

斑马鱼肌肉高表达基因近端非编码元件分析及功能检测

鲁程成 赵一凡 范纯新 王建

IDENTIFICATION OF PROXIMAL CIS-REGULATORY ELEMENT FOR MUSCLE HIGHLY EXPRESSED GENES IN ZEBRAFISH

LU Cheng-Cheng, ZHAO Yi-Fan, FAN Chun-Xin, WANG Jian

在线阅读 View online: <https://doi.org/10.7541/2022.2020.290>

您可能感兴趣的其他文章

Articles you may be interested in

斑马鱼 *IRF11* 基因的鉴定、亚细胞定位及表达特征

IDENTIFICATION, SUBCELLULAR LOCALIZATION AND EXPRESSION CHARACTERIZATION OF ZEBRAFISH *IRF11*

水生生物学报. 2017, 41(1): 1–8 <https://doi.org/10.7541/2017.1>

斑马鱼 *ftr56* 基因克隆表达及功能研究

MOLECULAR CLONING, EUKARYOTIC EXPRESSION AND FUNCTION STUDY OF *FTR56* FROM *DANIO RERIO*

水生生物学报. 2020, 44(1): 20–25 <https://doi.org/10.7541/2020.003>

fbxo32 基因在 *traf6* 缺失斑马鱼肝脏中的表达分析

THE EXPRESSION ANALYSIS OF *FBXO32* GENE IN LIVER OF *TRAF6* MUTANT ZEBRAFISH

水生生物学报. 2021, 45(5): 945–950 <https://doi.org/10.7541/2021.2020.043>

斑马鱼和鲤 *miR-1-2/133a-1* 基因间序列活性和调控研究

STUDY ON THE INTERGENE SEQUENCE ACTIVITY AND REGULATION OF *MIR-1-2/133A-1* IN ZEBRAFISH AND CARP

水生生物学报. 2021, 45(3): 507–513 <https://doi.org/10.7541/2021.2020.040>

敲除 *trim47* 基因斑马鱼脑和脾的比较转录组学分析

COMPARATIVE TRANSCRIPTOMIC ANALYSIS OF BRAIN AND SPLEEN IN *TRIM47* KNOCKOUT ZEBRAFISH (*DANIO RERIO*)

水生生物学报. 2021, 45(3): 495–506 <https://doi.org/10.7541/2021.2020.036>

两种抗雌激素药物对雌性斑马鱼脂肪代谢的影响

EFFECTS OF TWO ANTI-ESTROGEN DRUGS EXPOSURE ON LIPID METABOLISM OF FEMALE ZEBRAFISH (*DANIO RERIO*)

水生生物学报. 2017, 41(1): 95–100 <https://doi.org/10.7541/2017.12>



关注微信公众号，获得更多资讯信息

doi: 10.7541/2022.2020.290

斑马鱼肌肉高表达基因近端非编码元件分析及功能检测

鲁程成^{1,2} 赵一凡^{1,2} 范纯新^{1,2} 王建^{1,3}

(1. 上海海洋大学国家海洋生物科学国际联合研究中心, 上海 201306; 2. 上海海洋大学国际海洋研究中心海洋生物系统与神经科学研究所, 上海 201306; 3. 上海海洋大学水产种质资源发掘与利用教育部重点实验室, 上海 201306)

摘要: 为鉴定鱼类肌肉组织特异性顺式调控元件, 通过分析斑马鱼多个组织的转录组数据, 筛选出肌肉高表达基因及低表达基因。通过MEME对肌肉高表达基因和低表达基因非编码区序列特征进行分析, 在5个肌肉高表达基因的转录起始位点上游发现了序列保守的DNA区域, 包含6个排列顺序一致的DNA基序。将其中一段目标片段插入具有Tol2转座子元件的基础启动子驱动的eGFP编码基因的上游, 构建表达载体。注射载体至斑马鱼受精卵获得转基因胚胎, 分析胚胎中eGFP荧光的表达模式, 发现该DNA片段具有增强基因在肌肉特异性表达的功能。Tomtom预测该DNA区域可能作为Myod等多个转录因子的结合位点。研究结果有助于理解鱼类肌肉基因表达的遗传基础, 并为利用生物信息学方法预测组织特异性转录调控元件提供新思路。

关键词: 顺式调控元件; 肌肉; 基因表达; 斑马鱼

中图分类号: Q344⁺.1

文献标识码: A

文章编号: 1000-3207(2022)03-0292-11



肌肉是鱼类最主要的结构和功能组织之一, 且经济鱼类的肌肉相关性性状很大程度影响着其品质与产量。因此鱼类肌肉形成和生长性状相关因素一直是水生生物学和水产养殖领域的热点问题。肌肉重要组成蛋白如肌球蛋白重链Myh (Myosin heavy chain)家族、Melusin等编码基因的表达及其调控, 在肌肉组织形成及功能维持中起到关键作用^[1,2]。

基因组非编码区存在大量顺式调控元件(Cis-regulatory element), 是调控基因表达的重要遗传因素, 如增强子(Enhancer)可以增强相关基因的转录水平。增强子常带有细胞及组织特异性^[3,4], 同一基因也可受不同增强子调控在不同细胞中表达^[3]。调控元件与生物体许多重要表型相关^[4-6], 如调控元件的变异可导致动物外耳及骨骼发育异常^[7]。肌肉组织特异性调控元件可以作为特定转录因子的结合位点在肌肉形成中发挥关键作用^[8]。如在肌肉发育过程中, 许多肌肉特异性基因的调控区可以与Myod (Myogenic differentiation 1)、Myf5 (Myogenic factor 5)等生肌调节因子(Myogenic regulatory factors, MRFs)家族蛋白结合而调控下游基因

表达^[9]。这种调控机制也可诱导成熟细胞类型变化, 如Myod可诱导成纤维细胞等多种细胞向成肌细胞转化, 且转化后的细胞中大量表达Myh等肌肉重要组成蛋白^[10]。可见, 研究肌肉特异性调控元件有助于解释肌肉形成的遗传基础。

表达模式相同的基因可能受同类转录因子及其对应顺式元件的转录调控, 因此同一组织表达的多个基因的顺式调控元件之间可能存在相同或相似的序列特征。许多肌肉表达基因的调控区存在E-box (CANNTG)DNA基序(Motifs), 与MRFs家族蛋白的碱性螺旋-环-螺旋(Basic helix-loop-helix, bHLH)结构域具有很高的亲和力^[11]。另外, 小鼠、线虫等物种的几种组织特异表达基因的调控区也发现了相同的DNA基序^[12,13]。此外, 部分顺式调控元件的DNA序列具有物种间保守性, 有时甚至高于编码区序列^[14,15]。可见, 分析肌肉表达基因调控区序列的保守性特征可以预测调控元件的存在。

本文通过分析两组斑马鱼不同组织的转录组测序数据, 筛选出在肌肉中高表达的基因, 并根据序列保守特征预测这些基因上游的调控元件。结

收稿日期: 2021-01-04; 修订日期: 2021-05-26

基金项目: 国家自然科学基金(31702329和31772406)资助 [Supported by the National Natural Science Foundation of China (31702329 and 31772406)]

作者简介: 鲁程成(1995—), 男, 硕士; 主要研究方向为生物信息学。E-mail: m180100067@st.shou.edu.cn

通信作者: 王建(1982—), 女, 博士, 硕士生导师; 主要研究方向为分子遗传学。E-mail: j_wang@shou.edu.cn

果在5个肌肉高表达基因的近端获得了序列特征相似DNA区域。进一步利用斑马鱼体内荧光报告基因表达检测系统对其中一段序列进行检测,发现其具有增强报告基因在肌肉组织中表达的能力,其中可能包含一类肌肉组织特异型增强子。

1 材料与方法

1.1 实验动物与胚胎

本实验所用野生型AB斑马鱼与TU斑马鱼均购自国家斑马鱼中心。TB斑马鱼由野生型AB与TU品系杂交获得。所有实验用鱼均饲养于14h:10h光暗周期环境中,养殖水温为26—28℃,每天早晚各喂食一次卤虫。

本实验所用胚胎由2对5月龄TB成鱼内交获得。胚胎饲养于Blue water(0.06 g红海盐、0.01 mg/L亚甲基蓝)中,于28.5℃进行培养。

1.2 斑马鱼肌肉高表达基因筛选

通过NCBI的SRA数据库下载得到两组包含成体野生斑马鱼肌肉及其他组织的RNA-seq数据(PRJNA255848^[16]和PRJNA263496^[17])。使用fastq-dump转化得到fastq文件,并使用Trimmomatic^[18]进行质控,去除序列5'端14 bp,其他保持默认参数。通过Ensembl数据库(release-94)下载获得斑马鱼基因组参考序列文件(Danio_rerio.GRCz11.dna_sm.primary_assembly.fa.gz)及注释文件(Danio_rerio.GRCz11.94.gtf)。使用HISAT2^[19]将质控后的fastq文件对比到基因组,并使用StringTie^[20]进行基因表达量计数。使用edgeR^[21]进行组织间基因表达差异分析,筛选肌肉相对其他6个组织(脑、鳃、心、肝、肾和肠)高表达($\log_2FC \geq 2$)的基因并取交集。两组RNA-seq分析结果取交集得到肌肉高表达基因。使用相同方法,以 $\log_2FC \leq -2$ 为标准筛选肌肉低表达基因。通过<http://geneontology.org/>^[22]对候选基因进行GO功能富集,富集结果使用ggplot2^[23]进行绘图展示。

1.3 肌肉高表达基因近端保守DNA元件预测

根据基因组注释文件,获得肌肉高表达及低表达基因的近端调控区(基因起始位点上游5000 bp及下游1000 bp)在基因组中的位置信息。使用bedtools getfasta工具从基因组参考序列文件中提取目标区域的DNA序列。根据ANCORA^[24](<http://ancora.genereg.net/>)数据库,以90pc_50col为标准,在上述斑马鱼肌肉高表达基因的近端调控区查找跨物种保守元件(HCNE, Highly conserved noncoding elements),进一步获得物种间调控区保守的肌肉高表达基因。

使用MEME^[25]工具(<http://meme-suite.org/tools/meme>)中的Discriminative mode算法,以种间保守肌肉高表达基因的近端调控区DNA为基本序列(Primary sequences),以肌肉低表达基因对应区域DNA作为对照序列(Control sequences),预测肌肉高表达基因间的近端保守DNA基序。使用MAFFT(<https://www.ebi.ac.uk/Tools/msa/mafft/>)进行DNA多序列比对及序列相似度计算。使用Tomtom^[26]工具(<http://meme-suite.org/tools/tomtom>),以JASPAR CORE (2018) vertebrates^[27]数据库为参考,预测DNA基序中的转录因子结合位点。

1.4 荧光定量PCR(qPCR)验证目标基因表达情况

取TB斑马鱼成鱼雌雄各5尾,解剖获得肌肉、脑、心、肝、肾和肠6种组织,参照TRIzol(ThermoFisher, 15596026)说明书的步骤提取各组织的总RNA。参照HiScript III 1st Strand cDNA Synthesis Kit(诺唯赞, R312-01)试剂盒步骤将RNA逆转录为cDNA。以elfa为内参对目标基因进行qPCR相对定量检测,每个基因每个组织样品进行3次重复,引物信息见表1。qPCR反应按照 ChamQ Universal SYBR qPCR Master Mix(诺唯赞, Q711-02)说明进行,反应体系为20 μ L,反应条件为: 95℃, 10s; 60℃, 30s; 40个循环。结果使用R以 $2^{-\Delta\Delta C_t}$ 进行绘图展示和统计分析,用方差分析加Tukey HSD多重比较检验基因在各组织间的表达差异。

1.5 荧光蛋白表达报告质粒构建

带有Tol2和EGFP的质粒由上海海洋大学刘志伟实验室惠赠,使用无缝克隆技术参照pTol2-GT2MP-EGFP质粒^[28]的结构进行改造,并在Tol2内部插入人源 β -globin绝缘子序列(HBB-5'HS5和HBB-3'HS1)^[29],在Gata2启动子上游插入候选调控元件。具体步骤:使用1%的间氨基苯甲酸乙酯甲磺酸盐(Ethyl 3-aminobenzoate methanesulfonate, MS222, sigma, E10521-10G)麻醉TU斑马鱼后取尾鳍,用DNeasy Blood & Tissue Kit (QIAGEN, 69504)提取斑马鱼基因组DNA。使用高保真酶Phanta(诺唯赞, P515-01), Gata2-F/R引物(表1)以基因组DNA为模板扩增得到Gata2启动子, Vector-gata-F/R(表1)引物扩增质粒骨架,并使两者间存在15—20 bp的重叠区。用Dpn I (NEB, #R0176V)消化去除PCR产物中的质粒模板,然后利用单片段快速克隆试剂盒(ClonExpress II One Step Cloning Kit, 诺唯赞)连接以上2个片段。连接产物转化至大肠杆菌, 37℃培养12h,使用高纯度质粒小提中量试剂盒(天根, DP107-02)提取质粒pTol2-GT2MP-EGFP。以此质粒为模板,进一步利用含有人源 β -

表 1 引物信息表

Tab. 1 PCR Primers

引物名称 Primer name	引物序列 Primer sequence (5'—3')	用途 Usage
elfa-qPCR -F	TCAGGACGCTGTAGATTCGC	荧光定量 PCR
elfa-qPCR -R	CCGCTAGCATTACCCTCC	
itgbl1-qPCR-F	TGGATCCCTCAGGAGACTGG	
itgbl1-qPCR-R	GCCGTGCACACGTTTATCTC	
obs11b-qPCR-F	AGTGTACTGTAGAAGAAGCTCC A	质粒构建
obs11b-qPCR-RT	GTGTGCGTGAATCCTGCTT	
zgc:92429-qPCR-F	GACGCCTTAAAGGGCTGGTC	
zgc:92429-qPCR-R	ATGATCTCCTCTGCTGTGCC	
myh6-qPCR-F	CAGCCTGGATGATCTACACCTAC	
myh6-qPCR-R	RGCTGCTGCTCTTCTTGCTACT	
mylk4a-qPCR-F	GGCCTCGCCAGAAAGTATCA	
mylk4a-qPCR-R	GTCTCGTTATCGTCGTCGCC	
Gata2-F	CTGCGCTGAATGATGAGTCTCT	
Gata2-R	CTCAAGTGTCGCGCTTAGAA	
dr334-F	CATTCGTTTCCCTTCGGCTTATT	
dr334-R	GCACTGTCACCTTACAACAAGAA	
Vector-gata-F	CTAAGCGCGGACACTTGAGCCGC CACCATGTCTAGAGTGAGCAAGG GCGA	
Vector-gata-R	CTCATCATTCAGCGCAGGGGTCA GGGCCCAAGTGATC	
Insulator-F	TTCTTCCCTCTAGTGGTCATAACA GCAGCTTCAGCTACCTCTCGCACT GTCACCTTACAACAAG	
Insulator-R	ACTAGGGGTCAGAAGTAGTTCAT CAAACCTTCTTCCCTCCCTAAACA ACAACAATTGCATTCAATTT	
Vector-F	CACTAGAGGGAAGAAGATACCTG TGGAGCTAATGGTCCAAGATGGG GTCAGGGCCCAAGTG	
Vector-R	ACTTCTGACCCCTAGTGGTGTCCA GAAAAGACCATTAAAGGAATGG GATCATCATCGATAGAGAAATGT CGAAGGGAACGAATGCTGCGCT GAATGATGAGTCTC	
vector-HS-F	TTGTAAAGGTGACAGTGCGAGAG GTAGCTGAAGCTGC	
vector-HS-R		

globin绝缘子序列的引物Insulator-F/R和Vector-HS-F/R (表 1), 分别扩增GT2MP-EGFP和Tol2质粒骨架两DNA片段, 两片段利用单片段快速克隆试剂盒连接, 产物转化后提质粒, 经测序鉴定得到pTol2-HS-GT2MP-EGFP质粒作为空白对照(Empty vector control, EVC), EVC质粒序列上传至NCBI(MW 698954)。

以基因组DNA为模板, 用引物dr334F/R(表 1)扩增出候选调控区dr334。同时通过含部分dr334序列的引物vector-HS-F/R (表 1), 利用反向PCR扩增出pTol2-HS-GT2MP-EGFP质粒骨架。用单片段快速克隆试剂盒将dr334插入pTol2-HS-GT2MP-EGFP质粒的HBB-5'HS5绝缘子和GT2MP之间, 得

到质粒pTol2-dr334:GT2MP-EGFP (简称dr334)。

1.6 显微注射、计数及统计分析

收集TB斑马鱼内交产生的胚胎, 向1-细胞阶段的胚胎中注射dr344质粒和Tol2转座酶mRNA混合物, 同时以EVC质粒作为对照。每颗胚胎中约注射2 nL的混合物, 其中质粒约为50 pg, 转座酶mRNA约为100 pg。注射后的胚胎置于Blue water中于28.5℃进行培养。分别于12 hours post fertilization (hpf)、24 hpf、36 hpf和48 hpf对注射胚胎的荧光进行观察和拍照。首先用1% MS222对胚胎进行麻醉, 再用1%的甲基纤维素(Sigma, M0387-100G)固定胚胎后置于玻底培养皿, 在荧光显微镜(Axio Observer Z1, Zeiss)下进行拍照。最后, 对具有不同荧光模式组胚胎数目使用R函数fisher.test()进行Fisher精确检验(Fisher's exact test), 计算获得对应P值及优势比(Odds ratio, OR)。

2 结果

2.1 成年斑马鱼肌肉高表达候选基因筛选

经调查发现PRJNA255848和PRJNA263496两项研究中均包含斑马鱼肌肉、脑、心、肝、肾和肠的转录组测序数据。利用该两项研究的转录组数据, 分别进行上述组织间基因表达差异分析, 筛选得到肌肉相对其他组织高表达($\log_2FC \geq 2$)的基因。在各组织比较结果取交集后, 在PRJNA255848数据中得到247个基因, 其在肌肉中表达量均高于其他5个组织; 在PRJNA263496中得到370个肌肉高表达基因。进一步对两组肌肉高表达基因取交集, 获得183个肌肉高表达的基因用于后续研究(图 1A)。为确定这些肌肉高表达基因参与的主要功能, 我们通过基因本体(GO)富集分析, 发现其中157个基因(85.8%)获得GO注释结果, 富集GO集中在骨骼肌细胞增殖调节, 骨骼肌收缩调节, 骨骼肌肌球蛋白粗丝组装, 骨骼肌纤维发育, 钙离子跨膜转运等(图 1B)。用类似方法, 我们筛选到了234个肌肉低表达对照基因。

2.2 多个肌肉高表达基因上游存在序列相似的DNA基序簇

我们通过Ancora数据库查找了上述斑马鱼肌肉高表达基因近端调控区中的跨物种保守元件(HCNE), 发现24个基因的调控区中存在跨物种非编码保守元件(表 2), 提示这些基因可能存在种间保守的转录调控机制。通过查找zfin(<http://zfin.org/>)表达数据库数据发现, 除个别基因数据缺失外, 绝大多数基因都在48hpf在肌肉中有表达。为探讨是否存在基因间保守的表达调控, 我们以该24个近端调控区序列为查找目标, 以234个肌肉低表达基因的近

端调控区序列作为对照, 使用MEME的Discriminative mode算法进行比较, 在5个基因[*itgb11*, *obs11b*, *zgc:9242*, *myh6*(Ensembl 103版中被注释为*myh7l*), *mylk4a*]的上游均发现了一段序列相似的保守DNA片段。片段内部均包含成簇排列且顺序一致的6个DNA基序(图2), 片段长度为292—347 bp(表3)。多序列比对发现5个DNA片段两两之间的序列相似度为78.62%—84.19%。这些DNA片段在肌肉高表达基因间的序列保守性(图2A), 提示其可能在调控基因组织特性表达中具有一定功能。根据片段区域信息(表3)在Ancora网站查询, 我们发现这些区域与Ancora中保守元件不重合(结果未展示), 提示其可能具有物种特异性。荧光定量PCR结果发现, 这5个基因在斑马鱼成鱼肌肉中的表达量均明显高于其他5个组织(图2B), 进一步验证了转录组分析的结果。这些发现提示上述5个DNA片段可能对调控这些基因在肌肉中高表达起到作用。

2.3 目标DNA片段显示增强基因肌肉特异性表达功能

为了验证上述DNA基序簇的转录调控功能, 我们选取了在肌肉中表达丰富的*zgc:92429* 基因^[30]上游的DNA基序簇区域作为研究对象。我们克隆得到了该基因上游区域334 bp的DNA片段, 经测序验证, 该片段序列与斑马鱼GRCz11版本基因组对应区域序列一致。将其插入在pTol2-GT2MP-EGFP的Gata2 最小启动子的上游, 构建了荧光蛋白表达报告质粒, 该质粒命名为pTol2-*dr334*:GT2MP-EGFP(图3A, 简称dr334)。我们同时构建了不含克

隆片段的载体pTol2-HS-GT2MP-EGFP作为对照(EVC)。分别将两种质粒注射到同批次的TB斑马鱼1细胞期胚胎中, 并在12hpf、24hpf、36hpf和48hpf四个发育时期通过荧光显微镜进行观察。发现在各时期, 均有不同比例的胚胎显示出肌肉组织特异荧光信号(图3B)。另外, 根据鱼体内荧光信号情况, 各时期的胚胎可分为3类: (1) 没有荧光信号(N); (2) 有荧光信号但非肌肉组织特异性表达(NSF); (3) 荧光信号有明显的肌肉组织特异性(MF)(图3C)。对各时期三类胚胎进行计数统计(表4和图3D), 发现在四个时期, 注射dr334的胚胎可观察到的荧光个体(NSF+MF)相对无荧光个体(N)的比例均高于对照组(Odds ratio范围: 1.643—3.881), 且在24hpf($P=0.006$)和48hpf($P<0.001$)具有极显著统计学差异。另外, dr334组胚胎的肌肉特异信号个体(MF)相对非肌肉特异信号个体(N+NSF)的比例均高于对照组(Odds ratio范围: 1.311—6.487), 且在发育后期更加明显, 48hpf时具有极显著统计学差异($P<0.001$)。这说明该长度为334 bp的DNA片段不仅具有增强报告基因表达的作用, 且其增强作用带有肌肉组织特异性。

2.4 DNA基序簇内存在肌肉形成相关转录因子结合位点

我们发现的DNA基序簇片段, 在多个肌肉高表达基因间序列保守, 且在体内显示增强基因在肌肉中表达的功能, 提示可能存在同类的转录因子与该区域结合调控这些基因在肌肉中高表达。我们通过Tomtom(<http://meme-suite.org/tools/tomtom>)对目

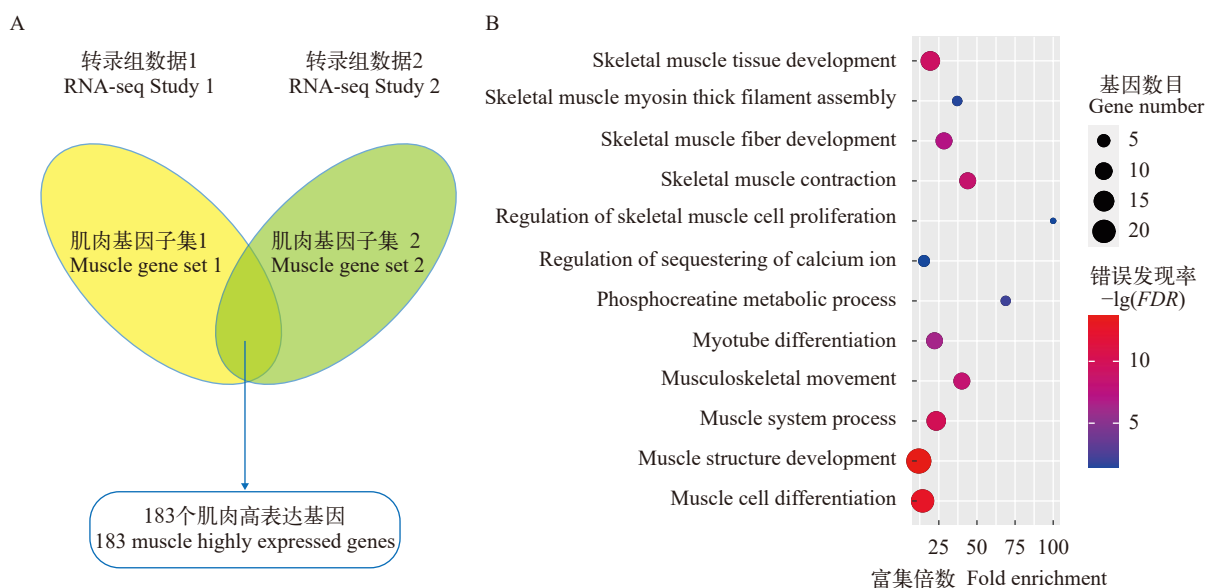


图1 转录组分析筛选肌肉高表达基因(A)和肌肉高表达基因GO生物过程富集结果(B)

Fig. 1 Screening for genes that are highly expressed in muscle (A) and GO biological process enrichment of the high-expression gene of muscle (B)

标DNA基序进行转录因子结合位点预测。结果在 motif 2、motif 3和motif 4中分别发现了 Xbp1(X-box binding protein 1)^[31]和twist2(twist family bHLH transcription factor 2)^[32]、bhlha15(basic helix-loop-helix family, member a15)^[33]、myod1(myogenic differentiation 1)^[34]、twist1(twist family bHLH transcription factor 1)^[35]等多个肌肉形成相关转录因子

的结合位点(图 4), 预测P值分别为1.34e-02、2.16e-03、1.71e-03、1.09e-02和1.64e-02, 提示该区域可能作为这些转录因子的靶点, 参与基因表达调控。

3 讨论

3.1 斑马鱼肌肉高表达基因的非编码元件

鱼类肌肉形成及相关基因的表达调控是水生

表 2 Ancora数据库中具有基因近端跨物种保守元件的斑马鱼肌肉高表达基因

Tab. 2 Muscle highly expressed genes with proximal CNEs in Ancora database

基因ID Gene ID	NCBI 登录号 NCBI accession number	基因名称 Gene name	基因全称 Gene description	对应物种 Species	ZFIN数据库是否肌 肉表达 Muscle expression in ZFIN
ENSDARG00000040985	554050	<i>itgbl1</i>	integrin, beta-like 1	CC	是, 0.75—48hpf
ENSDARG00000077388	606585	<i>obs1b</i>	obscurin like cytoskeletal adaptor 1b	CI, CC	是, 0.75—48hpf
ENSDARG00000030176	445063	<i>zgc:92429</i>	<i>zgc:92429</i>	AM	是, 10.22—48hpf
ENSDARG00000098747	100329748	<i>myh6</i>	myosin heavy chain 6	TR	肌肉特异性表达, 0.75—90dpf
ENSDARG00000091260	566845	<i>mylk4a</i>	myosin light chain kinase family, member 4a	AM, LO, GA	NA
ENSDARG00000014259	30436	<i>eya1</i>	EYA transcriptional coactivator and phosphatase 1	CI, AM, CC, LO, OL, GA, TN, HS, MM, TR	是, 0.75—90dpf
ENSDARG00000002582	246222	<i>tbx15</i>	T-box transcription factor 15	AM	是, 0.75—90dpf
ENSDARG00000003081	321053	<i>mybphb</i>	myosin binding protein Hb	AM	是, 10.33—48hpf
ENSDARG00000005841	394068	<i>tnni2a.2</i>	troponin I type 2a (skeletal, fast), tandem duplicate 2	AM	是, 5.25—72hpf
ENSDARG00000013755	58037	<i>actn3a</i>	actinin alpha 3a	AM	是, 0.75—90dpf
ENSDARG00000014976	553696	<i>lims2</i>	LIM and senescent cell antigen-like domains 2	GA	是, 10.33—48hpf
ENSDARG00000016391	386968	<i>calcocol1b</i>	calcium binding and coiled-coil domain 1b	TN, AM	NA
ENSDARG00000019342	404273	<i>chrnd</i>	cholinergic receptor, nicotinic, delta (muscle)	AM, LO, OL, TN, TR	是, 24—90dpf
ENSDARG00000020890	30441	<i>tmod4</i>	tropomodulin 4 (muscle)	AM, GA	肌肉特异性表达, 72hpf
ENSDARG00000026473	404627	<i>six1b</i>	SIX homeobox 1b	CI, AM, CC, LO	是, 5.25—72hpf
ENSDARG00000039304	494168	<i>six1a</i>	SIX homeobox 1a	CI	是, 10.33—72hpf
ENSDARG00000034588	564977	<i>scn4ab</i>	sodium channel, voltage- gated, type IV, alpha, b	LO	是, 0.75hpf, 5.25hpf, 24hpf, 48hpf, 90.0 dpf
ENSDARG00000035458	494489	<i>atp2a1l</i>	ATPase sarcoplasmic/endoplasmic reticulum Ca ²⁺ transporting 1, like	AM, OL, GA, TN, TR	否, 90dpf
ENSDARG00000044155	100000492	<i>mafaa</i>	v-maf avian musculoaponeurotic fibrosarcoma oncogene homolog Aa	CI, AM, CC, GA, OL, TN, TR	是10.33—48hpf
ENSDARG00000053773	606659	<i>vgl12b</i>	vestigial-like family member 2b	CC	是, 10.33—72hpf
ENSDARG00000056209	572909	<i>myoz1a</i>	myozenin 1a	CC	肌肉特异性表达 10.33—72hpf
ENSDARG00000077157	100006620	<i>synpo2b</i>	synaptopodin 2b	CI	是, 10.33—48hpf
ENSDARG00000095217	393472	<i>zgc:66156</i>	<i>zgc:66156</i>	AM	是, 10.33—72hpf
ENSDARG00000096257	100007086	<i>si:ch73-367p23.2</i>	<i>si:ch73-367p23.2</i>	AM	NA

注: CI. 草鱼(*Ctenopharyngodon idellus*); AM. 洞穴鱼(*Astyanax mexicanus*); CC. 鲤(*Cyprinus carpio*); LO. 斑点雀鳊(*Lepisosteus oculatus*); OL. 青鳉(*Oryzias latipes*); GA. 棘鱼(*Gasterosteus aculeatus*); TN. 河豚(*Tetraodon nigroviridis*); HS. 人(*Homo sapiens*); MM. 小鼠(*Mus musculus*); TR. 红鳍东方鲀(*Takifugu rubripes*)

生物学和水产科学研究中的热点问题。组织特异性DNA顺式调控元件可以作为特定转录因子的结合位点, 调控基因在特定组织表达^[7]。为研究可调控基因在肌肉中表达的DNA功能元件, 我们通过分析公共数据库中的转录组和基因组数据, 在5个肌肉高表达基因的近端均发现了一段保守的DNA区域, 猜测其可能作为功能元件参与基因在肌肉中表达的调控。我们进一步克隆获得了*zgc:92429*对应区段的DNA片段, 并将其连接到eGFP荧光报告基因上游, 使用to12转座系统将克隆DNA片段转入斑马鱼胚胎基因组。通过观察eGFP在体内表达情况,

发现实验组胚胎在发育后期大比例显示出肌肉组织特异性荧光信号, 说明该DNA片段可能作为转录调控元件增强基因肌肉特异性表达。

我们观察到注射了实验组和对照组质粒的胚胎均可能会表现为3种情况: (1)没有荧光信号; (2)非特异性荧光信号; (3)肌肉组织特异荧光信号。报告基因表达除受我们关注的DNA元件影响外, 还可能会受到其他因素的影响。一方面, 注射和转座实验存在一定系统误差, 部分胚胎未将eGFP基因连同调控元件整合到基因组中, 造成没有荧光蛋白生成或者随机瞬时生成部分荧光蛋白;

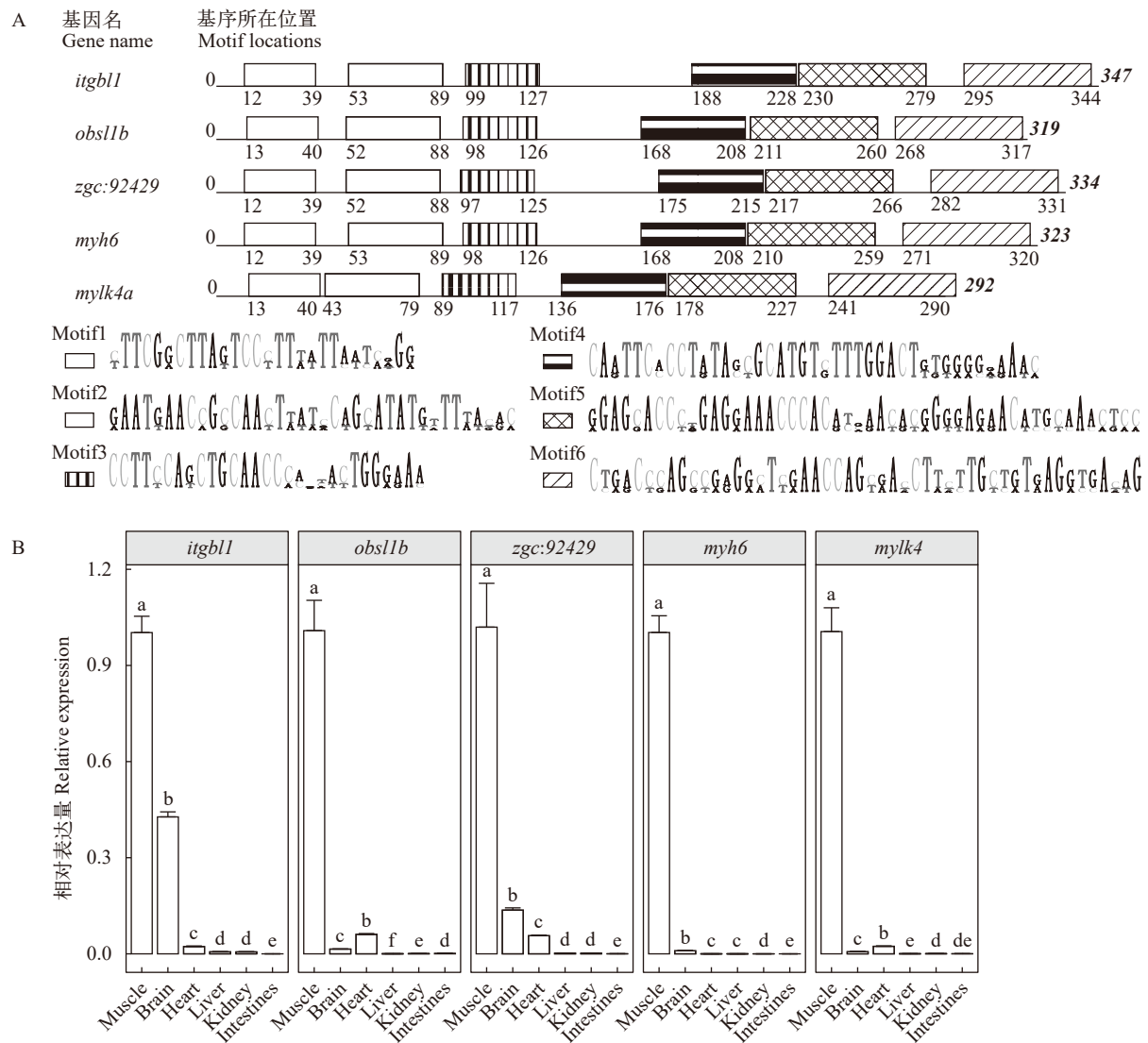


图2 肌肉高表达基因及其近端保守DNA片段

Fig. 2 Muscle highly expressed genes and proximal conserved DNA sequences

A. 5个基因近端的顺序相同的DNA基序簇, 不同线条填充的矩形代表不同种基序; B. 5个基因在不同组织的qPCR相对定量结果, Y轴为各组织相对肌肉的表达量变化($2^{-\Delta\Delta C_t}$), 误差线为标准误; 不同字母表示差异显著

A. DNA motif clusters with same orders for five genes. The rectangles with different line patterns represent the different motifs; B. The relative expression of 5 genes in different tissues. Y axes are relative expression to muscle ($2^{-\Delta\Delta C_t}$). Error bars are standard errors of mean. Different letters indicate significant differences

表 3 基因近端保守区序列信息

Tab. 3 The sequence information of proximal conserved element for 5 genes

片段区域 Region	基因ID Gene ID	基因名称 Gene name	与基因起始 位点的距离 Distance to gene start site (bp)	长度 Length (bp)
9:31751740-31752087	ENSDARG0000040985	<i>itgbl1</i>	-541	347
9:41789619-41789938	ENSDARG0000077388	<i>obs11b</i>	183	319
7:23748122-23748456	ENSDARG0000030176	<i>zgc:92429</i>	-2478	334
24:40746974-40747297	ENSDARG0000098747	<i>myh6</i>	-2325	323
2:392063-392355	ENSDARG0000091260	<i>mylk4a</i>	-3611	292

另一方面, eGFP基因插入到基因组其他增强元件附近的时候也可能被激活表达, 载体设计中加入的绝缘子可能未完全阻断其他功能元件的影响。因此各组胚胎的荧光信号表达模式均存在一定随机性。经对大量胚胎计数统计分析后发现, 实验组中肌肉组织特异性荧光信号的比例显著高于对照组, 说明目标区域的DNA片段具有增强报告基因在肌肉中表达的作用。如对该系统进行进一步优化, 其有望作为基因工程的工具, 高效引导基因在肌肉中特异表达。

另外, 我们发现的DNA基序簇所在基因组区域(表 3), 并未与Ancora数据库中跨物种保守元件重合, 提示其转录调控功能可能具有物种特异性。另一方面也说明, 对基因间序列保守性分析进行非编码调控元件预测, 可以作为跨物种序列保守分析的补充, 对基因组非编码区进行功能注释。

3.2 多个肌肉高表达基因可能协同作用参与肌肉结构组成

本文发现5个基因上游存在相似的DNA序列, 这些基因均在肌肉的结构和功能中起到重要作用。*zgc:92429*在斑马鱼肌肉中高度表达^[36], 在人类、小鼠等物种的同源基因为*Itgbl1bp2*。该基因在小鼠骨骼肌和心肌中高表达且作为一种肌肉特异性信号蛋白^[37]。该基因产物通过与一种细胞膜受体整合素1(*itgbl1*)的胞质内区域结合, 调节肌肉的生长、收缩和修复^[38]。同时, 我们在*itgbl1*上游也发现了同样排列顺序的DNA基序簇, 提示以上2个基因的表达受到同类信号的调控, 有助于其协同作用。在斑马鱼中, *Obs11b*(Obscurin like cytoskeletal adaptor 1b)大量存在于肌节中参与细胞基质、细胞和胞内细胞骨架连接的稳定^[39]。肌球蛋白重链6(Myh6)构成II型肌球蛋白的一部分, II型肌球蛋白在肌节中为肌肉收缩提供所需的机械力^[40,41]。需

要注意的是, 在多数文献的报道中*myh6*在胚胎时期的心肌中表达丰富^[42], 由于在较新版本的ensembl数据库中也将该基因注释为*myh7l*, 经过调查发现*myh7*基因在成年时也会在骨骼肌中表达^[43,44], 这也部分验证了我们的实验结果, 我们推测该基因可能在不同时期有不同的表达模式。*mylk4*基因作为MYLK家族的一员, 在肌肉发育中起重要作用^[45], 该基因会在心力衰竭中下调并可能引起肌丝的磷酸化的下降, 从而影响心肌细胞的骨架结构^[46]。此外, 以上5个基因均在人或小鼠中参与肌节的构成^[41-43,49,50], 说明这些基因可能在斑马鱼肌肉中协同作用维持肌肉的正常形态和功能, 其表达可能具有相近的转录调控机制。

3.3 调控元件可能与Myod为中心的多个转录因子组成的复合物相结合

上述5个基因的上游发现了相同的DNA基序簇, 提示可能存在同类的转录因子通过与这些候选区域结合, 调控这些基因的表达。通过Tomtom (<http://meme-suite.org/tools/tomtom>)预测候选区域DNA基序的转录因子结合位点, 发现该区域可能存在Myod1、Xbp1、Bhlha15、Twist1和Twist2等多个肌肉形成相关的转录因子的结合位点, 提示该区域可能作为这些转录因子的靶点, 参与基因表达调控。Myod作为MRF的成员之一在特异肌基因转录调控中起到总开关作用, 推动肌源性细胞谱系的形成^[49], 且可诱导非肌肉细胞(如成纤维细胞、成软骨细胞、视网膜色素上皮细胞等)向成肌细胞转化^[8]。有研究发现在C2C12细胞中, *xbp1*被敲降后Myod等生肌调节因子的表达下调, 并且细胞向成肌细胞的分化受到抑制, 猜测Xbp1可能通过诱导Cdk5来调控*myod*家族基因参与成肌细胞早期分化^[50]。*bhlha15*是一种*myod*的负调控因子, 该基因通过与Myod形成异二聚体或者与自身形成二聚体来占据E-box区域, 从而使得胚胎时期的肌肉分化增殖保持动态平衡^[33]。Twist同样属于bHLH转录因子家族^[51], 小鼠Twist可以与Myod结合抑制后者的作用^[52]且表达*twist2*的肌源性干细胞被报道作为一种新的干细胞类型参与了肌肉的生长和再生^[53]。结合上述研究, 我们发现的候选转录调控元件可能通过与Myod为中心的多种转录因子组成的调控复合体结合, 调控基因在早期发育的肌肉细胞中定向表达。但基于目前的分析, 我们尚不清楚除上述候选转录因子结合位点外, 该保守DNA区段的是否还存在其他重要的功能区, 也不了解这些候选结合位点是否同时发挥作用。此外, 也不排除该元件部分区域具有启动子活性。研究该片段的具体转录调控机制, 还需要

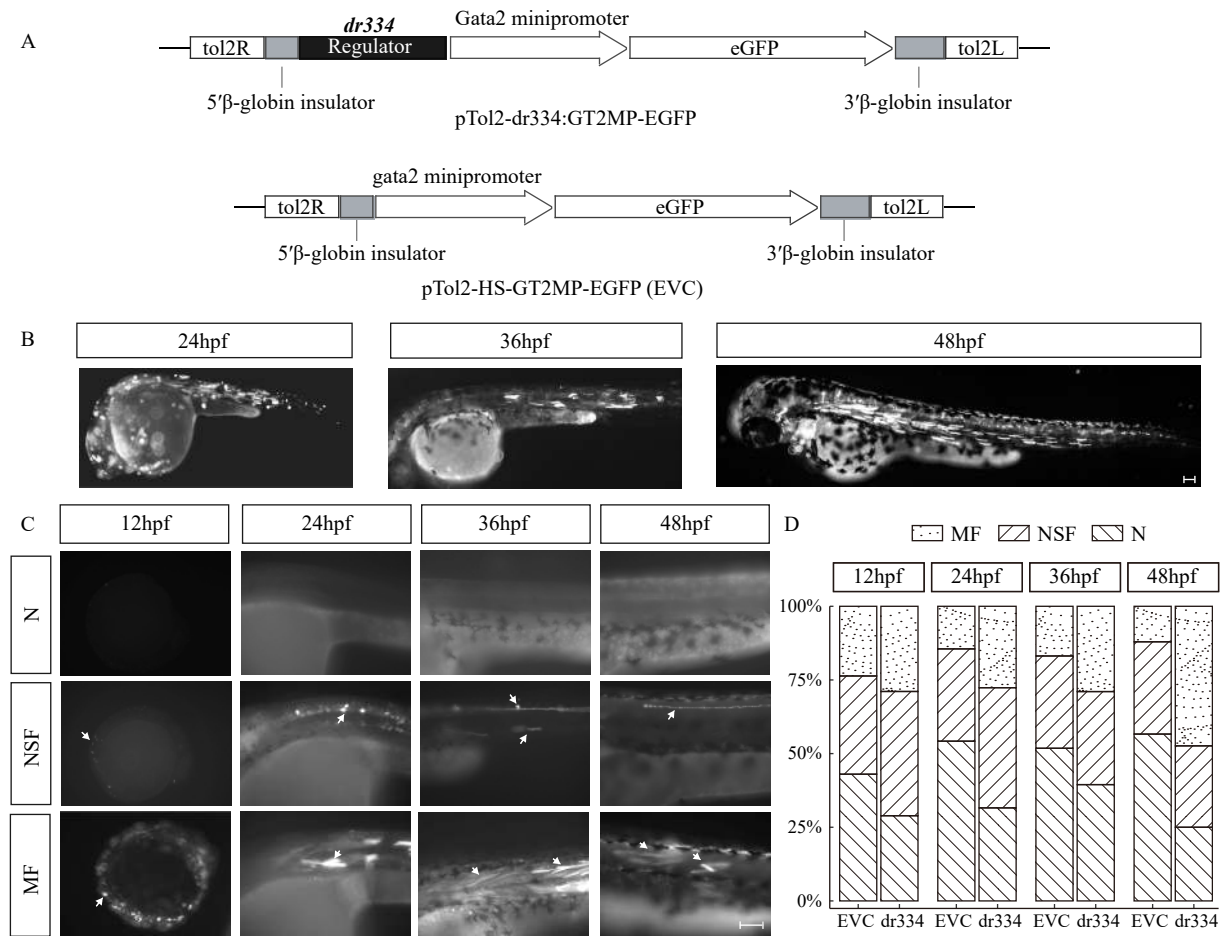


图 3 荧光报告载体结构及注射后胚胎荧光信号

Fig. 3 Structure of eGFP expression reporter plasmids and fluorescence signals in zebrafish embryos

A. dr334质粒(上)及对照质粒(下)功能区的结构; B. 3个发育时期显示肌肉特异性荧光信号的胚胎, 标尺=1 mm; C. 4个发育时期3种荧光信号类型胚胎的局部图, 标尺=1 mm; N代表没有荧光信号, NSF代表有非肌肉特异性信号, MF代表有肌肉特异性荧光信号; D. 4个发育时期3种荧光信号类型胚胎的比例; EVC表示注射空载体质粒的胚胎; dr334表示注射dr334质粒的胚胎

A. Structure of functional region for dr334 plasmid (up) and control (bottom) plasmids; B. Embryos with muscle specific fluorescent signal in three developmental stages, Scale bar = 1 mm; C. Embryos with three types of fluorescent signals for four developmental stages; N is no fluorescence signal, NSF is non-specific fluorescence, and MF is muscle specific fluorescence. Scale bar = 1 mm. D. Embryo counts proportions upon fluorescence types or four developmental stages. EVC represents empty vector control plasmids were injected in to embryos; dr334 represents dr334 plasmids were injected

表 4 实验组与对照组的荧光观察结果计数分析

Tab. 4 Embryo counts upon fluorescence levels and proportion test between empty vector control and dr334 vector

时期Stage	对照组胚胎数 Control			实验组胚胎数 dr334			荧光 vs. 无荧光** Fluorescence vs. Non- Fluorescence		肌肉 vs. 非肌肉*** Muscle vs. Non-Muscle***	
	N	NSF	MF*	N	NSF	MF*	P-value	Odds Ratio (95%CI)	P-value	Odds Ratio (95%CI)
12hpf	40	31	22	24	35	24	0.061	1.849 (0.948—3.66)	0.493	1.311 (0.633—2.726)
24hpf	45	26	12	24	31	21	0.006	2.550 (1.279—5.177)	0.051	2.247 (0.959—5.479)
36hpf	43	26	14	30	24	22	0.152	1.643 (0.838—3.251)	0.088	1.999 (0.883—4.655)
48hpf	47	26	10	19	21	36	0.000	3.881 (1.894—8.19)	0.000	6.487 (2.798—16.267)

注: N. 未观察到荧光; NSF. 非特异性组织荧光表达; MF*. 肌肉中特异性荧光表达。**dr334与对照组的荧光(NSF+MF)相对于非荧光(N)的比值比。***dr334与对照组肌肉特异荧光(MF)相对非肌肉荧光(N+NSF)的比值比。P值来自Fisher'精确检验

Note: N. no fluorescence observed; NSF. fluorescence observed by non-specific tissue expressed; MF*. fluorescence expressed specifically in muscle. **Odds ratio of Fluorescence (NSF+MF) for dr334 group, compared to Non- Fluorescence (N). ***Odds ratio of Muscle (MF) for dr334 group, compared to Non-muscle (N+NSF). P-values are from Fisher's exact test

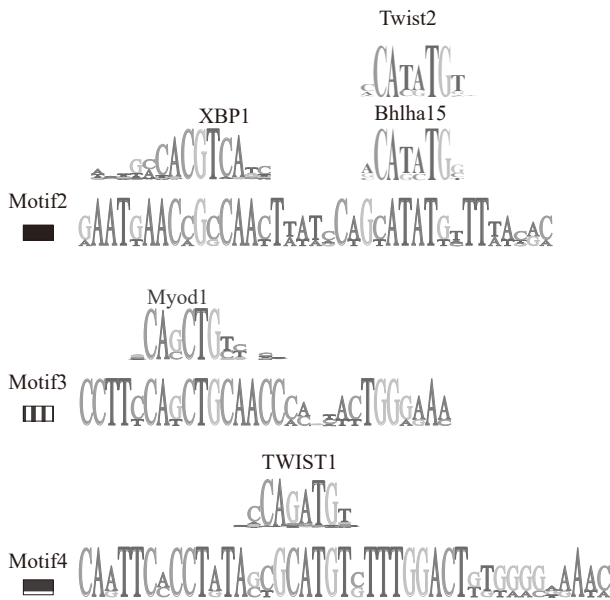


图4 Tomtom预测得到的DNA基序簇中肌肉相关的转录因子结合位点

Fig. 4 Muscle-related transcription factor binding sites in motif clusters predicted by Tomtom

更多信息学分析以及更多实验证据,如碱基突变功能验证、DNA与蛋白作用检测、启动子活性检测等实验,进行深入探讨。

综上所述,本研究通过生信分析在多个斑马鱼肌肉高表达基因上游发现一段序列保守的DNA区域,体内荧光报告基因表达检测发现该DNA片段可能是一段肌肉组织特异的转录调控元件,可能通过与Myod为中心的多种转录因子结合调控基因表达。该发现为进一步研究肌肉形成相关基因表达调控的分子机制奠定了基础。该片段也有望作为基因工程的工具,引导基因在肌肉中特异表达。

参考文献:

- [1] Zhou R X, Huang B, Meng T, *et al.* Cloning and ontogenetic expression analysis of the alkali myosin light chain gene in *Siniperca chuatsi* [J]. *Acta Hydrobiologica Sinica*, 2010, **34**(5): 927-934. [周瑞雪, 黄斌, 蒙涛, 等. 鳊碱性肌球蛋白轻链基因cDNA的克隆及其发育表达分析 [J]. 水生生物学报, 2010, **34**(5): 927-934.]
- [2] Ennion S, Gauvry L, Butterworth P, *et al.* Small-diameter white myotomal muscle fibres associated with growth hyperplasia in the carp (*Cyprinus carpio*) express a distinct myosin heavy chain gene [J]. *The Journal of Experimental Biology*, 1995, **198**(7): 1603-1611.
- [3] Blankvoort S, Witter M P, Noonan J, *et al.* Marked diversity of unique cortical enhancers enables neuron-specific tools by enhancer-driven gene expression [J]. *Current Biology*, 2018, **28**(13): 2103-2114.
- [4] Shima Y, Sugino K, Hempel C M, *et al.* A Mammalian enhancer trap resource for discovering and manipulating neuronal cell types [J]. *eLife*, 2016(5): e13503.
- [5] Symon A, Harley V. SOX9: A genomic view of tissue specific expression and action [J]. *The International Journal of Biochemistry & Cell Biology*, 2017(87): 18-22.
- [6] Carroll S B. Evo-Devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution [J]. *Cell*, 2008, **134**(1): 25-36.
- [7] Turner E E, Cox T C. Genetic evidence for conserved non-coding element function across species-the ears have it [J]. *Frontiers in Physiology*, 2014(5): 7.
- [8] Marcovitz A, Jia R, Bejerano G. "Reverse Genomics" predicts function of human conserved noncoding elements [J]. *Molecular Biology and Evolution*, 2016, **33**(5): 1358-1369.
- [9] Soleimani V D, Nguyen D, Ramachandran P, *et al.* Cis-regulatory determinants of MyoD function [J]. *Nucleic Acids Research*, 2018, **46**(14): 7221-7235.
- [10] Choi J, Costa M L, Mermelstein C S, *et al.* MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes [J]. *Proceedings of the National Academy of Sciences*, 1990, **87**(20): 7988-7992.
- [11] Berkes C A, Tapscott S J. MyoD and the transcriptional control of myogenesis [J]. *Seminars in Cell & Developmental Biology*, 2005, **16**(4-5): 585-595.
- [12] Burghoorn J, Piasecki B P, Crona F, *et al.* The in vivo dissection of direct RFX-target gene promoters in *C. elegans* reveals a novel cis-regulatory element, the C-box [J]. *Developmental Biology*, 2012, **368**(2): 415-426.
- [13] Jaeger S A, Chan E T, Berger M F, *et al.* Conservation and regulatory associations of a wide affinity range of mouse transcription factor binding sites [J]. *Genomics*, 2010, **95**(4): 185-195.
- [14] Woolfe A, Goodson M, Goode D K, *et al.* Highly Conserved Non-Coding Sequences Are Associated with Vertebrate Development [J]. *PLoS Biology*, 2005, **3**(1): e7.
- [15] Polychronopoulos D, King J W D, Nash A J, *et al.* Conserved non-coding elements: developmental gene regulation meets genome organization [J]. *Nucleic Acids Research*, 2017, **45**(22): 12611-12624.
- [16] Pasquier J, Cabau C, Nguyen T, *et al.* Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database [J]. *BMC Genomics*, 2016, **17**(1): 368.
- [17] Hu P, Liu M, Zhang D, *et al.* Global identification of the genetic networks and *cis*-regulatory elements of the cold response in zebrafish [J]. *Nucleic Acids Research*, 2015, **43**(19): 9198-9213.
- [18] Bolger A M, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data [J]. *Bioinforma-*

- tics*, 2014, **30**(15): 2114-2120.
- [19] Kim D, Paggi J M, Park C, *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype [J]. *Nature Biotechnology*, 2019(37), 907-915.
- [20] 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.
- [21] Robinson M D, McCarthy D J, Smyth G K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data [J]. *Bioinformatics*, 2010, **26**(1): 139-140.
- [22] Ashburner M, Ball C A, Blake J A, *et al.* Gene Ontology: tool for the unification of biology [J]. *Nature Genetics*, 2000, **25**(1): 25-29.
- [23] Ginestet C. ggplot2: Elegant graphics for data analysis [J]. *Journal of the Royal Statistical Society*, 2011, **174**(1): 245-246.
- [24] Engström P G, Fredman D, Lenhard B. Ancora: a web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes [J]. *Genome Biology*, 2008, **9**(2): R34.
- [25] Bailey T L, Boden M, Buske F A, *et al.* MEME Suite: tools for motif discovery and searching [J]. *Nucleic Acids Research*, 2009, 37(Web Server issue): W202-W208.
- [26] Gupta S, Stamatiyannopoulos J A, Bailey T L, *et al.* Quantifying similarity between motifs [J]. *Genome Biology*, 2007, **8**(2): R24.
- [27] Khan A, Fornes O, Stigliani A, *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework [J]. *Nucleic Acids Research*, 2018, **46**(D1): D260-D266.
- [28] Xue Y L, Xiao A, Wen L, *et al.* Generation and Characterization of blood vessel specific EGFP transgenic zebrafish via Tol2 transposon mediated enhancer trap screen [J]. *Progress in Biochemistry and Biophysics*, 2010, **37**(7): 720-727.
- [29] Farrell C M, West A G, Felsenfeld G. Conserved CTCF insulator elements flank the mouse and human β -globin loci [J]. *Molecular and Cellular Biology*, 2002, **22**(11): 3820-3831.
- [30] Kudoh T, Tsang M, Hukriede N A, *et al.* A gene expression screen in zebrafish embryogenesis [J]. *Genome Research*, 2001, **11**(12): 1979-1987.
- [31] Yamamoto K, Yoshida H, Kokame K, *et al.* Differential contributions of ATF6 and XBP1 to the activation of endoplasmic reticulum stress-responsive cis-acting elements ERSE, UPRE and ERSE-II [J]. *Journal of Biochemistry*, 2004, **136**(3): 343-350.
- [32] Kophengnavong T, Michnowicz J E, Blackwell T K. Establishment of distinct MyoD, E2A, and Twist DNA binding specificities by different basic region-DNA conformations [J]. *Molecular and Cellular Biology*, 2000, **20**(1): 261-272.
- [33] Lemerrier C, To R Q, Carrasco R A, *et al.* The basic helix-loop-helix transcription factor Mist1 functions as a transcriptional repressor of MyoD [J]. *The EMBO Journal*, 1998, **17**(5): 1412-1422.
- [34] Cao Y, Yao Z, Sarkar D, *et al.* Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming [J]. *Developmental Cell*, 2010, **18**(4): 662-674.
- [35] Chang A T, Liu Y, Ayyanathan K, *et al.* An evolutionarily conserved DNA architecture determines target specificity of the TWIST family bHLH transcription factors [J]. *Genes & Development*, 2015, **29**(6): 603-616.
- [36] Thisse B, Thisse C. Fast Release Clones: A High Throughput Expression Analysis. ZFIN Direct Data Submission [DB]. <http://zfin.org>, 2004.
- [37] Palumbo V, Segat L, Padovan L, *et al.* Melusin gene (IT-GB₁BP₂) nucleotide variations study in hypertensive and cardiopathic patients [J]. *BMC Medical Genetics*, 2009, **10**(1): 140.
- [38] Brancaccio M, Guazzone S, Menini N, *et al.* Melusin is a new muscle-specific interactor for beta (1) integrin cytoplasmic domain [J]. *Journal of Biological Chemistry*, 1999, **274**(41): 29282-29288.
- [39] Geisler S B, Robinson D, Hauringa M, *et al.* Obscurin-like 1, OBSL1, is a novel cytoskeletal protein related to obscurin [J]. *Genomics*, 2007, **89**(4): 521-531.
- [40] Klos M, Mundada L, Banerjee I, *et al.* Altered myocyte contractility and calcium homeostasis in alpha-myosin heavy chain point mutations linked to familial dilated cardiomyopathy [J]. *Archives of Biochemistry and Biophysics*, 2017(615): 53-60.
- [41] Granados-Riveron J T, Ghosh T K, Pope M, *et al.* Alpha-cardiac myosin heavy chain (MYH6) mutations affecting myofibril formation are associated with congenital heart defects [J]. *Human Molecular Genetics*, 2010, **19**(20): 4007-4016.
- [42] Shih Y H, Zhang Y, Ding Y, *et al.* Cardiac transcriptome and dilated cardiomyopathy genes in zebrafish [J]. *Circulation Cardiovascular Genetics*, 2015, **8**(2): 261-269.
- [43] Chopra S S, Stroud D M, Watanabe H, *et al.* Voltage-gated sodium channels are required for heart development in zebrafish [J]. *Circulation Research*, 2010, **106**(8): 1342-1350.
- [44] Bagatto B, Franc J, Liu B, *et al.* Cadherin2 (N-cadherin) plays an essential role in zebrafish cardiovascular development [J]. *BMC Developmental Biology*, 2006(6): 23.
- [45] Zheng L, Zhang G M, Dong Y P, *et al.* Genetic variant of MYLK4 gene and its association with growth traits in Chinese cattle [J]. *Animal Biotechnology*, 2019, **30**(1): 30-35.
- [46] Herrero I, Roselló-Lleti E, Rivera M, *et al.* RNA-sequencing analysis reveals new alterations in cardiomyocyte

- cytoskeletal genes in patients with heart failure [J]. *Laboratory Investigation*, 2014, **94**(6): 645-653.
- [47] Hsieh F C, Lu Y F, Liao I, *et al.* Zebrafish *VCAP1X2* regulates cardiac contractility and proliferation of cardiomyocytes and epicardial cells [J]. *Scientific Reports*, 2018, **8**(1): 7856.
- [48] Gerull B, Gramlich M, Atherton J, *et al.* Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy [J]. *Nature Genetics*, 2002, **30**(2): 201-204.
- [49] Zammit P S. Function of the myogenic regulatory factors Myf5, MyoD, Myogenin and MRF4 in skeletal muscle, satellite cells and regenerative myogenesis [J]. *Seminars in Cell & Developmental Biology*, 2017(72): 19-32.
- [50] Tokutake Y, Yamada K, Hayashi S, *et al.* IRE1-XBP1 pathway of the unfolded protein response is required during early differentiation of C2C12 Myoblasts [J]. *International Journal of Molecular Sciences*, 2019, **21**(1): 182.
- [51] Thisse B, Messal M E, Perrin-Schmitt F. The twist gene: isolation of a Drosophila zygote gene necessary for the establishment of dorsoventral pattern [J]. *Nucleic Acids Research*, 1987, **15**(8): 3439-3453.
- [52] Hamamori Y, Wu H Y, Sartorelli V, *et al.* The basic domain of myogenic basic helix-loop-helix (bHLH) proteins is the novel target for direct inhibition by another bHLH protein, Twist [J]. *Molecular and Cellular Biology*, 1997, **17**(11): 6563-6573.
- [53] Liu N, Garry G A, Li S, *et al.* A Twist2-dependent progenitor cell contributes to adult skeletal muscle [J]. *Nature Cell Biology*, 2017, **19**(3): 202-213.

IDENTIFICATION OF PROXIMAL CIS-REGULATORY ELEMENT FOR MUSCLE HIGHLY EXPRESSED GENES IN ZEBRAFISH

LU Cheng-Cheng^{1,2}, ZHAO Yi-Fan^{1,2}, FAN Chun-Xin^{1,2} and WANG Jian^{1,3}

(1. International Research Center for Marine Biosciences, Ministry of Science and Technology, Shanghai Ocean University, Shanghai 201306, China; 2. Institute for Marine Biosystem and Neuroscience, International Center for Marine Studies, Shanghai Ocean University, Shanghai 201306, China; 3. Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Ministry of Education, Shanghai Ocean University, Shanghai 201306, China)

Abstract: The related factors to the formation and growth traits of fish muscle are important topics in aquatic biology and aquaculture research. The expression and regulation of muscle component genes are essential to its tissue function maintenance and trait control, and the identification of cis-regulatory elements in muscle tissue helps explain the genetic basis of muscle formation. Conserved DNA sequences may be found among cis-regulatory elements, whose regulating genes have similar expression patterns. To predict the regulatory elements for zebrafish muscle gene expression, we explored the conservation features for DNA sequences in proximal non-coding regions of muscle highly expressed genes. By analyzing RNA-seq data of multiple zebrafish tissues from public database, we located muscle highly expressed genes as targets and lowly expressed genes as control, respectively. The GO enrichment analysis of these highly expressed genes confirmed their functions associated with muscle development. By discriminative mode of MEME motif discovery tools and using non-coding region of lowly expressed genes as background, we found five target conserved DNA regions, including six DNA motifs of the same sequence, around 300 bp in length, close to gene start sites of five muscle highly expressed genes. Meanwhile, DNA sequences of these five target regions had high pairwise identities (78.62%—84.19%). The results of qPCR confirmed the remarkably higher expression of these five genes in muscle than other tissues. We constructed an eGFP expression reporter plasmid containing the tol2 transposon system. One of the target regions, a 334 bp fragment at upstream of *zgc:9242*, was cloned into the plasmid at upstream of the eGFP driven by the base promoter. After the plasmid was injected into zebrafish embryos, a greater proportion of muscle specific fluorescence was observed in embryos carrying the target DNA fragment than in the control group (odds ratio=6.487, $P=0.000$ at 48 hpf), indicating that the 334 bp DNA fragment may enhance muscle gene expression. Using Tomtom motif comparison tools, we also found the candidate binding sites for MyoD and other transcription factors within DNA motifs. Our findings suggest that the DNA motif cluster fragments might act as transcriptional regulatory elements to specifically enhance zebrafish muscle gene expression. These results can help us better understanding the genetic basis for fish muscle gene expression and provide a new strategy for predicting tissue specific cis-regulatory elements by bioinformatics.

Key words: Cis-regulatory elements; Muscle; Gene expression; Zebrafish