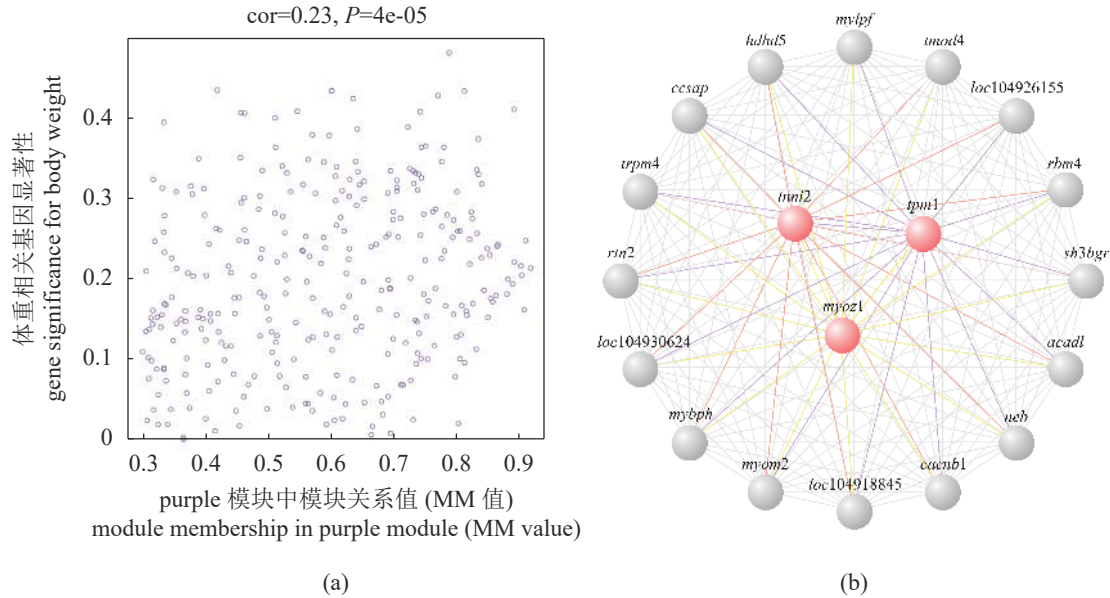


图 8 purple 模块核心基因挖掘

(a) purple 模块 MM 值与体重相关 GS 值间的相关性散点图; (b) purple 模块中核心基因的互作网络图: 图中央的红色圆圈表示核心基因。

Fig. 8 Core gene mining of purple module

(a) the scatter plot of correlation between purple module MM value and body weight related GS value; (b) the interaction network diagram of core genes in the purple module: the red circle in the center of the diagram indicates the core genes.

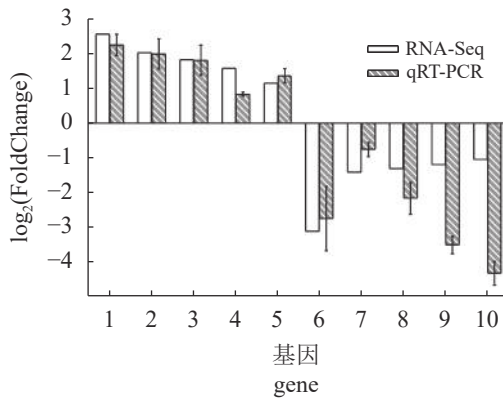


图 9 RNA-Seq 和 qRT-PCR 的基因表达模式

Fig. 9 Gene expression patterns of RNA-Seq and real-time fluorescence quantitative PCR

1. *kiaa754*, 2. *slbp*, 3. *gdf3a*, 4. *ccnp*, 5. *kiz*, 6. *tnn*, 7. *pnp*, 8. *mmp13*, 9. *npr1*, 10. *epd*.

基因家族的成员,也是肌钙蛋白复合物的组成部分,包括肌钙蛋白 T、肌钙蛋白 C 和肌钙蛋白 I 亚单位,肌钙蛋白复合物和原肌球蛋白一起负责横纹肌收缩的钙依赖性调节<sup>[43-44]</sup>; *des* 编码的是肌肉特异性 III 型中间丝,其对正常的肌肉结构和功能至关重要,在维持肌节结构、相互连接 Z 盘和形成肌原纤维方面起着至关重要的作用,它们不仅与肌膜细胞骨架相连,还与细胞核和线粒体相连,从而在活动过程中为肌纤维提供力量<sup>[45]</sup>,说

明该基因对肌肉的生长发育至关重要,本实验中该基因在生长较快个体组中呈现下调表达,具体机理还需深入研究。

本实验将筛选得到的 227 个差异表达基因进行 GO 和 KEGG 通路富集分析,结果显示这些差异表达基因富集到的 GO term 和 KEGG 通路大多与肌肉生长发育和脂肪酸与糖类物质代谢等相关,如 GO term 有肌动蛋白丝组织、肌动蛋白细胞骨架组织、肌动蛋白丝基过程、微管细胞骨架组织、微管组织中心、骨骼系统发育和发育生长等,KEGG 通路有细胞和生长因子、脂肪酸生物合成、转化生长因子-β 信号通路、柠檬酸循环、糖酵解/糖异生、胰高血糖素信号通路、胰岛素信号通路和肌动蛋白细胞骨架的调控等。如上所述有研究者在定位到的 QTL 区间内发现了与鲫生长相关的潜在候选功能基因 *tgf-β*,本实验在对差异表达基因进行 KEGG 分析时富集到了转化生长因子-β 信号通路,并且如上所述差异表达基因 *gdf9* 编码的是 TGF-β 超家族蛋白质的分泌配体,说明该通路和该基因很有可能与大黄鱼的肌肉生长相关,值得深入研究。已经被证实的调控鱼类生长的最主要候选基因和相关通路主要包括两类,分别为生长轴相关的和肌肉生长调控因子家族中的基因,其中生长轴相关的包含胰岛素样生长因子基因 (*igf-I* 和 *igf-II*),这些基因对调控鱼类生长代谢至关

重要, 同时在生长轴上还存在一些可以调控生长激素和胰岛素样生长因子合成与应用的其他激素, 这些激素中包括胰岛素, 而在本实验中不仅富集到了与肌肉生长直接相关的通路, 还富集到了胰高血糖素信号通路和胰岛素信号通路, 值得深入研究, 胰岛素样生长因子基因对鱼类新陈代谢过程和糖分的摄取, 以及在调节生长激素基因 (*gh*) 的表达和肌肉水平上的细胞生长有着重要的作用, 研究表明, 除了上述生长轴相关的和肌肉生长调控因子家族中的基因外, 还有其他基因也被证明可能通过调节能量代谢或肌肉生长相关的其他生理网络而显著影响生长, 因此与这些相关的一些基因和通路值得深入研究<sup>[23, 46]</sup>。

本实验使用生长较快与较慢大黄鱼的比较转录组表达量数据构建了无尺度加权基因共表达网络进行 WGCNA 分析, 选择的最适软阈值为 4, 无尺度网络拟合指数  $R^2$  为 0.9, 共获得了 33 个共表达模块。结果显示, 与体重性状正相关性较强的模块为 midnightblue 和 purple 模块, 负相关性较强的模块为 darkgrey、green 和 magenta 模块, 其中正相关性最强的模块为 midnightblue 模块, 相关性系数为 0.47, 负相关性最强的模块为 magenta 模块, 相关性系数为 -0.27, 对这些模块的核心基因进行挖掘, 每个共表达模块选取模块关系值 (MM 值) 前 3 的基因作为该模块的核心基因, 通过对这些基因进行功能注释, 发现 purple 模块中的核心基因可能与肌肉生长相关, 分别为 *myoz1*、*tpm1* 和 *tnni2*, 该模块与体重性状的相关性系数为 0.32。有研究表明, *myoz1* 对猪肌纤维生长发育有一定的作用, 钙调神经磷酸酶参与的很多信号通路可以影响肌纤维的生长发育, *myoz1* 可以通过抑制钙调神经磷酸酶的活性来影响猪肌纤维的生长发育, 促进猪快肌纤维向慢肌纤维转化<sup>[47]</sup>; *tpm1* 可以表达出对骨骼肌和平滑肌、大脑和细胞骨架特异的异构体, 研究表明 *tpm1* 的表达与猪背最长肌生长发育及品种间肌肉表型差异有关<sup>[48]</sup>。如上所述, *tnni2* 编码的是一种快速抽动的骨骼肌蛋白, 是肌钙蛋白复合物的组成部分, 肌钙蛋白复合物和原肌球蛋白一起负责横纹肌收缩的钙依赖性调节。总而言之, 通过对 WGCNA 分析中与体重性状相关性较强模块中核心基因的深度挖掘表明, purple 模块值得重点关注并深入研究。

本实验通过对同一个网箱内养殖的生长速率存在差异的 412 尾 2 龄大黄鱼个体的肌肉组织进行比较转录组分析, 共筛选得到了 227 个差异表

达基因, 基因注释率分别为 99.12% (NCBI-nr) 和 81.94% (Swiss-Prot), 通过功能注释结果和 WGCNA 分析初步筛选出可能与大黄鱼肌肉生长相关的候选功能基因, 如 *gdf9*、*ckm*、*tnni2*、*des* 和 *tpm1* 等。通过 GO 和 KEGG 通路富集分析发现了大量可能与大黄鱼肌肉生长相关的信息, 未来将对其进行深度挖掘, 结合本课题组开展的大黄鱼生长性状相关的全基因组关联分析对大黄鱼生长性状相关的功能基因进行精确定位, 并利用模式生物斑马鱼对鉴定到的功能基因进行转基因功能验证。

(作者声明本文无实际或潜在的利益冲突)

### 参考文献 (References):

- [1] 伍汉霖, 钟俊生. 中国石首鱼类的再研究 [C]//中国海洋湖沼学会鱼类学分会、中国动物学会鱼类学分会 2012 年学术研讨会论文摘要汇编. 兰州: 中国海洋湖沼学会, 2012.  
Wu H L, Zhong J S. The revision of the Sciaenoid fishes of China[C]//Abstracts of 2012 Academic Seminar of Ichthyology Branch of Chinese Society of Oceanography and Limnology and Ichthyology Branch of Chinese Zoological Society. Lanzhou: China Ocean and Limnology Society. 2012 (in Chinese).
- [2] 林新濯. 中国近海三种主要经济鱼类的生物学特性与资源现状[J]. 水产学报, 1987, 11(3): 187-194.  
Lin X Z. Biological characteristics and resources status of three main commercial fishes in offshore waters of China[J]. Journal of Fisheries of China, 1987, 11(3): 187-194 (in Chinese).
- [3] 李玉尚. 600年来中国带鱼种群历史专题导言[J]. 科学与管理, 2012, 32(1): 21-22.  
Li Y S. An introduction to the history of hairtail population in China in the past 600 years[J]. Science and Management, 2012, 32(1): 21-22 (in Chinese).
- [4] Zheng W B, Liu G Z, Ao J Q, et al. Expression analysis of immune-relevant genes in the spleen of large yellow croaker (*Pseudosciaena crocea*) stimulated with poly I: C[J]. Fish & Shellfish Immunology, 2006, 21(4): 414-430.
- [5] 农业农村部渔业渔政管理局, 全国水产技术推广总站, 中国水产学会. 2021 中国渔业统计年鉴 [M]. 北京: 中国农业出版社, 2021.  
Fisheries and Fisheries Administration, Ministry of Agriculture and Rural Affairs, National Fisheries Technology



- logy Extension Station, China Society of Fisheries. 2021 China fishery statistical yearbook[M]. Beijing: China Agriculture Press, 2021 (in Chinese).
- [6] Ye H, Wang X Q, Gao T X, *et al.* EST-derived microsatellites in *Pseudosciaena crocea* and their applicability to related species[J]. *Acta Oceanologica Sinica*, 2010, 29(6): 83-91.
- [7] Hong W S, Zhang Q Y. Review of captive bred species and fry production of marine fish in China[J]. *Aquaculture*, 2003, 227(1-4): 305-318.
- [8] Ni J, You F, Xu J H, *et al.* Single nucleotide polymorphisms in intron 1 and intron 2 of *Larimichthys crocea* growth hormone gene are correlated with growth traits[J]. *Chinese Journal of Oceanology and Limnology*, 2012, 30(2): 279-285.
- [9] 薛良义, 李婷, 孙升, 等. 大黄鱼肌肉生长抑制素基因外显子 I 遗传多态性分析[J]. *水产科学*, 2008, 27(10): 503-506.
- Xue L Y, Li T, Sun S, *et al.* Genetic polymorphism analysis of exon I in myostatin gene from large yellow croaker *Pseudosciaena crocea*[J]. *Fisheries Science*, 2008, 27(10): 503-506 (in Chinese).
- [10] 赵文, 高峰英, 石振广. 达氏鳢肌肉组织转录组测序和功能分析[J]. *水产学报*, 2014, 38(9): 1255-1262.
- Zhao W, Gao F Y, Shi Z G. The transcriptome sequencing and functional analysis of muscle tissue of Kaluga (*Huso dauricus*)[J]. *Journal of Fisheries of China*, 2014, 38(9): 1255-1262 (in Chinese).
- [11] 罗辉, 叶华, 肖世俊, 等. 转录组学技术在水产动物研究中的运用[J]. *水产学报*, 2015, 39(4): 598-607.
- Luo H, Ye H, Xiao S J, *et al.* Application of transcriptomics technology to aquatic animal research[J]. *Journal of Fisheries of China*, 2015, 39(4): 598-607 (in Chinese).
- [12] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics[J]. *Nature Reviews Genetics*, 2009, 10(1): 57-63.
- [13] 张毅, 吴万亿, 刘霞宇, 等. 利用WGCNA鉴定甘薯耐盐相关共表达网络及核心基因[J]. *河南农业科学*, 2021, 50(6): 16-27.
- Zhang Y, Wu W Y, Liu X Y, *et al.* Identification of salt tolerance co-expression modules and hub genes in *Ipomoea batatas* by WGCNA[J]. *Journal of Henan Agricultural Sciences*, 2021, 50(6): 16-27 (in Chinese).
- [14] 汪颖, 高芳, 刘兆新, 等. 利用WGCNA鉴定花生主茎生长基因共表达模块[J]. *作物学报*, 2021, 47(9): 1639-1653.
- Wang Y, Gao F, Liu Z X, *et al.* Identification of gene co-expression modules of peanut main stem growth by WGCNA[J]. *Acta Agronomica Sinica*, 2021, 47(9): 1639-1653 (in Chinese).
- [15] Li E C, Li C. Use of RNA-seq in aquaculture research[J]. *Poultry, Fisheries & Wildlife Sciences*, 2014, 2(2): e108.
- [16] 唐艺. 基于 dual RNA-seq 的大黄鱼与变形假单胞菌互作的研究 [D]. 厦门: 集美大学, 2020.
- Tang Y. Studies on the interaction between *Larimichthys crocea* and *Pseudomonas plecoglossicida* based on dual RNA-seq[D]. Xiamen: JiMei University, 2020 (in Chinese).
- [17] 张恒泽, 高丽婷, 万里, 等. 假单胞菌减毒活疫苗免疫后的大黄鱼脾脏转录组分析[J]. *基因组学与应用生物学*, 2020, 39(1): 123-131.
- Zhang H Z, Gao L T, Wan L, *et al.* Transcriptome analysis of spleen of the large yellow croaker (*Larimichthys crocea*) vaccinated with live attenuated vaccine of *Pseudomonas plecoglossicida*[J]. *Genomics and Applied Biology*, 2020, 39(1): 123-131 (in Chinese).
- [18] 曹款. 大黄鱼性腺性别分化的转录组学研究 [D]. 厦门: 集美大学, 2016.
- Cao K. Characterization of gonadal transcriptomes from large yellow croaker *Larimichthys crocea*[D]. Xiamen: JiMei University, 2016 (in Chinese).
- [19] Eldem V, Zararsız G, Erkan M, *et al.* De novo assembly and comprehensive characterization of the skeletal muscle transcriptomes of the European anchovy (*Engraulis encrasicolus*)[J]. *Marine Genomics*, 2015, 20: 7-9.
- [20] 邵芳, 卢祥云, 顾志良. 鱼类MSTN基因的研究进展[J]. *常熟理工学院学报(自然科学版)*, 2011, 25(8): 76-80.
- Shao F, Lu X Y, Gu Z L. Research progress of fish Myostatin gene[J]. *Journal of Changshu Institute of Technology (Natural Science Edition)*, 2011, 25(8): 76-80 (in Chinese).
- [21] 苗曼宁, 郭益文, 胡德宝, 等. MSTN基因编辑牛肌肉转录组测序及生物信息学分析[J]. *中国畜牧兽医*, 2021, 48(7): 2271-2281.
- Miao M N, Guo Y W, Hu D B, *et al.* Transcriptome sequencing and bioinformatics analysis of MSTN gene-edited bovine muscles[J]. *China Animal Husbandry & Veterinary Medicine*, 2021, 48(7): 2271-2281 (in Chinese).
- [22] Banerjee-Basu S, Baxevanis A D. Molecular evolution of the homeodomain family of transcription factors[J].



- [Nucleic Acids Research](#), 2001, 29(15): 3258-3269.
- [23] Xu Y X, Zhu Z Y, Wang C M, *et al.* Characterization of two parvalbumin genes and their association with growth traits in Asian seabass (*Lates calcarifer*)[C]//International Journal of Modern Physics: Conference Series. The World Bank, 2006.
- [24] 王兰梅, 朱文彬, 傅建军, 等. 福瑞鲤2号不同生长速率个体肌肉组织转录组分析[J]. 水产学报, 2021, 45(1): 79-87.
- Wang L M, Zhu W B, Fu J J, *et al.* De novo transcriptome analysis and comparison of the FFRC No.2 strain common carp (*Cyprinus carpio*) associated with its muscle growth[J]. Journal of Fisheries of China, 2021, 45(1): 79-87 (in Chinese).
- [25] Chen S F, Zhou Y Q, Chen Y R, *et al.* Fastp: an ultrafast all-in-one FASTQ preprocessor[J]. [Bioinformatics](#), 2018, 34(17): i884-i890.
- [26] Dobin A, Davis C A, Schlesinger F, *et al.* STAR: ultrafast universal RNA-seq aligner[J]. [Bioinformatics](#), 2013, 29(1): 15-21.
- [27] Kovaka S, Zimin A V, Pertea G M, *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2[J]. [Genome Biology](#), 2019, 20(1): 278.
- [28] Pertea M, Kim D, Pertea G M, *et al.* Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown[J]. [Nature Protocols](#), 2016, 11(9): 1650-1667.
- [29] Pertea M, Pertea G M, Antonescu C M, *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads[J]. [Nature Biotechnology](#), 2015, 33(3): 290-295.
- [30] Liao Y, Smyth G K, Shi W. Featurecounts: an efficient general purpose program for assigning sequence reads to genomic features[J]. [Bioinformatics](#), 2014, 30(7): 923-930.
- [31] Love M I, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2[J]. [Genome Biology](#), 2014, 15(12): 550.
- [32] Buchfink B, Xie C, Huson D H. Fast and sensitive protein alignment using DIAMOND[J]. [Nature Methods](#), 2015, 12(1): 59-60.
- [33] Huerta-Cepas J, Forslund K, Coelho L P, *et al.* Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper[J]. [Molecular Biology and Evolution](#), 2017, 34(8): 2115-2122.
- [34] Huerta-Cepas J, Szklarczyk D, Forslund K, *et al.* eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences[J]. [Nucleic Acids Research](#), 2016, 44(D1): D286-D293.
- [35] Eddy S R. Accelerated profile HMM searches[J]. [PLoS Computational Biology](#), 2011, 7(10): e1002195.
- [36] Yu G, Wang L G, Han Y, *et al.* clusterProfiler: an R package for comparing biological themes among gene clusters[J]. [Omics-a Journal of Integrative Biology](#), 2012, 16(5): 284-287.
- [37] Langfelder P, Horvath S. Fast R functions for robust correlations and hierarchical clustering[J]. [Journal of Statistical Software](#), 2012, 46(11): 1-17.
- [38] Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis[J]. [Statistical Applications in Genetics and Molecular Biology](#), 2005, 4(1): 17.
- [39] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis[J]. [BMC Bioinformatics](#), 2008, 9(1): 559.
- [40] Zheng Q, Zhang Y, Chen Y, *et al.* Systematic identification of genes involved in divergent skeletal muscle growth rates of broiler and layer chickens[J]. [BMC Genomics](#), 2009, 10(1): 87.
- [41] Kedem A, Fisch B, Garor R, *et al.* Growth differentiating factor 9 (*GDF9*) and bone morphogenetic protein 15 both activate development of human primordial follicles *in vitro*, with seemingly more beneficial effects of *GDF9*[J]. [The Journal of Clinical Endocrinology & Metabolism](#), 2011, 96(8): e1246-e1254.
- [42] Liu H Y, Fu B D, Pang M X, *et al.* A high-density genetic linkage map and QTL fine mapping for body weight in crucian carp (*Carassius auratus*) using 2b-RAD sequencing[J]. [G3 Genes| Genomes| Genetics](#), 2017, 7(8): 2473-2487.
- [43] Mullen A J, Barton P J R. Structural characterization of the human fast skeletal muscle troponin I gene (*TNNI2*)[J]. [Gene](#), 2000, 242(1-2): 313-320.
- [44] Sheng J J, Jin J P. *TNNI1*, *TNNI2* and *TNNI3*: evolution, regulation, and protein structure-function relationships [J]. [Gene](#), 2016, 576(1): 385-394.
- [45] Hnia K, Ramspacher C, Vermot J, *et al.* Desmin in muscle and associated diseases: beyond the structural function[J]. [Cell and Tissue Research](#), 2015, 360(3): 591-608.

- [46] De-Santis C, Jerry D R. Candidate growth genes in fin-fish-where should we be looking?[J]. *Aquaculture*, 2007, 272(1-4): 22-38.
- [47] 赵香丽. *MYOZ1* 基因在肌纤维发育中功能的初步研究 [D]. 武汉: 华中农业大学, 2011.
- Zhao X L. The Basic research of the role of *MYOZ1* during the porcine muscle fiber development[D]. Wuhan: Huazhong Agricultural University, 2011 (in Chinese).
- [48] 梁如意, 李岩, 唐中林, 等. *TPM1* 基因在猪背最长肌生长发育过程中的动态表达[J]. *中国农业科学*, 2012, 45(11): 2267-2272.
- Liang R Y, Li Y, Tang Z L, *et al.* Dynamic expression of *TPM1* gene during *Longissimus dorsi* muscle development in pigs[J]. *Scientia Agricultura Sinica*, 2012, 45(11): 2267-2272 (in Chinese).

## Comparative analysis of transcriptome of muscle tissue of individuals with different growth rate of *Larimichthys crocea*

ZHANG Bo, JIANG Dan, ZHANG Dongling, WANG Zhiyong, FANG Ming\*

(Key Laboratory of Healthy Mariculture in the East China Sea, Ministry of Agriculture and Rural Affairs, Fisheries College of JiMei University, Xiamen 361021, China)

**Abstract:** In order to explore the molecular regulation mechanism of growth traits of *Larimichthys crocea*, in this experiment, 182 individuals with fast growth rate [average body weight ( $527.1 \pm 83.6$ ) g] and 230 individuals with slow growth rate [average body weight ( $238.4 \pm 52.3$ ) g] were selected from *L. crocea* cultured in the same cage, a total of 412 individuals were studied. The muscle tissues were subject to total RNA extraction and sequenced by transcriptome, and the differential expression of genes between the two groups was analyzed. According to  $P_{adj} < 0.05$  and  $abs[\log_2(\text{FoldChange})] > 1$ , 227 differentially expressed genes were screened, including 125 up-regulated genes and 102 down-regulated genes. The annotation rates of differentially expressed genes in NCBI-nr and UniProt (Swiss-Prot) databases were 99.12% and 81.94%, respectively. Based on the functional annotation results of differentially expressed genes, candidate functional genes that may be related to muscle growth, such as *gdf9*, *ckm*, *tnni2* and *des*, were screened. GO and KEGG analysis predicted some information related to muscle growth, such as actin filament organization, actin cytoskeleton organization, actin filament-based process, microtubule organizing center and developmental growth in GO term, and cytokines and growth factors, fatty acid biosynthesis, transforming growth factor-beta signaling pathway (TGF- $\beta$  signaling pathway), insulin signaling pathway and regulation of actin cytoskeleton in KEGG pathway. WGCNA analysis showed that there was a strong correlation between purple module and body weight traits, and its core genes *myoz1*, *tpm1* and *tnni2* may be related to muscle growth. These results provide a reference for further verification of the functions of related genes and the analysis of the molecular regulation mechanism of growth traits of *L. crocea*.

**Key words:** *Larimichthys crocea*; muscle growth; transcriptome; differentially expressed genes; weighted gene co-expression network analysis (WGCNA)

**Corresponding author:** FANG Ming. E-mail: fangming618@126.com

**Funding projects:** National Natural Science Foundation of China (31672399, U1705231, 31872560, 32172964); Natural Science Foundation of Fujian Province (2020J01672, 2021J02045)