

单细胞测序数据分析揭示MyoD介导成纤维细胞转分化为肌细胞的时序特征

阿力木江·吐地玉苏甫¹ 万鑫¹ 谷莹莹¹ 胡晓娣¹ 朱大海¹ 秦岩² 张勇^{1*} 曾平^{2*}

⁽¹⁾中国医学科学院基础医学研究所, 北京协和医学院基础医学院, 北京 100005;

⁽²⁾中国医学科学院北京协和医院保健医疗部, 北京 100730)

摘要 MyoD是骨骼肌细胞谱系特异性表达的转录因子, 对骨骼肌细胞命运决定发挥重要调控作用。在成纤维细胞中异位表达MyoD即可将其转分化为成肌细胞。但是, MyoD介导成纤维细胞转分化为肌细胞的时序调控特征并不十分清楚, 特别是对MyoD如何调控成纤维细胞相关基因表达下调缺乏系统性分析。该研究基于已发表的公共多组学数据集开展生物信息学再分析, 旨在单细胞分辨率下系统解析MyoD介导小鼠胚胎成纤维细胞(mouse embryonic fibroblast, MEF)转分化为骨骼肌细胞的时序特征, 揭示转分化过程中细胞命运转变的分子基础。该文整合了MyoD介导MEF细胞转分化不同时间点的单细胞转录组数据(scRNA-seq)、MyoD结合位点的数据(MyoD CUT&RUN)以及染色质开放性的数据(SHARE-seq)。采用Monocle2拟时序分析, 发现了MyoD诱导MEF细胞转分化轨迹呈现3个有序阶段: 早期成肌细胞(*Cdh15*⁺/*Mef2a*⁺)、肌细胞(*Myog*⁺/*Actc1*⁺)和终末分化的肌细胞(*Myh3*⁺/*Ckm*⁺)。其中, 转分化24 h时以早期成肌细胞为主(约60%), 72 h时终末分化的肌细胞占比达60%。对MyoD CUT&RUN与SHARE-seq的数据联合分析表明: MyoD通过直接结合骨骼肌分化基因启动子E-box元件, 开放染色质, 从而激活成肌分化相关基因表达; 而MEF细胞相关基因的沉默与其启动子区染色质可及性降低相关, 但MyoD无法直接靶向结合该类启动子, 提示成纤维细胞相关基因表达下调存在E-box非依赖的间接抑制机制。总之, 该研究阐明了MyoD介导MEF转分化经历早期成肌细胞、肌细胞、终末分化肌细胞3个有序阶段, 发现了MyoD可通过不同表观遗传途径激活成肌分化命运或抑制成纤维细胞身份, 揭示了转录因子驱动细胞命运重编程的双重调控机制。

关键词 MyoD; 成纤维细胞; 转分化; 单细胞RNA测序; 染色质可及性

Single-Cell Transcriptomic Analysis Reveals the Temporal Dynamics of MyoD-Mediated Fibroblast-to-Myocyte Transdifferentiation

Alimujiang Tudiyusufu¹, WAN Xin¹, GU Yingying¹, HU Xiaodi¹, ZHU Dahai¹,

QIN Yan², ZHANG Yong^{1*}, ZENG Ping^{2*}

⁽¹⁾Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing 100005, China; ⁽²⁾Health and Medical Department, Peking Union Medical College Hospital, Chinese Academy of Medical Science, Beijing 100730, China)

收稿日期: 2026-05-25 接受日期: 2026-06-22

国家自然科学基金面上项目(批准号: 32570756)和中央高水平医院临床科研业务费资助项目(批准号: 2025-PUMCH-C-043)资助的课题

*通信作者。Tel: 010-65105081, E-mail: yongzhang@ibms.pumc.edu.cn; magsie@126.com

Received: May 25, 2026 Accepted: June 22, 2026

This work was supported by the National Natural Science Foundation of China (Grant No.32570756) and the National High Level Hospital Clinical Research Funding (Grant No.2025-PUMCH-C-043)

*Corresponding authors. Tel: +86-10-65105081, E-mail: yongzhang@ibms.pumc.edu.cn; magsie@126.com

Abstract MyoD is a myogenic lineage-specific transcription factor essential for skeletal muscle cell fate determination. Ectopic expression of MyoD in fibroblasts is sufficient to induce transdifferentiation into myoblasts. However, the temporal regulatory features of MyoD-mediated transdifferentiation, particularly the systematic downregulation of fibroblast-associated genes, remain poorly understood. This bioinformatic reanalysis of publicly available multi-omics datasets aimed to characterize, at single-cell resolution, the temporal dynamics and molecular mechanism underlying MyoD-mediated transdifferentiation of MEFs (mouse embryonic fibroblasts) into skeletal muscle cells. This article analyzed publicly available GEO datasets including scRNA-seq (single-cell RNA-sequencing) data, MyoD CUT&RUN data, and chromatin accessibility (SHARE-seq) data from MyoD-mediated MEF transdifferentiation. Pseudo-time analysis using Monocle2 revealed three sequential stages along the transdifferentiation trajectory: early myoblasts (*Cdh15*⁺/*Mef2a*⁺), myocytes (*Myog*⁺/*Actc1*⁺), and terminally differentiated myocytes (*Myh3*⁺/*Ckm*⁺). At 24 h post-induction, early myoblasts represented approximately 60% of cells, whereas by 72 h, terminally differentiated myocytes accounted for 60% of the population. Integrative analysis of MyoD CUT&RUN and SHARE-seq data showed that MyoD directly binds to E-box elements in promoters of skeletal muscle differentiation genes, leading to chromatin opening and activation of myogenic gene expression. In contrast, silencing of MEF-related genes correlated with reduced chromatin accessibility at their promoter regions but lacked direct MyoD binding, suggesting an indirect, E-box-independent repression mechanism. Collectively, this study delineates an ordered three-stage transdifferentiation process from MEFs to terminally differentiated myocytes and demonstrates that MyoD employs distinct epigenetic pathways to activate myogenic cell fate and repress fibroblast identity. The finding reveals a dual regulatory mechanism underlying transcription factor-driven cell reprogramming.

Keywords MyoD; fibroblast; transdifferentiation; single-cell RNA-seq; chromatin accessibility

细胞命运决定是发育生物学的核心科学问题。直接重编程(direct reprogramming)是指在不经历诱导多能干细胞(induced pluripotent stem cell, iPSC)阶段的情况下,通过异位表达关键转录因子,将一种已分化细胞直接转变为另一种细胞类型的过程^[1]。这一策略为再生医学提供了绕开iPSC诱导的新途径,但其基本原理,包括原始细胞身份的消除机制和细胞命运转变的时序规律,仍有待深入阐明^[2-3]。

MyoD(myogenic differentiation)是决定骨骼肌细胞分化命运的关键转录因子(master transcription factor),属于碱性螺旋-环-螺旋(basic helix-loop-helix, bHLH)蛋白家族。1987年, Davis等^[4]首次发现单一转录因子MyoD的异位表达即可将小鼠胚胎成纤维细胞(mouse embryonic fibroblast, MEF)转分化为骨骼肌细胞,开创了转录因子驱动细胞命运重编程的研究先河。作为先锋转录因子(pioneer transcription factor)^[5], MyoD通过识别增强子盒(enhancer-box, E-box, CANNTG)顺式调控元件,结合相关基因的启动子或增强子,进而招募SWI/SNF

染色质重塑复合物及共激活因子p300/CBP, 开放靶基因启动子的染色质结构, 激活*Myog*、*Myh3*等骨骼肌分化基因的表达^[6-7]。此外, Wang等^[8]进一步发现MyoD作为三维基因组结构组织者(3D genome structure organizer)发挥作用, 通过重塑拓扑关联域(topologically associating domain, TAD)边界和增强子-启动子远距离成环(chromatin looping), 在全基因组尺度建立骨骼肌谱系特异性的染色质空间构象, 揭示了成肌分化相关基因表达的三维基因组水平的调控机制。

近年来, 单细胞多组学技术的快速发展为解析细胞命运转分化时序特征提供了有力的工具^[9]。SHARE-seq技术可在同一细胞中同步检测转录组与染色质可及性^[10], CUT&RUN技术可高分辨率图谱化转录因子结合位点。最近, Hersbach等^[11]建立了Collide-seq技术平台, 利用组合式、可追踪的重编程因子表达体系, 在单细胞水平同时记录转录因子身份与转录组状态, 对MyoD等多种重编程因子介导MEF细胞转分化过程中的不同时间点(24 h、48 h、72 h)进行了scRNA-seq、CUT&RUN和SHARE-seq

综合分析。该研究的核心发现是: 不同重编程因子诱导的成纤维细胞基因沉默缺乏普适性的机制。然而, 该研究侧重多因子碰撞条件下的细胞命运决策, 对MyoD单因子转分化的时序特征及表观遗传调控机制尚未进行深度的单细胞拟时序解析。

因此, 本研究对Hersbach等^[11]公开的数据集进行深度分析, 聚焦于MyoD单因子转分化体系, 整合多时间点scRNA-seq、CUT&RUN和SHARE-seq三套数据, 采用Monocle2^[12]拟时序分析、差异表达基因(differentially expressed genes, DEGs)富集分析、ChIPseeker峰注释及单细胞ATAC-seq技术等联合分析, 在单细胞分辨率下系统解析MyoD介导MEF细胞转分化的时序特征和表观遗传调控机制。

1 数据来源与分析方法

1.1 原始数据来源

本研究所有分析数据均来源于Hersbach等^[11]公开的三套GEO数据集: ①scRNA-seq数据集GSE211826(GSM6504514: 24 h; GSM6504515: 48 h; GSM6504516: 72 h); ②CUT&RUN数据集GSE210181(72 h MyoD结合数据); ③SHARE-seq数据集GSE211863(72 h单细胞转录组与染色质可及性联合检测数据)。参考文献[11]的生物信息分析代码已在GitHub(theislab/collideseq_reproducibility)公开, 本研究在此基础上新增了Monocle2(v2.30)拟时序分析、ChIPseeker(v1.28)峰注释, 并通过自定义R脚本完成多组学整合分析; 相关领域研究者若有合理需求, 可向作者索取本分析流程。

1.2 scRNA-seq数据预处理

从GEO数据库下载已完成比对至mm10参考基因组的h5格式矩阵文件, 按数据来源文章中的方法, 使用Python包(3.10.20) scanpy(v1.12.0)对数据进行预处理。预处理时, 首先过滤、去除高线粒体DNA、表达量低的基因和表达基因数量太少的细胞, 之后利用R包(4.6.0) SoupX(v1.6.2)去除环境RNA。然后将每个细胞依据外源转染的转录因子和外源荧光蛋白的表达情况, 通过最大后验估计验证, 将每个细胞分到相应转录因子表达组。完成上述预处理后, 选取表达MyoD组(788个细胞)和转染空载质粒的对照组(200个细胞)作为新的数据组用于后续分析。

1.3 scRNA-seq无监督分群

首先, 对数据进行标准化(normalize_total, tar-

get_sum=10 000), 筛选高变基因(top 2 000 highly variable genes), 进行 $\log(1+x)$ 对数转换。然后, 利用主成分分析(principal component analysis, PCA)降维($n_{\text{comps}}=50$), 采用Harmony方法($\theta=2$)消除3个时间点之间的批次效应。在校正后的低维嵌入(embedding)基础上构建K近邻算法(K-nearest neighbors, KNN)图($n_{\text{neighbors}}=15$), 利用鲁汶(Louvain)算法(resolution=0.5)进行无监督聚类, 并以均匀流形逼近与投影(uniform manifold approximation and projection, UMAP)进行二维可视化。每个细胞亚群按其特异性标记基因进行注释。

1.4 Monocle2拟时序分析

利用R包(4.6.0) Monocle2(v2.30)进行无监督单细胞轨迹分析^[12]。在进行轨迹分析时, 首先将对照成纤维细胞组与MyoD阳性的细胞组之间的差异表达基因($q<0.01$, 共2 465个基因)作为关键基因, 再根据每个细胞中这些关键基因的表达特征, 利用Monocle2算法[判别式降维树(discriminative dimensionality reduction tree, DDRTree)]计算每个细胞伪时间值并对细胞进行排序, 模拟出细胞随时间变化的发育轨迹。

1.5 差异表达基因分析与GO富集分析

利用Python包(3.10.20) diffxpy(v0.7.4)对细胞亚群进行差异分析。细胞亚群0分别与细胞亚群1、2、3进行差异分析, 分别计算其差异表达基因($q<0.01$, $|\log_2\text{FC}|>0.5$), 之后将三组上调基因和下调基因合并为最终总的上下调基因。对上下调基因分别进行基因本体条目(GO term)富集分析并将结果可视化(gProfiler2, 校正后 $P<0.05$)。

1.6 MyoD CUT&RUN数据分析

从GEO数据库中下载原始FastQ文件, 首先用FastQC(v0.11.9)对测序数据进行质控, 之后利用Trim Galore去掉测序接头(adapter)和低质量序列(quality<20)。用Bowtie2(v2.4.2, --no-unal --no-mixed --no-discordant -X 700)将序列比对到mm10小鼠基因组, 比对后利用SAMtools[比对质量值(mapping quality, MAPQ) ≥ 20]过滤低质量比对, 并用Picard MarkDuplicates去除PCR重复序列。利用MACS2(v2.2.7, $q<0.05$, --nomodel --extsize 200 --gsize mm)以IgG对照为背景分别对2个重复样本进行峰识别(call peak), 2个重复样本所得的峰取交集(bedtools intersect)作为高可信度MyoD结合峰, 用于后续分析。MyoD结

合峰利用IGV(v2.12)进行可视化。利用R包(4.6.0) ChIPseeker(v1.28)对MyoD结合峰进行基因组注释,将转录起始位点(transcription start site, TSS)上下游1 Kb作为启动子区域定义,筛选出启动子区域存在MyoD结合峰的基因,使用clusterProfiler包(v4.14.6)进行GO term富集分析(校正后 $P < 0.05$)。

1.7 SHARE-seq数据预处理

从GEO数据库中下载已完成比对到小鼠基因组的SHARE-seq的h5文件和单细胞ATAC峰注释文件,按数据来源的文章中的方法^[1],使用Python包(3.10.20) scanpy(v1.12.0)对其进行预处理。首先在SHARE-seq数据中将RNA数据与ATAC数据分别提取出来,单独对其进行预处理。对RNA-seq数据进行过滤并去除高线粒体DNA、表达量低的基因和表达基因数量太少的细胞,之后利用R包(4.6.0) SoupX (v1.6.2)去除环境RNA。在ATAC-seq数据中去掉低质量的细胞和峰,根据ATAC峰注释文件对单细胞中每个峰进行注释。然后在RNA-seq数据中,将每个细胞依据外源转染的转录因子和外源荧光蛋白的表达情况,通过最大后验估计验证,将每个细胞分到相应转录因子表达组。完成上述预处理后,在数据中选取转染MyoD组和转染空载质粒的对照组细胞作为新的数据组用于后续分析。

1.8 SHARE-seq数据无监督分群

将SHARE-seq数据按照RNA-seq特征进行分群。再对RNA-seq数据进行标准化、筛选高变基因、 $\log(1+x)$ 对数转换,之后利用Louvain算法和UMAP对数据降维并进行无监督分群。每个细胞亚群按其特异性基因进行注释。将每个细胞根据RNA分群结果直接应用至ATAC数据,计算并可视化相应基因在不同细胞亚群中的表达和染色质开放性信号。

1.9 数据与代码可用性

本研究使用的所有原始数据均可从NCBI GEO数据库中公开获取,编号分别为:scRNA-seq数据GSE211826、MyoD CUT&RUN数据GSE210181、SHARE-seq数据GSE211863。生物信息学分析基于Hersbach等^[1]公开的开源代码框架,使用Monocle2(v2.30)、ChIPseeker(v1.28)等标准生物信息学软件包及自定义脚本完成;所有分析参数与软件版本已在方法学中列出,相关领域研究者若有合理需求,可向作者索取本分析流程。

1.10 统计学分析

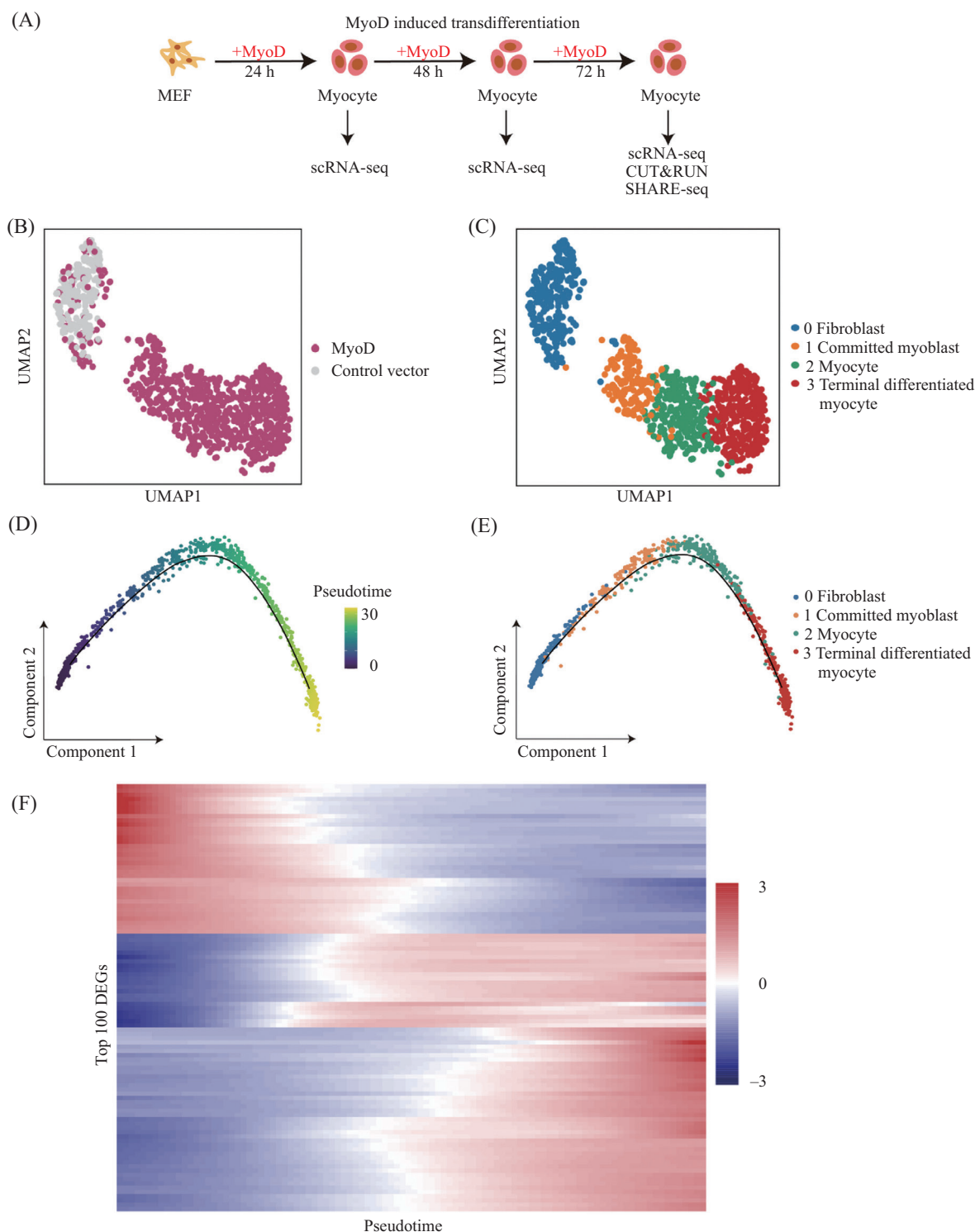
实验数据以平均数 $\pm$ 标准差($\bar{x} \pm s$)表示。采用Shapiro-Wilk检验进行正态性分析,两组间比较采用曼-惠特尼U检验(Mann-Whitney U检验),分类变量比较采用 χ^2 检验或Fisher精确检验,多重检验采用Benjamini-Hochberg方法进行假发现率(false discovery rate, FDR)校正。所有统计学分析使用R包(4.6.0)和Python包(3.10.20)完成, $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 拟时序分析发现MyoD介导MEF细胞转分化经历3个阶段

为了探讨MyoD介导MEF细胞转分化过程中细胞命运转变的时序特征,我们从GEO公共数据库中下载了Hersbach等^[1]发表的多种转录因子介导MEF转分化不同时间点的scRNA-seq数据GSE211826[GSM6504514(24 h)、GSM6504515(48 h)、GSM6504516(72 h)](图1A)。整合3个时间点的数据并进行标准化处理,消除样本批次之间的差异。数据预处理之后选取成功表达MyoD的细胞(Myod阳性)和对照组即感染空载体的MEF细胞(Myod阴性)进行后续的系统性分析。全部单细胞样本中Myod阳性的细胞有788个,对照组细胞有200个(图1B)。经数据降维,无监督分群得到4个细胞亚群(0、1、2、3群,图1C)。其中,0群细胞主要来自对照组,也包含少部分Myod阳性细胞,而1、2、3群全部为Myod阳性细胞(图1C)。0、1、2、3群细胞数量分别为:279、159、240、310。

为了探究细胞命运转变的轨迹,我们利用Monocle2算法根据每个细胞基因表达谱的变化特征进行拟时间轨迹计算。结果显示:所有细胞排列出没有分歧的轨迹,表明MyoD介导转分化过程中细胞命运转变是随时间逐渐完成的过程(图1D)。将拟时间轨迹上的全部细胞按细胞亚群注释,结果显示:0群细胞处于拟时间轨迹初始位置,1、2群依次分布在其后,3群在末端(图1E)。这表明,转分化过程中细胞命运由0群逐渐转变为1群,再到2群,最后转变为3群。进一步分析MyoD阳性细胞与对照组差异最大的100个基因在拟时间轨迹轴上表达的变化,可以看到近1/3的早期高表达基因随着转分化进程逐渐下调,而其余早期低表达基因逐渐上调



A: MyoD介导MEF细胞转分化的数据集来源。B: MyoD阳性细胞和对照组MEF细胞的UMAP图。C: 所有细胞无监督分群后细胞亚群的UMAP图。D: 所有细胞在拟时序轨迹上的分布图, 细胞颜色表示拟时序值。E: 细胞亚群在细胞分化轨迹上的分布图, 对图1D中所有细胞依据其所属亚群进行分色标注。F: 前100个差异表达基因(DEGs)在拟时序的细胞中表达热图, 纵坐标为MyoD阳性细胞与对照组细胞之间差异最大的前100个差异表达基因, 横坐标为每个细胞按拟时序分布。

A: dataset source of MyoD-mediated fibroblast transdifferentiation. B: UMAP plot of MyoD-positive cells and control fibroblasts. C: UMAP plot showing subpopulations by unsupervised clustering of all cells. D: distribution of all cells along the pseudotime trajectory. Cell colors represent pseudotime values. E: distribution of cell subpopulations along the differentiation trajectory. Each cell in fig.1D is colored according to its subcluster identity. F: heatmap of the top 100 DEGs (differentially expressed genes) ordered by pseudotime. The y-axis lists the top 100 genes showing the largest expression differences between MyoD-positive cells and fibroblast control. The x-axis represents cells arranged in increasing pseudotime order.

图1 拟时序分析发现MyoD介导MEF细胞转分化经历3个阶段

Fig.1 MyoD-mediated fibroblast transdifferentiation proceeds through three stages by pseudotime analysis

(图1F)。这些结果支持MyoD介导成纤维细胞命运转变经历3个阶段:亚群1、2、3分别处于细胞命运转变的不同时期,并按0、1、2、3这个顺序完成细胞命运转变。

2.2 利用差异表达基因数据集注释3个阶段的细胞命运

为了探讨转分化过程中不同阶段的细胞亚群的特征,我们通过差异表达基因找到每个亚群的标记基因(marker基因)并根据marker基因注释亚群细胞类型。因为0群细胞大部分为对照成纤维细胞并且在拟时间轨迹上处于最初始位置,将0群细胞分别与1、2、3群细胞做差异分析。以热图展示全部差异表达基因在4个亚群中的表达情况,0群细胞高表达的基因在1、2、3群细胞中持续下调;0群细胞不表达的基因在1、2、3群持续上调(图2A)。GO term富集分析显示:上调基因主要富集在骨骼肌细胞分化等生物学过程中;下调基因主要富集在细胞黏附、细胞外基质的组织等成纤维细胞相关通路中(图2B)。这些结果符合MyoD介导MEF转分化过程中骨骼肌细胞相关基因逐渐上调、成纤维细胞相关基因逐渐下调的特征。

每个细胞亚群的marker基因和它们对应的生物学功能如图2C。0群的marker基因为: *Fbln2*、*Thy1*、*Col8a1*、*Serpine1*, 这些基因是成纤维细胞marker基因, GO terms富集在细胞外基质、细胞黏附等成纤维细胞相关通路中。1群的marker基因为: *Cdh15*、*Des*、*Itga7*、*Mef2a*, 富集在骨骼肌细胞早期分化相关通路中。2群的marker基因为: *Myog*、*Tnni1*、*Mef2c*、*Actc1*, 富集在骨骼肌细胞分化等肌细胞相关通路中。3群的marker基因为: *Tnnc2*、*Myh3*、*Myh7*、*Ckm*, 富集在肌肉收缩、肌肉器官发育等骨骼肌细胞分化晚期相关通路中。

根据每个亚群的marker基因来注释细胞(图2D): 0群为成纤维细胞, 这些细胞包括对照组细胞和少量命运没有转分化的细胞; 1群为早期命运决定的成肌细胞; 2群为肌细胞; 3群为终末分化的肌细胞。UMAP图显示, *Thy1*、*Fbln2*主要在成纤维细胞中表达, 在早期命运决定的成肌细胞中出现下调趋势(图2E)。 *Cdh15*、*Mef2a*、*Myog*、*Actc1*、*Myh3*、*Ttn*等与骨骼肌细胞谱系相关的基因, 依次在成肌细胞、分化肌细胞及终末成熟肌细胞阶段启动表达, 且伴随细胞命运定向分化进程, 基因表达水平逐步上调(图2F~图2H)。

细胞亚群注释结果和前面的拟时间轨迹结果表明, MyoD介导MEF转分化过程中的成纤维细胞先后经历早期成肌细胞、肌细胞、终末分化的肌细胞等阶段。

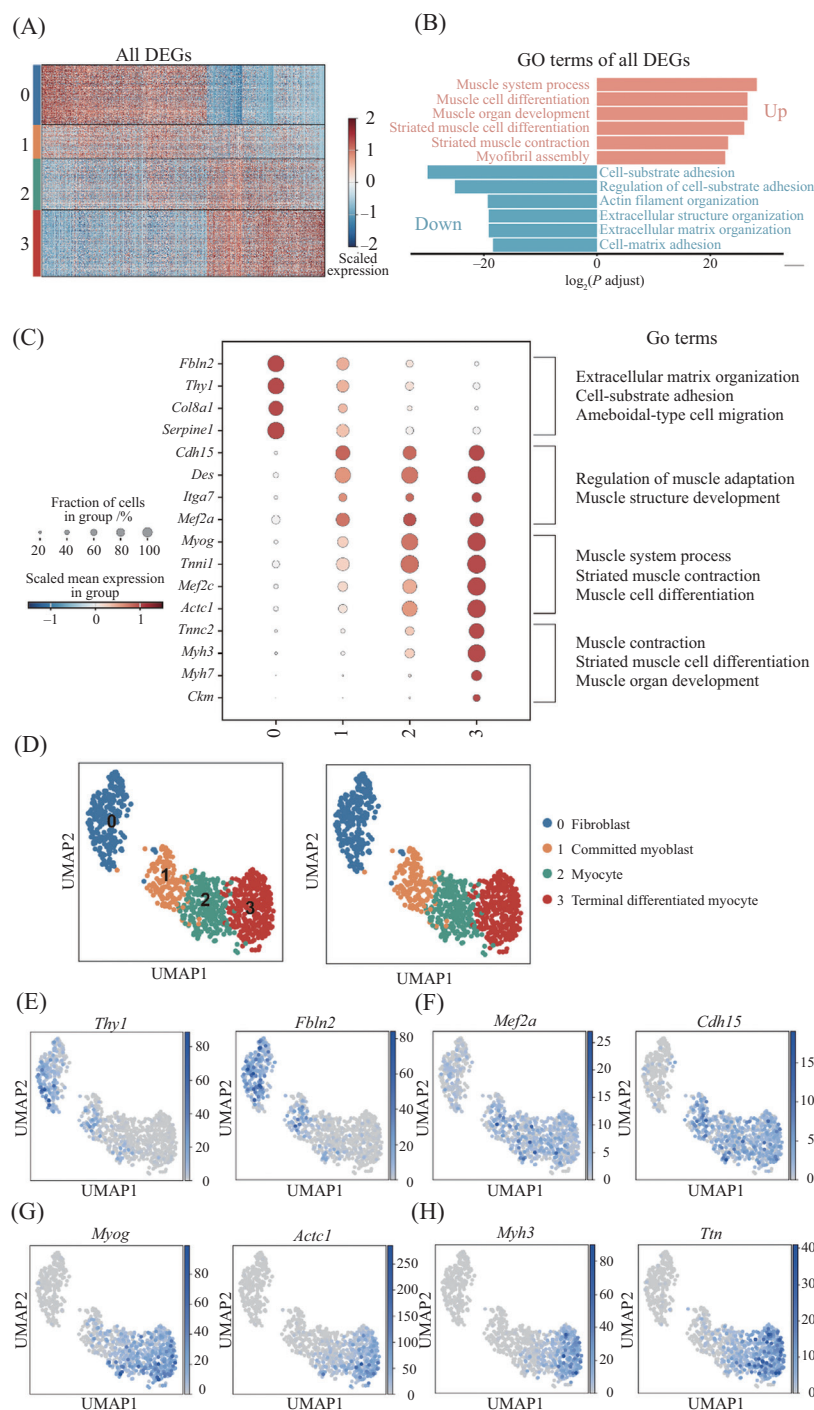
2.3 时间序列的数据分析揭示了3个肌细胞亚群的时序变化特征

为了进一步研究细胞亚群时序性特点, 我们分析了24 h、48 h、72 h等3个肌肉细胞亚群的变化情况(图3A)。24 h时除了对照成纤维细胞之外, 只有早期的成肌细胞和肌细胞2个亚群。48 h时早期成肌细胞、肌细胞、终末分化的肌细胞等3个亚群都存在并以后2个亚群为主。72 h时同样3个亚群细胞都有, 但大部分细胞在终末分化肌细胞亚群中。这结果与前面拟合时间轨迹结果一致。统计所有MyoD阳性细胞中不同亚群的比例和该比例在不同时间点变化情况(图3B): 24 h、48 h、72 h均有约10% MyoD阳性细胞仍没有进行转分化, 处于成纤维细胞状态。24 h时约60%细胞处于早期成肌细胞状态, 72 h时大部分细胞处于终末分化肌细胞状态, 约占60%。UMAP展示 *Fbln2*、*Cdh15*、*Myog*、*Myh3*等基因在24 h、48 h、72 h每个细胞中的表达情况(图3C~图3F)。这些结果表明: 在MyoD介导转分化过程中, 从早期成肌细胞逐渐变成肌细胞, 最后变成终末分化的肌细胞, 与前面拟时间轨迹计算结果一致。

2.4 MyoD结合成肌分化相关基因启动子影响染色质开放性介导MEF细胞转分化为成肌细胞命运

为了探讨MyoD介导MEF细胞命运转分化过程中MyoD结合基因组DNA与单细胞层面的基因表达和染色质开放性变化的关联, 我们进一步分析了Hersbach等^[11]发表的MEF细胞转分化72 h的MyoD CUT&RUN数据(GSE210181)和SHARE-seq数据(GSE211863)。

首先, 分析转分化过程中MyoD结合峰与成肌分化相关基因表达的关系。转分化72 h MyoD CUT&RUN的2个重复样本分别获得7 279个和10 674个MyoD结合峰, 将重复样本的结合峰取并集, 注释MyoD结合峰直接调控的基因。结果表明, 72 h MyoD结合峰中约11%的峰分布在基因的近端启动子区(proximal promoter), 大部分峰分布在增强子区(包括基因内含子或者远端基因间区域)(图4A)。对启动子区TSS上下游1 Kb存在MyoD结合峰

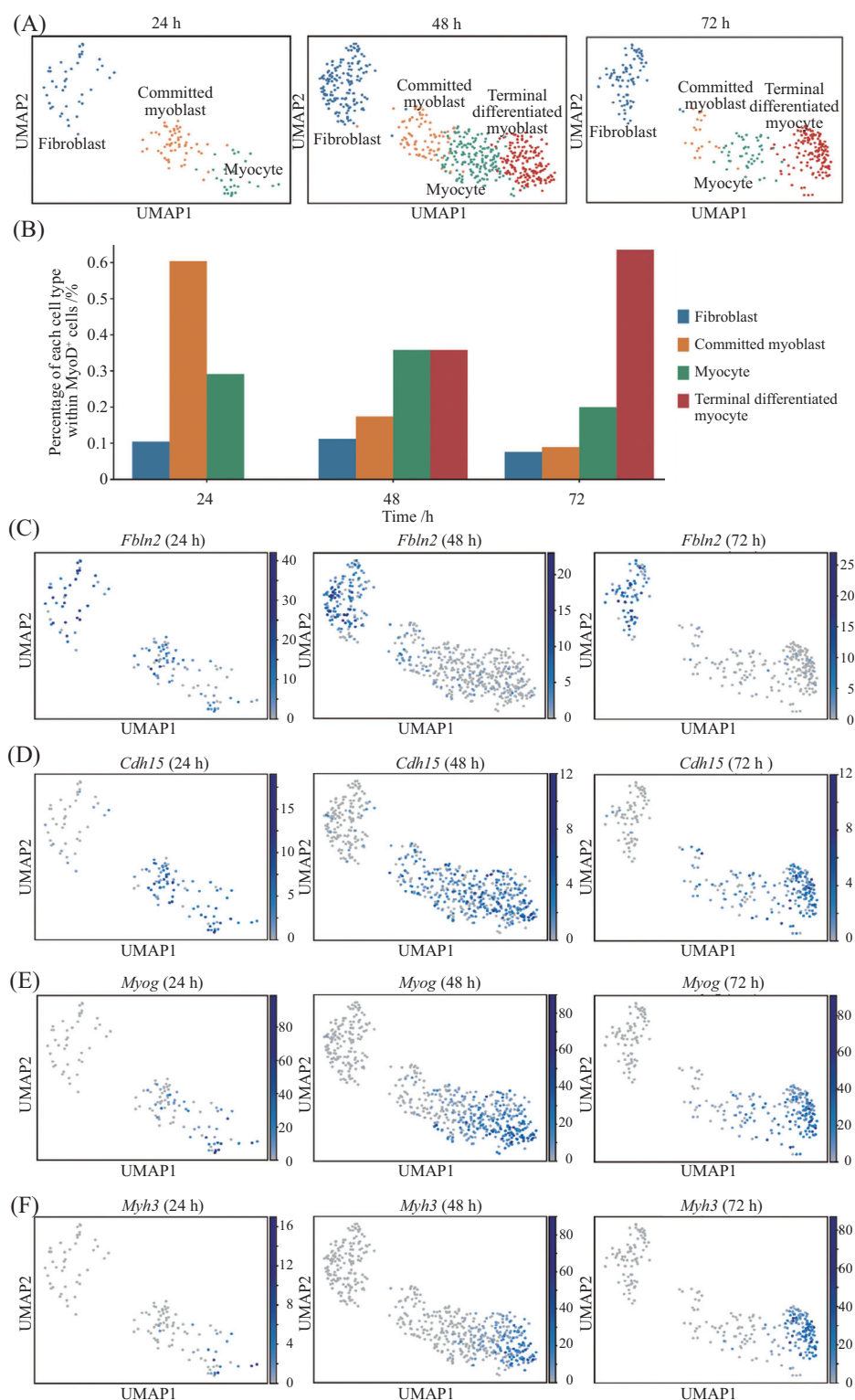


A: 4个细胞亚群的差异表达基因(DEGs)热图。差异表达基因为0群分别与1、2、3群比较所得差异表达基因的并集。DESeq2, $q < 0.01$, $|\log_2 FC| > 0.5$ 。B: 差异表达基因GO terms富集分析。gProfiler2, $q < 0.05$ 。C: 4个细胞亚群marker基因表达量气泡图及对应GO terms。D: 细胞亚群注释前后的UMAP对比图。左: Louvain无注释分群结果。右: 生物学注释。E~H: UMAP图展示成纤维细胞标记基因*Thy1*和*Fbln2*(E), 早期成肌细胞标记基因*Mef2a*和*Cdh15*(F), 肌细胞标记基因*Myog*和*Actc1*(G), 终末分化肌细胞标记基因*Myh3*和*Tnni*(H)的表达谱。

A: heatmap showing DEGs (differentially expressed genes) among four cell subpopulations. The set of genes includes the union of DEGs identified by comparing cluster 0 versus clusters 1, 2, and 3 (DESeq2, $q < 0.01$, $|\log_2 FC| > 0.5$). B: GO term enrichment analysis of the DEGs shown in fig.2A, performed using gProfiler2 ($q < 0.05$). C: bubble plot depicting the expression levels of marker genes for each of the four cell subpopulations, together with their corresponding enriched GO terms. D: UMAP plots comparing cell subclusters before and after biological annotation. Left: Louvain clustering (unannotated clusters). Right: UMAP with biologically annotated cell identities. E-H: UMAP plots showing the expression levels of fibroblast marker genes *Thy1* and *Fbln2* (E), early myoblast marker genes *Mef2a* and *Cdh15* (F), myocyte marker genes *Myog* and *Actc1* (G), and terminally differentiated myocyte marker genes *Myh3* and *Tnni* (H).

图2 利用各亚群标志基因注释3个阶段的细胞命运

Fig.2 Cell fate annotation of three stages based on marker genes

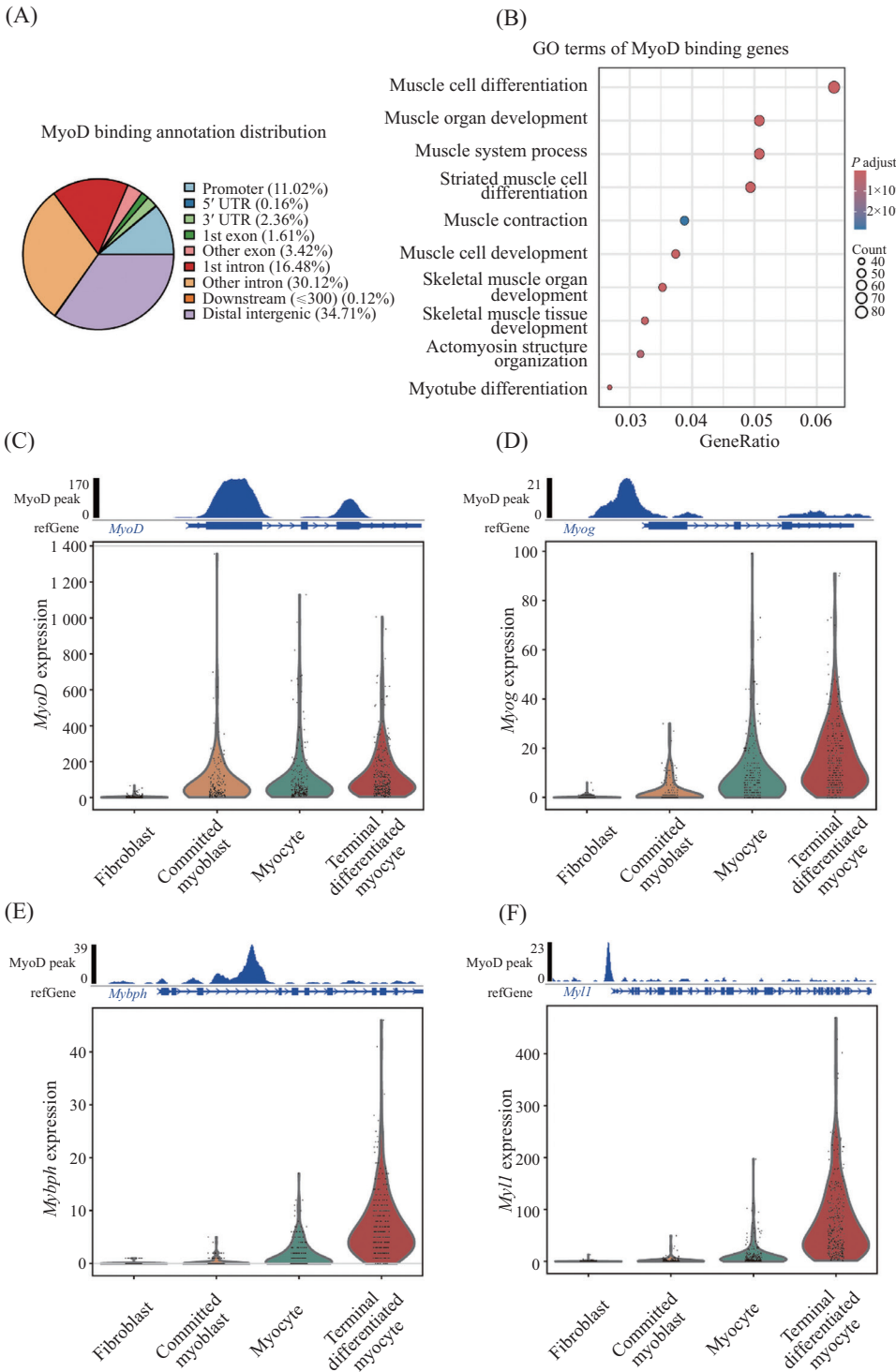


A: 24 h、48 h、72 h等3个时间点各细胞亚群分布的UMAP图(788个MyoD阳性细胞)。B: MyoD阳性细胞中各亚群比例随时间的变化的柱状图。C~F: UMAP图展示成纤维细胞标记基因*Fbln2*(C)、早期成肌细胞标记基因*Cdh15*(D)、肌细胞标记基因*Myog*(E)、终末分化肌细胞标记基因*Myh3*(F)在24 h、48 h、72 h各时间点的表达谱。

A: UMAP plots showing the distribution of 4 cell subclusters at 24 h, 48 h, and 72 h during MyoD-mediated transdifferentiation ($n=788$ MyoD-positive cells). B: Bar chart showing the proportional changes of the 4 cell subcluster over time. C-F: UMAP plots displaying the expression profiles of marker genes at 24 h, 48 h, and 72 h during MyoD-mediated transdifferentiation. Fibroblast marker gene *Fbln2* in fig.3C. Early myoblast marker gene *Cdh15* in fig.3D. Myocyte marker gene *Myog* in fig.3E. Terminally differentiated myocyte marker gene *Myh3* in fig.3F.

图3 拟时序分析揭示了3个肌细胞亚群的时序变化特征

Fig.3 Pseudotime analysis reveal temporal patterns of three myocyte subpopulations during transdifferentiation



A: 饼图展示MyoD结合峰在全基因组范围内的分布。B: 气泡图展示MyoD结合位点对应的基因的GO term富集情况图。选择在基因启动子区域有MyoD结合峰的全部基因进行GO term富集分析。C~F: 在成肌分化相关基因(*MyoD*、*Myog*、*Mybph*、*Myl1*)启动子区的MyoD结合峰图(上, IGV可视化结合峰图)与其在4个细胞亚群中的表达水平(下, 小提琴图)。其中, 图4C是*MyoD*基因, 图4D是*Myog*基因, 图4E是*Mybph*基因, 图4F是*Myl1*基因。

A: pie chart showing the genome-wide distribution of MyoD binding peaks. B: bubble plot displaying GO term enrichment analysis of genes with MyoD binding peaks in their promoter regions. All genes showing MyoD binding at promoter regions were subjected to GO enrichment analysis. C-F: MyoD binding peaks at the promoter regions of myogenic differentiation-related genes (upper panels: IGV visualization of peak profiles), and their expression levels across four cell subclusters (lower panels, violin plots) in fig.4C for *MyoD* gene, fig.4D for *Myog* gene, fig.4E for *Mybph* gene, and fig.4F for *Myl1* gene.

图4 MyoD结合成肌分化相关基因的启动子调控MEF细胞转分化

Fig.4 MyoD regulates MEF cell transdifferentiation by binding to promoters of myogenic differentiation-related genes

的基因进行GO term富集分析,结果显示MyoD结合的基因富集在骨骼肌细胞分化、骨骼肌发育等成肌分化相关的通路中(图4B)。进一步分析MyoD结合与成肌分化相关基因表达及细胞命运的关联,如图4C~图4F,我们分别展示了*MyoD*、*Myog*、*Mybph*、*Myhl*等成肌分化相关基因TSS附近MyoD的结合情况和这些基因在scRNA-seq不同细胞亚群中的表达情况,发现MyoD结合在这些基因的启动子附近并逐渐上调其表达,随着细胞转分化进程相关基因表达丰度逐渐增加。以上结果表明,MyoD通过直接结合成肌分化相关基因启动子激活基因表达实现MEF细胞转分化为骨骼肌细胞。

然后,系统分析MyoD介导MEF细胞转分化过程中染色质开放性的变化特征及其与基因表达的关联。MyoD介导MEF细胞转分化72 h的SHARE-seq数据,在单细胞水平同时记录了染色质开放性和基因表达谱。将SHARE-seq数据进行预处理后,选取所有表达MyoD的细胞和对照组成的纤维细胞进行后续分析。根据基因表达特征对SHARE-seq数据进行降维与无监督聚类,得到0、1、2、3这4个细胞亚群(图5A),各细胞亚群分别表达一组相关的marker基因(图5B)。基于每个亚群的marker基因进行亚群细胞类型注释,与前面的scRNA-seq结果一样,0、1、2、3这4个细胞亚群分别注释为成纤维细胞、早期成肌细胞、肌细胞和终末分化的肌细胞(图5C)。进一步分析染色质开放性与基因表达的关联,发现成肌分化相关marker基因(如*Myog*、*Des*、*Myh3*、*Myh7*等)表达量的变化与这些基因启动子区域染色质开放性的变化是一致的(图5D~图5G)。该结果表明,MyoD通过改变染色质开放性上调骨骼肌细胞分化相关基因表达介导MEF细胞转分化为成肌细胞。

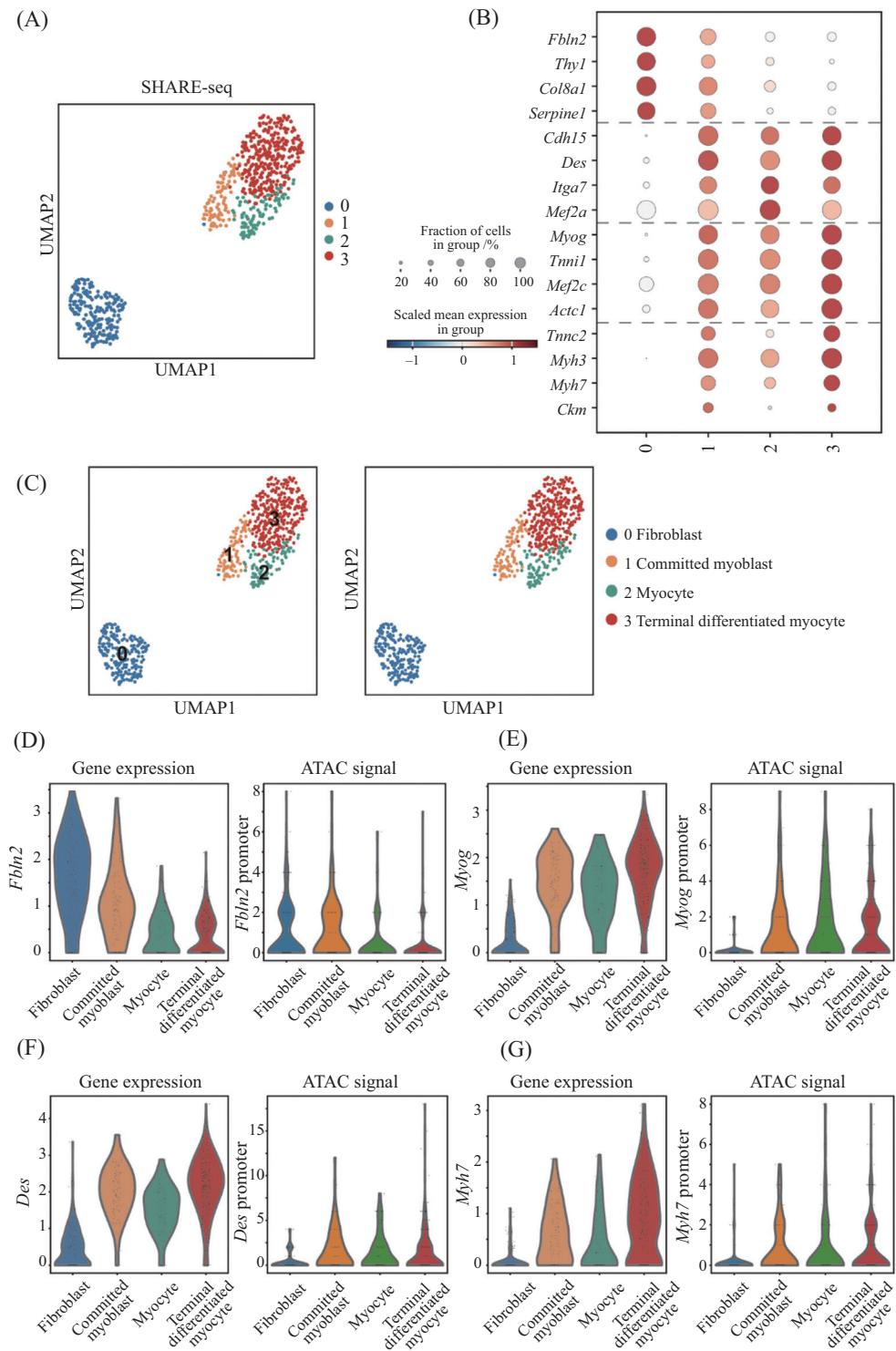
2.5 MyoD没有直接结合成纤维细胞相关基因启动子可能通过间接改变染色质开放性抑制其表达

为了研究MyoD在转分化过程中是否直接结合成纤维细胞基因的启动子来抑制其表达,我们系统分析了72 h MyoD CUT&RUN数据中成纤维细胞相关基因启动子区域的MyoD结合情况。首先分析在转分化72 h下调基因TSS位点上下游是否有MyoD结合?结果显示,这些下调基因TSS附近没有MyoD结合峰(图6A)。然后,我们选择在转分化过程中被抑制的成纤维细胞基因再计算这些成纤维细胞基因

TSS位点上下游是否有MyoD结合?结果显示,在这些被抑制的成纤维细胞基因附近同样没有MyoD结合(图6B)。比如,*Thyl*、*Fbln2*、*Colla1*等成纤维细胞相关的基因在转分化过程中表达下调,但MyoD并没有结合在这些基因的启动子TSS附近(图6C~图6E)。这些结果表明,在转分化过程中成纤维细胞相关基因的沉默不是由MyoD直接结合其启动子调控的。

为了在全基因组尺度上对上述发现进行定量验证,我们对191个上调基因和267个下调基因(差异表达基因筛选阈值为 $q < 0.01$, $|\log_2FC| > 0.5$)的启动子区域(TSS ± 1 Kb)的MyoD CUT&RUN信号强度进行了系统比较(图6F)。结果显示,成肌分化上调基因启动子的MyoD结合信号为 2.04 ± 2.14 ,而成纤维细胞下调基因启动子的MyoD结合信号为 0.92 ± 0.85 ,二者差异显著(曼-惠特尼U检验, $U = 16\ 005$, $Z = -6.80$, $P = 1.06 \times 10^{-11}$, 效应量 $r = 0.37$, 图6F)。进一步分析SHARE-seq ATAC-seq的峰重叠率发现,虽然成纤维细胞基因启动子的染色质开放比例(95.9%)略高于成肌分化基因启动子(91.1%, $\chi^2 = 4.43$, $P = 0.035$, 图6G),但成纤维细胞基因启动子区域的MyoD直接结合信号强度显著更低。该结果提示,成纤维细胞基因启动子区域存在“开放但缺乏MyoD直接结合”的特征,其表达抑制并非通过MyoD启动子直接占位实现。按全体458个基因的MyoD结合信号计算四分位边界后,Q1~Q4整体分布在两类基因之间存在显著差异(2 $\times$ 4列, χ^2 检验, $P = 7.77 \times 10^{-8}$)。对Q1~Q4分别进行Fisher精确检验并经FDR校正后,Q1[FDR = 1.58×10^{-4}]和Q4(FDR = 8.59×10^{-7})差异显著;其中,高结合信号区间(Q4)中肌源性上调基因占比为37.7%,成纤维细胞下调基因占比为16.1%,显示MyoD结合信号向肌源性上调基因显著富集(图6H)。这些定量证据为图6A~图6E所示的定性热图结果提供了基于统计检验的全基因组尺度验证,支持了MyoD对成纤维细胞基因的抑制经由E-box非依赖的间接机制实现的结论。

最后,我们系统分析了MyoD介导MEF转分化过程中染色质开放性的变化特征与基因表达变化之间的关系,以热图展示了72 h SHARE-seq中差异表达基因的表达谱和这些基因区域染色质开放性变化(图7A和图7B)。结果显示,基因表达的上下调与染色质开放或关闭基本一致,说明MyoD介导转

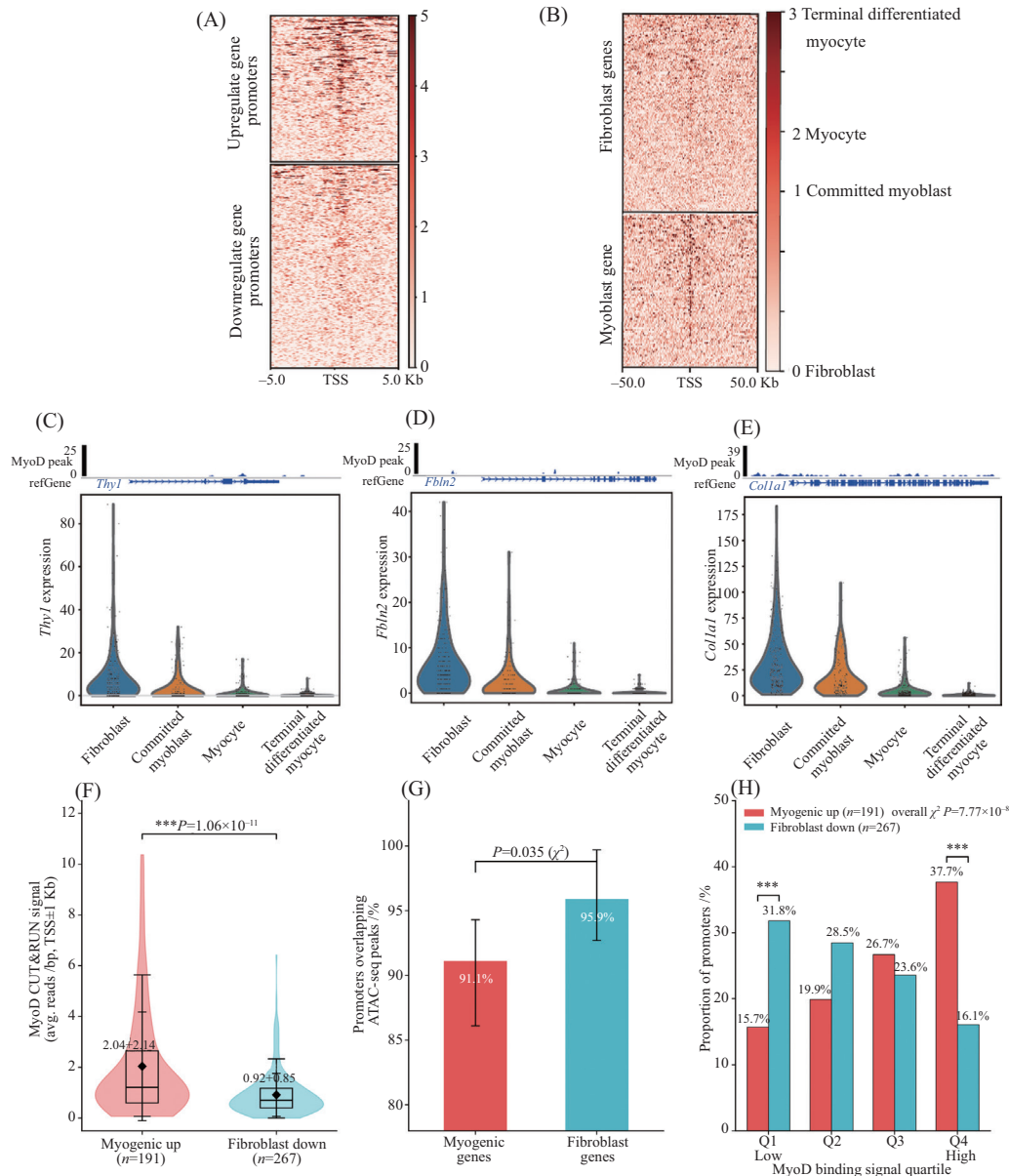


A: UMAP图展示基于基因表达谱的SHARE-seq细胞分群。所有细胞按其基因表达进行UMAP降维并进行无监督分群。B: 气泡图展示4个细胞亚群marker基因的表达量。C: 基于每个亚群的marker基因进行亚群细胞类型注释。左: Louvain无注释分群结果。右: 生物学注释分群。D~G: 小提琴图展示在4个亚群细胞中相关基因表达水平及其启动子区染色质开放性的ATAC信号。图5D为*Fbln2*基因, 图5E为*Myog*基因, 图5F为*Des*基因, 图5G为*Myh7*基因。

A: UMAP plot showing cell subclusters based on gene expression profiles from SHARE-seq. All cells were subjected to UMAP dimensionality reduction and unsupervised clustering according to their gene expression patterns. B: bubble plot displaying the expression levels of marker genes for the four cell subclusters. C: annotation of cell subcluster identities based on marker genes for each subcluster. Left: Louvain clustering (unannotated clusters). Right: UMAP with biologically annotated cell identities. D-G: Violin plots showing gene expression levels and promoter chromatin accessibility (ATAC signals) for the indicated genes across the four subclusters in fig.5D (*Fbln2*), fig.5E (*Myog*), fig.5F (*Des*), and fig.5G (*Myh7*).

图5 MyoD调节成肌分化相关基因的染色质开放状态

Fig.5 MyoD regulates the chromatin accessibility of myogenic differentiation-related genes

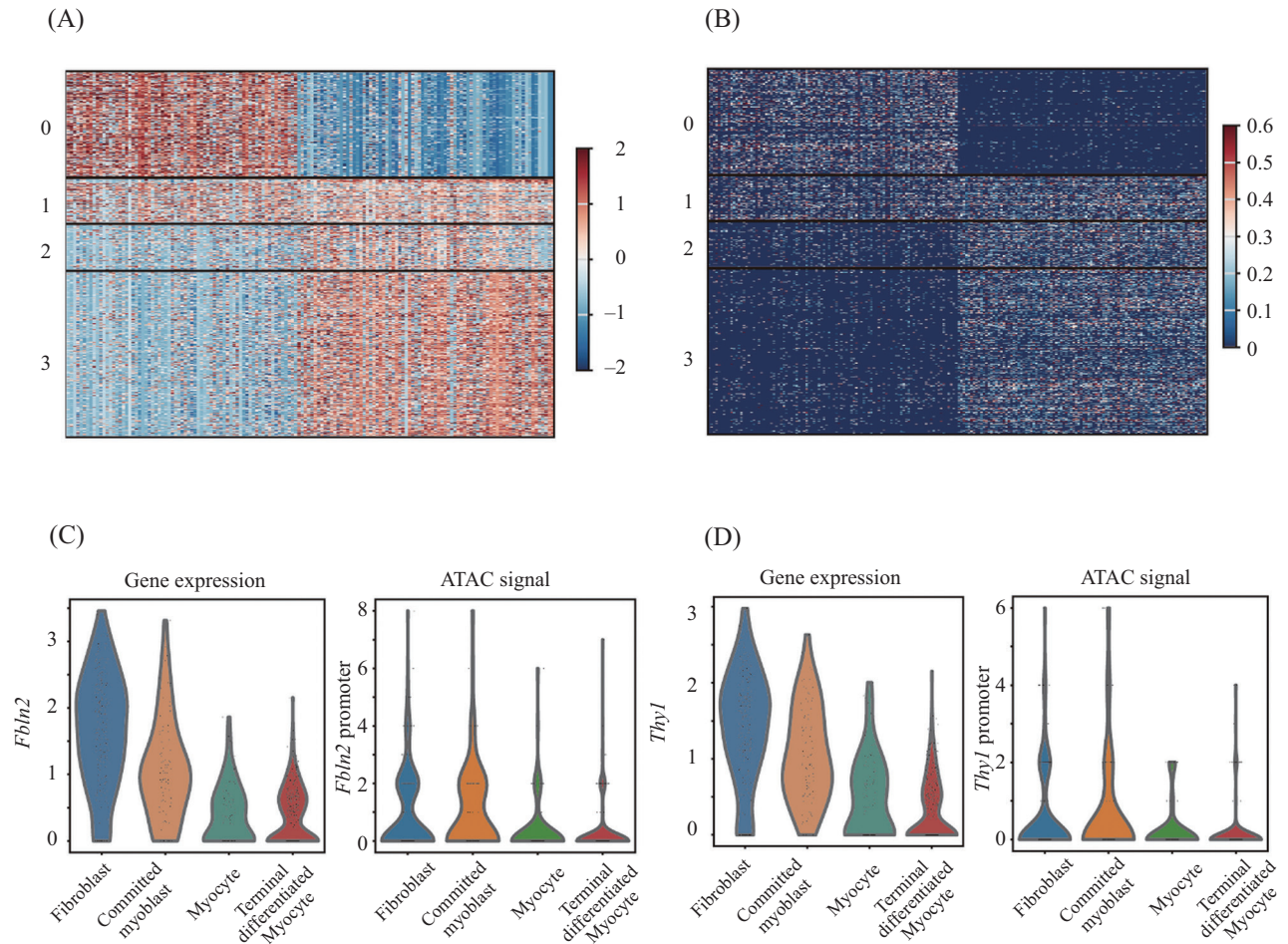


A: 热图展示scRNA-seq上调和下调基因启动子区域MyoD结合峰。上调和下调基因指的是: 在MyoD介导转分化过程中, 与0群细胞(成纤维细胞)相比, 所有上调的和下调的基因。B: 热图展示scRNA-seq成肌分化相关基因和成纤维细胞相关基因启动子区域MyoD结合峰。C~E: 在*Thyl*位点(C)、*Fbln2*位点(D)、*Colla1*位点(E)的MyoD结合峰(上图)与这些基因在4个细胞亚群中的表达水平(下图)。F: 小提琴-箱线图展示191个上调基因(红色)与267个下调基因(蓝色)差异表达基因筛选阈值: $q<0.01$, $|\log_2FC|>0.5$ 启动子区(TSS±1 Kb) MyoD CUT&RUN信号分布, 黑色菱形和误差线表示 $\bar{x} \pm s$ 。曼-惠特尼U检验, $P=1.06 \times 10^{-11}$, 效应量 $r=0.37$ 。G: SHARE-seq和ATAC-seq峰与两类基因启动子区的重叠率(%)。95%置信区间, χ^2 检验, $P=0.035$ 。H: MyoD结合信号四分位区间(Q1~Q4)中成肌分化相关基因与成纤维细胞基因启动子的占比分布。整体分布采用 2×4 , χ^2 检验; Q1~Q4分别采用Fisher精确检验并进行FDR校正, 图中标注FDR校正后显著的Q1和Q4。 $***P<0.001$ 。

A: heatmap showing MyoD binding peaks at promoter regions of upregulated and downregulated genes. Upregulated and downregulated genes were defined as those differentially expressed during MyoD-mediated transdifferentiation compared with cluster 0 (fibroblasts). B: heatmap showing MyoD binding peaks at promoter regions of myogenic differentiation-related genes and fibroblast-related genes. C-E: MyoD binding peaks at the *Thyl* (C), *Fbln2* (D), and *Colla1* (E) loci (upper panels) and the expression levels of these genes across four cell subclusters (lower panels). F: violin-boxplot showing the distribution of MyoD CUT&RUN signals at promoter regions (TSS±1 Kb) of 191 upregulated genes (red) and 267 downregulated genes (blue) identified using $q<0.01$ and $|\log_2FC|>0.5$. Black diamonds and error bars indicate $\bar{x} \pm s$. Mann-Whitney *U* test: $P=1.06 \times 10^{-11}$, effect size $r=0.37$. G: overlap percentage (%) of SHARE-seq ATAC-seq peaks with promoter regions of the myogenic and fibroblast genes. Error bars represent 95% confidence intervals, χ^2 test: $P=0.035$. H: proportion distribution of myogenic differentiation-related genes and fibroblast genes within quartiles (Q1-Q4) of MyoD binding signal intensity at promoters. The overall distribution was tested by a 2×4 , χ^2 test; each quartile was further tested using Fisher's exact test with FDR correction, and Q1 and Q4 remained significant after correction. $***P<0.001$.

图6 MyoD不是通过直接结合启动子的方式抑制成纤维细胞基因

Fig.6 MyoD represses fibroblast gene expression but does not directly bind to their promoters



A: 热图展示SHARE-seq结果中0、1、2、3四群细胞的差异表达基因。差异表达基因为0群细胞分别与1、2、3群细胞做差异分析得到的全部差异表达基因的总和。B: 热图展示SHARE-seq结果中0、1、2、3四群细胞的差异表达基因对应位点染色质开放性变化。C: 小提琴图展示*Fbln2*基因表达(左)与*Fbln2*基因启动子区的染色质开放性ATAC信号(右)。D: 小提琴图展示*Thy1*基因表达(左)与*Thy1*基因启动子区的染色质开放性ATAC信号(右)。

A: heatmap showing differentially expressed genes among clusters 0, 1, 2, and 3 from SHARE-seq data. The gene set includes the union of all DEGs identified by comparing cluster 0 with clusters 1, 2, and 3, respectively. B: heatmap showing changes in chromatin accessibility at loci corresponding to the differentially expressed genes shown in fig.7A from SHARE-seq data. C: violin plots showing *Fbln2* gene expression (left) and chromatin accessibility (ATAC signal) at the *Fbln2* gene promoter (right). D: violin plots showing *Thy1* gene expression (left) and chromatin accessibility (ATAC signal) at the *Thy1* gene promoter (right).

图7 MyoD通过降低染色质开放性抑制成纤维细胞基因表达

Fig.7 MyoD represses fibroblast gene expression by reducing chromatin accessibility

分化过程中基因表达的变化是染色质开放性的变化引起的, 基因表达下调是染色质开放性的关闭引起的。比如, 成纤维细胞相关基因*Fbln2*、*Thy1*等在4个细胞亚群中逐渐下调, 这些基因启动子区域的染色质开放性也逐渐关闭(图7和图7D)。总之, 这些结果表明, MyoD介导MEF转分化过程中染色质开放性的降低可能是抑制成纤维细胞相关基因表达的原因。

3 讨论

MyoD介导细胞命运转分化是重编程领域研究

最为经典的体系之一。自Davis等^[4]于1987年发现单一转录因子MyoD即可将成纤维细胞直接转化为骨骼肌细胞以来, 研究证实MyoD作为先锋转录因子(pioneer transcription factor)^[5], 可在多种非骨骼肌细胞(包括肝细胞、脂肪细胞和皮肤成纤维细胞等)中启动骨骼肌分化程序^[6-7], 成为转录因子驱动细胞命运重编程的范式性例证。然而, 该领域仍面临几个核心问题: 其一, MyoD介导转分化的时序调控机制及各阶段细胞亚群的分子标识尚不清晰; 其二, 转分化效率不完全——各研究中均有相当比例细胞抵抗命运转换并维持成纤维细胞状态, 其分子机制

尚不明确^[9,13-14]；其三，MyoD如何同时激活骨骼肌分化基因并抑制成纤维细胞特征基因的表达，特别是成纤维细胞基因的沉默是否依赖MyoD直接结合启动子，尚缺乏系统性单细胞水平分析^[15-16]。单细胞多组学技术的兴起为解决上述问题提供了有力工具^[9,12]，使在单细胞分辨率下同时追踪转录组变化与表观遗传状态成为可能。本研究对Collide-seq公开数据集^[11]进行深度分析，整合多时间点scRNA-seq、MyoD CUT&RUN和SHARE-seq三套数据，在单细胞分辨率下系统解析了MyoD介导MEF转分化的时序特征和表观遗传调控机制，核心发现包括：转分化遵循线性有序轨迹，历经早期成肌细胞、肌细胞和终末分化的肌细胞三阶段；MyoD通过直接结合骨骼肌分化基因启动子E-box元件激活肌源性转录；成纤维细胞基因的抑制则经由缺乏MyoD直接启动子结合的间接表观遗传机制实现。

第一，拟时序分析发现MyoD介导转分化过程中细胞命运转变是随时间逐渐完成的渐进性过程。全部细胞排列出没有分歧的线性轨迹，细胞亚群0、1、2、3按顺序处于细胞命运转变的不同时期，表明转分化是有序推进而非随机进行的。这与此前在心肌细胞直接重编程中观察到的线性轨迹一致^[17]，而与部分含分叉节点的非线性重编程轨迹不同^[18]，提示小鼠MEF系统中MyoD单因子条件下的命运转变具有高度确定性。24 h→48 h→72 h各亚群比例的平稳过渡(24 h约60%早期成肌细胞→72 h约60%终末分化肌细胞)进一步证实了这一渐进性特征。

第二，对3个转分化阶段的细胞亚群注释揭示了成纤维细胞向骨骼肌细胞命运转变的分子标识。0群(成纤维细胞)、1群(早期成肌细胞)、2群(肌细胞)、3群(终末分化肌细胞)的marker基因和GO term富集结果与已知骨骼肌发育生物学高度吻合，显示生肌调节因子(myogenic regulatory factors)驱动的分化程序被MyoD有序激活。值得注意的是，约10%的MyoD阳性细胞在3个时间点均维持成纤维细胞状态，与关于转分化不完全性与染色质重塑缺陷相关的报道一致^[13-14,19]，提示起始细胞的表观遗传状态是决定转分化效率的重要因素。

第三，CUT&RUN分析证实MyoD通过直接结合骨骼肌分化基因启动子E-box元件激活肌源性基因转录。约11%的MyoD结合峰定位于启动子区域，这些启动子对应基因的表达量随转分化进程单调递

增，支持MyoD作为先驱转录因子开放染色质的经典模型^[6-7,20]。约89%的MyoD结合峰位于远端内含子或基因间区域，与Wang等^[8]发现MyoD作为三维基因组结构组织者通过远端增强子-启动子空间互作调控基因表达的报道一致，提示MyoD的靶基因调控范围远超启动子直接结合所能覆盖的范畴。这一以远端结合为主的结合模式同时提示，MyoD对成纤维细胞基因的间接沉默可能通过远端调控元件活性变化或染色质三维结构重塑来实现。

最具价值的发现是成纤维细胞基因沉默的调控机制。*Thy1*、*Fbln2*、*Colla1*等成纤维细胞特征基因的启动子区域(TSS±1 Kb)缺乏MyoD直接结合，但其染色质可及性随转分化进程同步降低，基因表达同步下调，表明MyoD不通过经典的E-box依赖启动子结合方式抑制成纤维细胞基因。这一发现与原始Collide-seq研究在转录组层面观察到的“成纤维细胞特异性基因表达下调缺乏统一机制”在表观遗传层面形成互补^[11]。值得注意的是，Nicoletti等^[15-16]最新研究发现MyoD可通过E-box非依赖的方式结合并去活化成纤维细胞基因的远端超级增强子(super-enhancer)，进而引起启动子染色质关闭和基因抑制。这与本研究观察到的“成纤维细胞基因启动子染色质可及性降低但无MyoD启动子直接结合”高度吻合，提示Nicoletti机制可能是本体系统中成纤维细胞基因间接沉默的分子基础。本研究在单细胞ATAC-seq层面为上述机制提供了细胞群体层面的相关性证据。此外，Wang等^[8]前期发现MyoD可作为三维基因组结构组织者，与CTCF等结构蛋白互作，在成肌细胞增殖分化过程中重构肌肉谱系特异性染色质三维结构。基于这一非经典功能，在MEF转分化过程中MyoD可能同样通过染色质三维结构重塑，改变成纤维细胞基因相关远端调控元件与启动子的空间互作关系，从而间接导致这些基因启动子区域染色质可及性降低并下调基因表达。这一假说将MyoD经典的直接启动子结合功能与三维基因组组织者的非经典功能统一起来，为理解转分化过程中双向基因调控(激活肌源性基因、抑制成纤维细胞基因)提供了整合性框架，值得在未来研究中通过Hi-C等手段直接验证。

本研究存在以下局限性，需在解读结论时予以考量。首先，研究完全基于公共数据集再分析，所有表观遗传调控结论均为相关性证据，尚缺乏体内外

基因敲除、突变及成纤维细胞基因远端增强子删除等功能验证, 因此相关结论仍属于生信关联分析层面的机制提示, 因果关系仍需进一步实验证实。其次, SHARE-seq和CUT&RUN数据仅覆盖72 h单一时间点, 而scRNA-seq轨迹分析跨越24~72 h, 因此以72 h截面表观遗传数据推断全时程的染色质动态变化存在时间错位的固有局限, 未来需收集多时间点CUT&RUN和SHARE-seq数据以验证各拟时序阶段的表观遗传特征。再次, 成纤维细胞基因沉默的间接机制有待进一步精细解析: 本研究仅检查了启动子近端(TSS $\pm$ 1 Kb)的MyoD结合, MyoD是否通过结合成纤维细胞基因的远端调控元件(超级增强子等)发挥沉默功能尚需系统分析。此外, 染色质三维结构重塑(Hi-C或ChIA-PET)及相关修饰(H3K27me3、H3K9me2等)等在间接抑制中的具体作用亦有待实验验证。最后, 本研究总细胞数(MyoD组788个, 对照组200个)相对有限, 部分亚群(1群159个细胞)差异表达基因的统计功效有待扩大样本量予以确认。

从应用角度看, 本研究对MyoD介导MEF转分化阶段性细胞状态及其表观遗传特征的解析, 可为体外骨骼肌细胞分化、骨骼肌损伤修复策略优化以及体细胞重编程方案设计提供参考。尤其是早期成肌细胞、肌细胞和终末分化肌细胞3个阶段的标志基因及染色质开放性变化, 有助于后续研究选择合适的干预时间窗和评价指标; 但这些潜在应用仍需在外体诱导体系、动物损伤模型及功能实验中进一步验证。

总之, 本研究对Collide-seq公开数据集进行了深度再分析, 在单细胞分辨率下系统解析了MyoD介导MEF转分化的时序特征与表观遗传调控机制。Monocle2拟时序分析揭示, MEF转分化按线性、无分叉轨迹经历早期成肌细胞、肌细胞、终末分化的肌细胞3个有序阶段, 24 h以早期成肌细胞为主(约60%), 72 h终末分化的肌细胞占比达约60%。MyoD通过直接结合骨骼肌分化基因启动子E-box元件开放染色质并激活转录, 基因表达量与启动子染色质可及性在单细胞层面高度耦合。整合CUT&RUN与SHARE-seq数据发现, 成纤维细胞基因的沉默与其启动子染色质可及性降低相关, 但该类基因的启动子区域缺乏MyoD直接结合的位点, 表明MyoD通过不同表观遗传途径分别驱动肌源性基因激活与成纤维细胞基因沉默, 为既往Collide-seq研究中“成纤维

细胞基因下调缺乏统一机制”的实验观察提供了染色质层面的机制阐释。

参考文献 (References)

- [1] Vierbuchen T, Wernig M. Direct lineage conversions: unnatural but useful? [J]. *Nature Biotechnology*, 2011, 29(10): 892-907.
- [2] Srivastava D, DeWitt N. *In vivo* cellular reprogramming: the next generation [J]. *Cell*, 2016, 166(6): 1386-1396.
- [3] Keshri R, Detraux D, Phal A, et al. Next-generation direct reprogramming [J]. *Frontiers in Cell and Developmental Biology*, 2024, 12: 1343106.
- [4] Davis R L, Weintraub H, Lassar A B. Expression of a single transfected cDNA converts fibroblasts to myoblasts [J]. *Cell*, 1987, 51(6): 987-1000.
- [5] Zaret K S, Carroll J S. Pioneer transcription factors: establishing competence for gene expression [J]. *Genes & Development*, 2011, 25(21): 2227-2241.
- [6] Tapscott S J. The circuitry of a master switch: MyoD and the regulation of skeletal muscle gene transcription [J]. *Development*, 2005, 132(12): 2685-2695.
- [7] Battistelli C, Garbo S, Maione R. MyoD-induced trans-differentiation: a paradigm for dissecting the molecular mechanisms of cell commitment, differentiation and reprogramming [J]. *Cells*, 2022, 11(21): 3435.
- [8] Wang R Y, Chen F L, Chen Q, et al. MyoD is a 3D genome structure organizer for muscle cell identity [J]. *Nature Communications*, 2022, 13(1): 205.
- [9] Biddy B A, Kong W, Kamimoto K, et al. Single-cell mapping of lineage and identity in direct reprogramming [J]. *Nature*, 2018, 564(7735): 219-224.
- [10] Ma S, Zhang B, LaFave L M, et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin [J]. *Cell*, 2020, 183(4): 1103-1116.
- [11] Hersbach B A, Fischer D S, Masserdotti G, et al. Probing cell identity hierarchies by fate titration and collision during direct reprogramming [J]. *Molecular Systems Biology*, 2022, 18(9): e11129.
- [12] Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells [J]. *Nature Biotechnology*, 2014, 32(4): 381-386.
- [13] Liu W J, Saelens W, Rainer P, et al. Dissecting the impact of transcription factor dose on cell reprogramming heterogeneity using scTF-seq [J]. *Nature Genetics*, 2025, 57(10): 2451-2464.
- [14] Francesconi M, Di Stefano B, Berenguer C, et al. Single cell RNA-seq identifies the origins of heterogeneity in efficient cell transdifferentiation and reprogramming [J]. *eLife*, 2019, 8: e41627.
- [15] Nicoletti C, Massenet J, Pintado-Urbanc A P, et al. E-box independent chromatin recruitment turns MYOD into a transcriptional repressor [J]. *Genes & Development*, 2025, 39(21/22): 1279-1298.
- [16] Nicoletti C, Massenet J, Pintado-Urbanc A P, et al. MYOD represses gene expression from non-E-box motifs [J]. *Genes & Development*, 2025, 39(21/22): 1299-1317.
- [17] Liu Z Q, Wang L, Welch J D, et al. Single-cell transcriptomics

- reconstructs fate conversion from fibroblast to cardiomyocyte [J]. *Nature*, 2017, 551(7678): 100-104.
- [18] Cacchiarelli D, Qiu X J, Srivatsan S, et al. Aligning single-cell developmental and reprogramming trajectories identifies molecular determinants of myogenic reprogramming outcome [J]. *Cell Systems*, 2018, 7(3): 258-268.
- [19] Manandhar D, Song L Y, Kabadi A, et al. Incomplete MyoD-induced transdifferentiation is associated with chromatin remodeling deficiencies [J]. *Nucleic Acids Research*, 2017, 45(20): 11684-11698.
- [20] Cao Y, Kumar R M, Penn B H, et al. Global and gene-specific analyses show distinct roles for MyoD and MyoE at a common set of promoters [J]. *The EMBO Journal*, 2006, 25(3): 502-511.