

A comprehensive transcriptomic analysis of differentiating embryonic stem cells in response to the overexpression of Mesogenin 1

Dahai Liu^{1*}, Qiang Zhang^{1*}, Hong Zhang^{*1}, Ling Tang¹, Wei Li¹, Dongming Zhang¹, Guoying Wu¹, Shoudong Ye^{1,2}, Qian Ban^{1,2}, Kan He^{1,2}

¹Center for Stem Cell and Translational Medicine, School of Life Sciences, Anhui University, Hefei City, Anhui 230601, P. R. China

²Department of Biostatistics, School of Life Sciences, Anhui University, Hefei City, Anhui 230601, P. R. China

*Equally contributed

Correspondence to: Kan He; email: hekan_803@163.com

Keywords: somitogenesis, microarray, pathway, embryonic stem cells, Msgn1

Received: January 18, 2016

Accepted: September 22, 2016

Published: October 6, 2016

ABSTRACT

The mutation of somitogenesis protein Mesogenin 1 (Msgn1) has been widely used to study the direct link between somitogenesis and the development of an embryo. Several studies have used gene expression profiling of somitogenesis to identify the key genes in the process, but few have focused on the pathways involved and the coexpression patterns of associated pathways. Here we employed time-course microarray datasets of differentiating embryonic stem cells by overexpressing the transcription factor Msgn1 from the public database library of Gene Expression Omnibus (GEO). Then we applied gene set enrichment analysis (GSEA) to the datasets and performed candidate transcription factors selection. As a result, several significantly regulated pathways and transcription factors (TFs), as well as some of the specific signaling pathways, were identified during somitogenesis under Msgn1 overexpression, most of which had not been reported previously. Finally, significant core genes such as Hes1 and Notch1 as well as some of the TFs such as PPARs and FOXs were identified to construct coexpression networks of related pathways, the expression patterns of which had been validated by our following quantitative real-time PCR (qRT-PCR). The results of our study may help us better understand the molecular mechanisms of somitogenesis in mice at the genome-wide level.

INTRODUCTION

In vertebrates, somites give rise to the axial skeleton, skeletal muscle, and dorsal dermis, which arose from the presomitic mesoderm (PSM), a mesoderm-derived tissue lying on both sides of the neural tube [1, 2]. The PSM is dynamic: somites are generated at the anterior end, each posterior PSM cell moves from its original position toward anterior PSM and is finally integrated into the somites [3, 4]. The clock and wavefront model has been established to make the clock mechanism that somitogenesis relies on well known. Molecular evidence for the existence of this clock has been obtained on the basis of the periodic expression of several genes related to Notch signaling in the PSM.

The segmentation clock is considered to play an important role in the periodic activation of Notch signaling in the PSM [5].

In vertebrate embryos, somite segmentation is controlled by the molecular clock mentioned above, in the form of a transcriptional oscillator that operates in the PSM. Mesogenin 1 (*Msgn1*), a basic-Helix-Loop-Helix (bHLH) transcription factor that lay in the PSM [6, 7], is essential for differentiation, movement and maturation of PSM progenitor cells during somitogenesis [8]. *Msgn1* is expressed in the mesodermal compartment of the primitive streak and is necessary for the proper development of the mesoderm. *Msgn1* drives the changes in gene expression in the nascent PSM cells

and regulates the movements from the posterior region to the PSM. Besides, *Msgn1* controls the size of somites and manipulates the size and existence of the progenitor cells. *Msgn1* acts as a direct target gene of the Wnt/ β -catenin signaling pathway, which can activate Notch signaling genes to control the segmentation clock. *Msgn1* can mediate crosstalk between the Wnt and Notch signaling pathways during mammalian somitogenesis [9]. Thus, *Msgn1* is a pivotal ingredient of a transcriptional cascade and plays a fundamental role in the segmentation clock. *Msgn1* in the differentiating embryonic stem cells can overexpress in the presence of Doxycycline(Dox) which is a member of the tetracycline antibiotics group. Thus, we used Dox to verify the function of *Msgn1*.

Msgn1 can activate the expression of numerous genes including cyclic genes in Notch signaling pathway via Wnt signaling program, which belongs to a deeply conserved cell signaling system present in most multicellular organisms [10]. Notch signaling cascade is composed of a conserved family of extracellular ligands and two cellular factors that may associate with the Notch intracellular domain. Binding of ligand proteins to the extracellular domain can induce proteolytic cleavage and release of the intracellular domain, which activates some target genes, such as *Lfng* and *Hes7* [11]. It has been known that Notch signaling is of capital importance in embryogenesis, central nervous system development and function, cardiovascular development and endocrine development. In this project, we focus on its central role to somitogenesis. In 1995, Conlon RA et al found that Notch1 would play important roles in the coordination of the somites segmentation in mice [12]. Further studies have identified the role of Notch signaling in the segmentation clock mentioned above.

These studies supposed that the major function of Notch signaling does not act upon a single cell, but coordinates cell clocks and keep them timed. This hypothesis illustrated the function of Notch signaling during segmentation and has been supported by experiments in mice and zebrafish [13, 14]. In vertebrates, Notch signaling is in the direct regulation of segmentation. Tight control of Notch signaling was crucial for the proper implementation of somitogenesis [15]. Notch signaling keeps the oscillations of neighbouring presomitic mesoderm cells synchronized in somite segmentation [16].

Although the pivotal function of *Msgn1* in the segmentation clock has been revealed in previous studies, the multiple related pathways and associated transcription factors (TFs) were little known, especially in genome-wide. Here we employed time-course microarray datasets of differentiating embryonic stem cells by overexpressing the transcription factor *Msgn1* from the public database library of Gene Expression Omnibus (GEO). Then we applied gene set enrichment analysis (GSEA) on the datasets and performed candidate transcription factors selection to explore the molecular mechanisms of somitogenesis in mice at the genome-wide level.

RESULTS AND DISCUSSION

Overview of the significantly regulated pathways and TFs during somitogenesis under *Msgn1* overexpression

According to our GSEA of the data sets of 18 samples used in the microarray experiment by the comparison of *Msgn1* overexpression to wild-type mice, there were

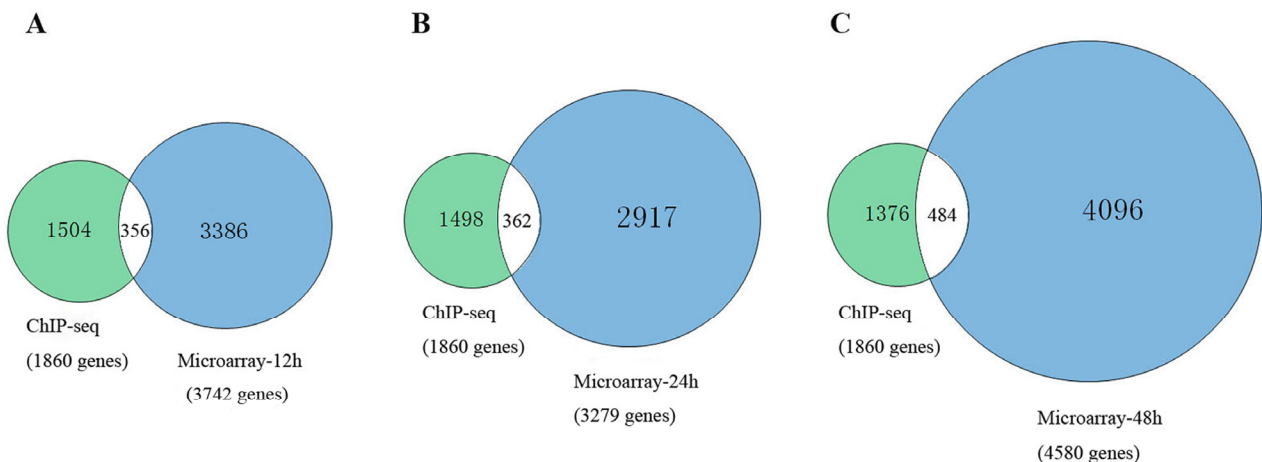


Figure 1. The comparison between related ChIP-seq results with our GSEA target genes.

100 significantly associated pathways identified with $p < 0.01$ at 12h, including 83 upregulated and 17 down-regulated pathways. Moreover, there were 113 significant pathways including 29 upregulated and 84 downregulated ones at 24h, as well as 183 significant pathways including 50 upregulated and 113 down-regulated ones at 48h. The details of each pathway are given in Table S-1. In order to predict TFs potentially involved in embryonic stem cell differentiation in the case of *Msgn1* overexpression, we implemented the analysis of TFBSs and the prediction of TFs using the significant genes in each identified pathway. Based on the cutoff value of TF importance, we have identified the associated TFs with potential target genes that are co-regulated in each of the above-mentioned 376 pathways. The details are shown in Table S-2.

To further explain the regulatory mechanism of *Msgn1* functions in the process of embryonic stem cells development, we have further combined the related ChIP-seq results with our GSEA target genes in different developmental stages, which were shown in Figure 1. As a result, there were 356 common genes identified at 12h, 362 common genes at 24h and 484 common genes at 48h. The details of common genes are shown in Table S-3.

Specific signaling pathways during somitogenesis under *Msgn1* overexpression

Based on the GSEA approach, we have identified several significantly related pathways during somito-

genesis in mice. By the comparison of each group, we have further identified 39 overlapping pathways significantly regulated in these three periods, including 8 signaling pathways (Figure 2). Most of these signaling pathways are related to immune system (3/8) and signal transduction (4/8) according to KEGG pathway maps in the KEGG database (www.genome.jp/kegg/). We next screened and reorganized the entire signaling pathways at three time points, the details of which were showed in Table S-4 and Table S-5. Fc epsilon RI signaling pathway is only significantly regulated in 12h, which belongs to the immune system. There are 2 specific signaling pathways significantly regulated in 24h, PPAR signaling pathway and Insulin signaling pathway, and both of them are concerned with the endocrine system. Similarly, there are 2 signaling pathways significantly regulated in 48h: Jak-STAT signaling pathway which is correlated with signal transduction and B cell receptor signaling pathway which is correlated with the immune system. Moreover, 2 signaling pathways play roles at both 12h and 24h, which are related to the nervous system and signal transduction respectively. In addition, 3 signaling pathways are regulated significantly at 12h and 48h. Among them, VEGF signaling pathway is related to the signal transduction, while Toll-like receptor signaling pathway and NOD-like receptor signaling pathway are belong to the immune system. In 24h and 48h overlapping signaling pathways, 3 signaling pathways functioned in the signal transduction while the other 2 signaling pathways act in the endocrine system.

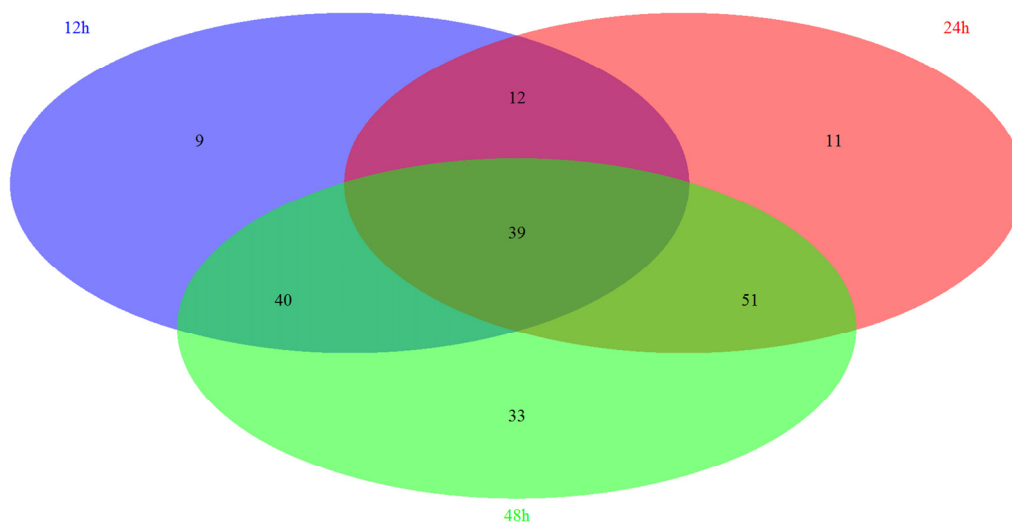


Figure 2. The summary of significantly regulated pathways based on GSEA under *Msgn1* overexpression. The Venn diagram showed the comparisons of each time point. There are 100 most significant pathways at 12 hour, 113 significant pathways at 24 hour and 163 significant pathways at 48 hour. By the comparison of 3 different time points, 39 significantly associated pathways were overlapped. By the comparison of 12h and 24h time points, 51 significant pathways were overlapped. Similarly, 79 and 90 overlapping significant pathways were identified respectively while comparing between 12h and 48h, between 24h and 48h.

Table 1. The signaling pathways significantly down or up-regulated in 12h, 24h and 48h during somitogenesis

Signaling pathway	Functional class	12h	24h	48h
p53 signaling pathway	Cell growth and death	↑	↑	↓
GnRH signaling pathway	Endocrine system		↓	↓
Adipocytokine signaling pathway	Endocrine system		↓	↓
PPAR signaling pathway	Endocrine system		↑	
Insulin signaling pathway	Endocrine system		↓	
Chemokine signaling pathway	Immune system	↑	↓	↓
RIG-I-like receptor signaling pathway	Immune system	↑	↑	↓
T cell receptor signaling pathway	Immune system	↑	↓	↓
Toll-like receptor signaling pathway	Immune system	↑		↓
NOD-like receptor signaling pathway	Immune system	↑		↑
Fc epsilon RI signaling pathway	Immune system	↑		
B cell receptor signaling pathway	Immune system			↓
Neurotrophin signaling pathway	Nervous system	↑	↓	
VEGF signaling pathway	Signal transduction	↑		↓
Calcium signaling pathway	Signal transduction	↓	↓	↓
Notch signaling pathway	Signal transduction	↓	↓	↑
Hedgehog signaling pathway	Signal transduction	↓	↓	↓
ErbB signaling pathway	Signal transduction	↑	↓	
MAPK signaling pathway	Signal transduction	↑	↓	↓
mTOR signaling pathway	Signal transduction		↓	↓
Wnt signaling pathway	Signal transduction		↓	↓
TGF-beta signaling pathway	Signal transduction		↓	↓
Jak-STAT signaling pathway	Signal transduction			↓

Here, all of the signaling pathways at 12h, 24h and 48h during somitogenesis were shown. We compared the signaling pathways with KEGG pathway maps, and obtained the functional classification of these signaling pathways presented in second column. Finally, we performed the down or up-regulated information of each signaling pathways at each time points.

According to Table 1, most signaling pathways are concerned with signal transduction (10/23) such as MAPK signaling pathway, Hedgehog signaling pathway and mTOR signaling pathway. Mitogen-activated protein kinases (MAPKs) signal is one of the important ways in the eukaryotic cell, which adjusts and controls the structure and function of the cell. MAPKs play a

vital role in the production of matrix metalloproteinases and the regulation of cartilage cell proliferation, apoptosis and differentiation [17]. The Hedgehog pathway is a key signaling pathway controlling key steps of embryonic development and patterning through the early stages of development as well as in the development of most tissues and organs in mammals



Figure 3. The coexpression networks of related pathways during somitogenesis. We analyzed the signaling pathways network of dynamic regulation during somitogenesis at three different time points including 12h, 24h and 48h. Fc epsilon RI signaling pathway was only significantly regulated in 12h, PPAR signaling pathway and Insulin signaling pathway were significantly regulated in 24h, Jak-STAT signaling pathway and B cell receptor signaling pathway were significantly regulated in 48h. 2 signaling pathways played roles at both 12h and 24h. Erb signaling pathway can activate Neurotrophin signaling pathway via MAPK signaling pathway. Jun, regulated by MAPK signaling pathway, induce neurotrophin activation. 5 signaling pathways regulated significantly at 24h and 48h, TGF- β signaling pathway activated Wnt signaling pathway via the expression of TGF- β 1. Wnt1 is target gene of Wnt signaling pathway and Wnt1 can induced the activation of mTOR signaling pathway. 3 signaling pathways regulated significantly at 12h and 48h. 9 signaling pathways regulated at three different time points. MAPK and Hedgehog signaling pathways can activate Notch signaling pathway by activating Akt2 and Gli2 respectively. Notch1 and Hes1 are the downstream effectors of Notch signaling pathway and Hes1 is the target gene of Notch1. Hes7 can inhibit Notch target genes, so we hypothesize that Hes1 may have the similar effect. For simplicity, several target genes, gene products, and regulatory interactions are not shown.

[18, 19]. MTOR is a core protein in early embryo development, which highlights the dynamic role of TOR signaling and presents additional functions beyond cell growth control during embryonic development [20]. In addition, some other signaling pathways like RIG-I-like receptor signaling pathway and Fc epsilon RI signaling pathway regulate in the immune system (7/23). The RIG-I-like receptor signaling pathway can help create the ability to recognize virus infection and to enhance a powerful antiviral response, which plays an important role in immune systems. RIG-I-like receptor signaling pathway can activate the production of antiviral cytokines and the establishment of an effective cellular antiviral state by triggering signal transduction pathways, which preserves neighboring cells against infection and initiates native and adaptive immune systems [21].

Notch signaling pathway is down-regulated in 12h and 24h, while it is upregulated in 48h. Wnt signaling pathway is down-regulated in 24h and 48h (shown in Table 1). It has been reported that chondrogenesis could be inhibited by Wnt/ β -catenin signaling pathway and Wnt/ β -catenin signaling pathway would be down-regulated during chondrogenesis under the induction of

TGF- β [22]. We presume that Notch signaling pathway may play a vital role in the later stage of somitogenesis. While in the middle and later periods, Wnt signaling pathway may reduce its effects on somitogenesis.

Network construction and dynamic regulation during somitogenesis

The signaling pathways network was reconstructed to reveal the dynamic regulation during somitogenesis at three different time points including 12h, 24h and 48h (shown in Figure 3). MAPK and Hedgehog signaling pathways can activate Notch signaling pathway by activating *Akt2* and *Gli2* respectively [23, 24]. Inhibition of the MAPK signaling pathway decreases the expression of *Notch1* and *Hes1*. *Notch1* and *Hes1* are downstream effectors of Notch signaling pathway and *Hes1* is the target gene of *Notch1* [25, 26]. *Hes7*, a segmentation clock, can inhibit Notch target genes, so we hypothesize that *Hes1* may have the similar effect. Besides, the relationship of genes and pathways specially involved in 24h and 48h was showed in Figure 2. TGF- β signaling pathway activated Wnt signaling pathway via the expression of *TGF- β 1* [22, 27]. *Wnt1* is

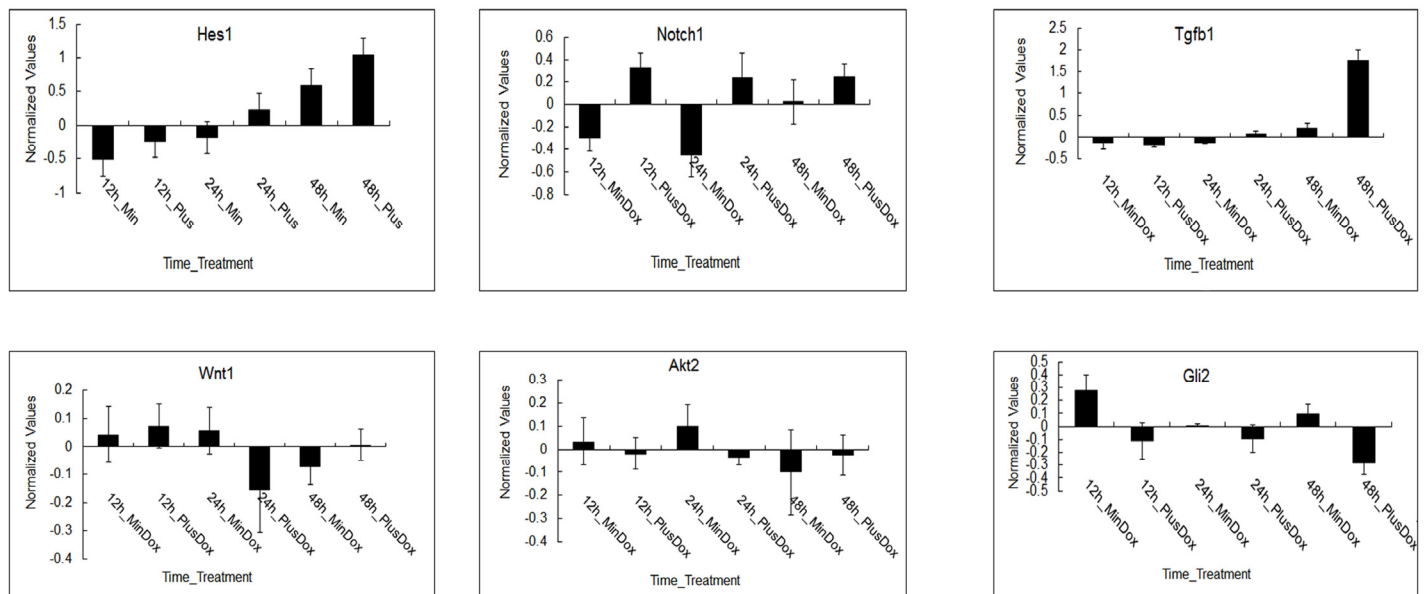


Figure 4. The expression patterns of regulated genes in major signaling pathways during somitogenesis under *Msgn1* overexpression. Microarray analysis of gene expression in the absence and presence of Dox over a 48 hour timecourse. log₂ ratios of the normalized expression levels of *Hes1*, *Notch1*, *Tgfb1*, *Wnt1*, *Akt2* and *Dli2* are presented. The expression of *Hes1* and *Tgfb1* was upregulated under *Msgn1* overexpression over a 48 hour timecourse. *Notch1* expression was variable, presumably reflecting the dynamic expression of a cyclic gene, but was generally elevated by *Msgn1* overexpression. *Tgfb1* expression was upregulated in treated groups compared with untreated ones at 24h, while *Wnt1* expression was just opposite. It may confirm that TGF- β signaling pathway could activate Wnt signaling pathway via the expression of *Tgfb1*. *Notch1* and *Hes1* expressions were both upregulated in the treated groups compared with untreated ones from 12h to 48h. The expression of *Akt2* and *Dli2* were both downregulated compared with control group while *Notch1* expression was upregulated in treated groups, presumably reflecting other signaling pathways, Wnt signaling pathway for example, activate Notch signaling pathway. Here error bars indicate standard deviation of 3 biological replicates.

target gene of Wnt signaling pathway and *Wnt1* can induce the activation of mTOR signaling pathway [28]. Members of the Transforming Growth Factor-beta (TGF β) superfamily of cytokines play crucial roles in pluripotency and differentiation of embryonic stem cells in vitro, which are essential for early mammalian embryonic development [29]. mTOR signaling pathway can influence muscle development in mice by majoring the accumulation of protein level in somites [30].

Erb signaling pathway can activate Neurotrophin signaling pathway via MAPK signaling pathway. The ErbB signaling pathway regulates proliferation, differentiation, cell motility, and survival. MAPK pathway is a downstream target of ErbB receptors. The highly conserved module of MAPK cascade participates in a variety of cellular functions, including cell proliferation, differentiation and migration. Neuro-

trophins belong to the category of trophic factors, which are involved in the differentiation, growth and survival of cells [31]. Several intracellular signaling pathways including the MAPK cascade would be stimulated by the activation of multiple ligands such as the epidermal growth factor (EGF) on ErbB receptor, which is regulated by Src [32]. From Jun Yamauchi et al.'s study, *Jun*, regulated by MAPK signaling pathway, induces neurotrophin activation [33]. All the genes mentioned above can be found in Tables 4, and they confirm our analysis once again.

Log₂ ratios of the normalized expression levels of *Hes1*, *Notch1*, *Tgfb1*, *Wnt1*, *Akt2* and *Dli2* are presented (shown in Figure 4). Error bars demonstrate standard deviation of 3 biological replicates. The expression of *Hes1* and *Tgfb1* was upregulated under *Msgn1* over-expression over a 48-hour time course. The expression

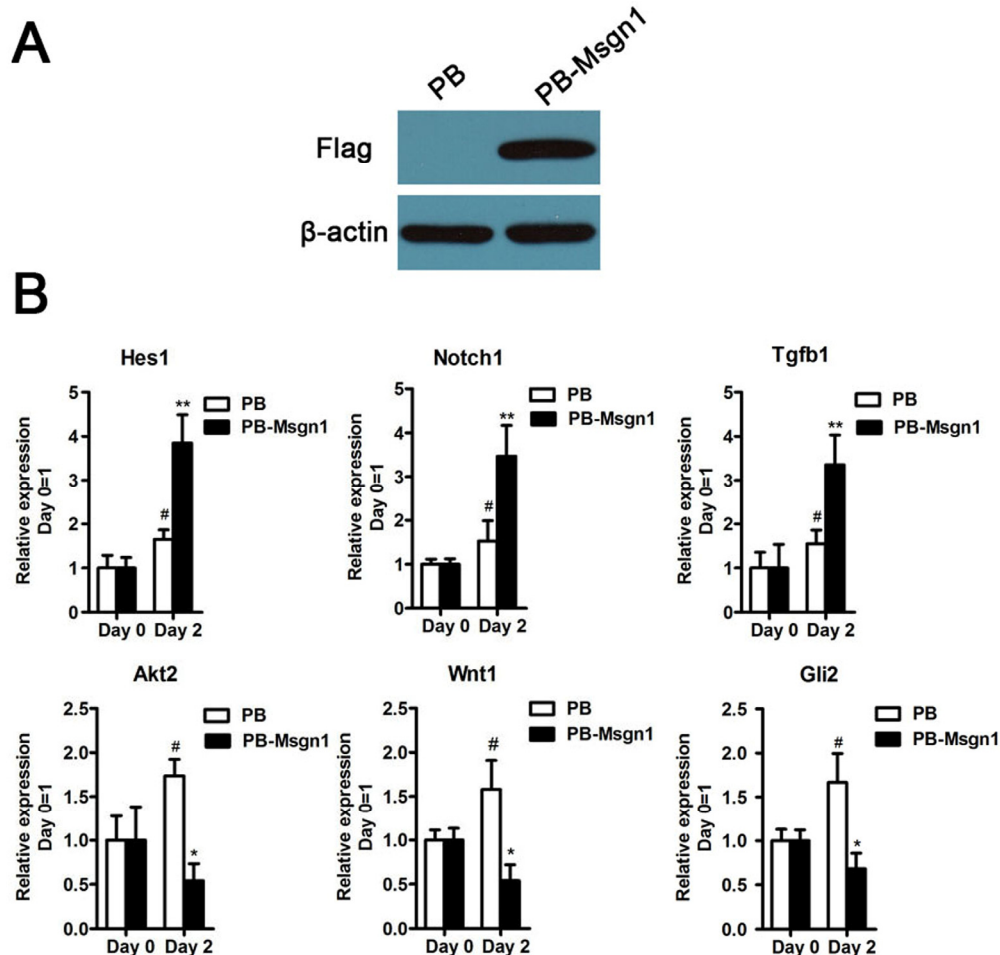


Figure 5. The validation results of qRT-PCR. (A) Flag-tagged *Msgn1* was introduced into 46C mESCs and the protein level of Flag-tagged *Msgn1* was determined by Western blot. β -actin was used as a loading control. (B) qRT-PCR analysis of the indicated gene expression for the indicated time in mESCs and mESCs-derived EBs. Data represent mean \pm s.d. of three biological replicates. #p < 0.05 vs PB:Day 0. *p < 0.05, **p < 0.01 vs PB-*Msgn1*:Day 0.

of *Notch1* was variable, presumably reflecting the dynamic expression of a cyclic gene, but was generally elevated by *Msgn1* overexpression. *Tgfb1* expression was upregulated in treated groups compared with untreated ones at 24h, while *Wnt1* expression was just opposite, which confirms Tgfb signaling pathway can activate Wnt signaling pathway via the expression of *TGF-β1*. The expressions of *Notch1* and *Hes1* were upregulated in treated groups compared with untreated ones from 12h to 48h, which was consistent with results shown in Figure 4. The expressions of *Akt2* and *Gli2* were downregulated compared with the control group while the *Notch1* expression was upregulated in treated groups, presumably reflecting other signaling pathways. The Wnt signaling pathway, for example, activates the Notch signaling pathway.

To further examine whether *Msgn1* could regulate these genes expressions during mESC differentiation, we generated a flag-tagged *Msgn1*-overexpressing mESC line using a PiggyBac vector (PB-*Msgn1*) in which *Msgn1* expression was efficiently enhanced (Figure 5A). According to our quantitative real-time PCR (qRT-PCR) results, the expression levels of above six genes at day 2 EBs could be detected. As shown in Figure 5B, the overexpression of *Msgn1* was able to induce higher expression levels of *Hes1*, *Notch1* and *Tgfb1* than the PB empty vector, while inhibiting the expressions of *Wnt1*, *Akt2* and *Gli2*. The validation of our qRT-PCR results was consistent with the previous gene expression profiles.

MATERIALS AND METHODS

Microarray data collection and preprocessing

First of all, we searched GEO (www.ncbi.nlm.nih.gov/geo/) for the gene expression profiling studies related to *Msgn1* overexpression. Data sets were included in our reanalysis if they met the following conditions: (1) the data were genome-wide, (2) *Msgn1* over expression datasets and *Msgn1* normal expression datasets were compared, (3) complete microarray raw or normalized data were available. Then, we got the dataset of GSE29848, contributed by Ravindra et al. [34]. In this data set, inducible A2lox-Flag *Msgn1* ES cells were differentiated to form embryoid bodies (EBs) for 2 days. Flag-*Msgn1* was induced on day 2 with doxycycline (DOX) and samples were collected at three time points, 12h, 24h and 48h after the addition of doxycycline (Table 2). Uninduced cells were used as controls, and experiments were performed in triplicate.

In order to assess the effect of preprocessing on the comparison, data preprocessing was performed using

software packages developed in Bioconductor ver. 2.6.0 and R ver. 2.10.1. Each Affymetrix data set was background-adjusted. Normalized log₂ probe-set intensities were calculated using the Robust Multichip Averaging (RMA) algorithm in the Affy package [35].

Gene set enrichment analysis

Our GSEA of pathways and genes was performed using the Category package in Bioconductor ver. 2.6.0 [36]. The target of GSEA is to reach a verdict as to whether the members of a gene set S are randomly distributed throughout the whole reference gene list L or are found primarily at the top or bottom of L. The relative robustness to noise and outliers in the data are remarkable advantages of GSEA. In our analysis, the gene sets of fewer than 10 genes were excluded. The t-statistic mean of the genes was calculated in each KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway. The threshold of the significance level p-values was chosen as 0.01 for the most significant pathways correlated with WD using a permutation test 1,000 times. Accordingly, the significant pathways and genes were then identified when comparing *Msgn1* overexpression and wildtype mice. The following category of identified signaling pathways is rooted in the KEGG pathway map br08901 of BRITE Functional Hierarchies in the KEGG database (http://www.genome.jp/keggbin/get_htext?br08901.keg). The annotation of significant genes in each signaling pathway was implemented by using the biomaRt package (<http://www.biomart.org/>) BioMart ver. 0.8 rc3 (version 0.8 of release candidate 3).

Regulatory elements and transcription factors of coregulated genes

To predict common regulatory elements (REs) for our input genes, we used a web server called DiRE (Distant Regulatory Elements of coexpressed genes, <http://dire.dcode.org/>), which is based on the Enhancer Identification (EI) method for predicting distant regulatory elements in higher eukaryotic genomes. DiRE predicts function-specific REs and scores the association of individual transcription factors (TFs) with the biological function shared by the group of input genes [37]. We selected a random set of 5,000 genes in the genome of mm9 (*Mus musculus*) as the source of background genes. TF occurrence for the centage of candidate REs which contain a conserved binding site for a special TF and TF importance for the product of TF occurrence and TF weight were two major parameters of our predicted TFs. The candidate associated TFs with the value of TF importance more than 0.05 would be selected.

Table 2. The summary of gene expression datasets.

Series	Contributor	Platform	Design		Replicates
			Time Points	Treatment	
GSE29848	Chalamalasetty et al.	Affymetrix Mouse Genome 430 2.0 Array	12h	MinDox	3
				PlusDox	3
			24h	MinDox	3
				PlusDox	3
			48h	MinDox	3
				PlusDox	3

It showed the summary of experiment design of dataset of GSE29848 used in our study. In Chalamalasetty et al.'s study, *Msgn1* was induced with doxycycline (PlusDox) and uninduced cells (MinDox) were used as controls. Samples were collected at three different time points including 12h, 24h and 48h. Experiments were performed in triplicate.

Table 3. The sequences of target genes primers for RT-PCR.

Genes	Forward(5'→3')	Reverse(5'→3')
Hes1	gagtgcataaacgaggtgac	cgttgatctgggcatgcag
Notch1	aggctacacaggaagcatgt	cgatgccatcctcacaagtg
Tgfβ1	cgcaacaacgccatctatga	actgctcccgaatgtctga
Wnt1	gtgacagtgacaacatcgat	ctggcgcatctcagagaaac
Akt2	acagccctcaagtatgcctt	accacatctctcagatgcaa
Gli2	atgccaaccagaacaagcag	tcttgacctgctccgctta

Cell culturing and differentiation

46C mouse embryonic stem cells (mESCs) were provided by Qi-Long Ying (University of Southern California), which were cultured on 0.1% gelatin-coated dishes. mESC medium contains DMEM(Sigma), 10% FBS (HyClone), 1× MEM nonessential amino acids (Invitrogen), 1× L-glutamine (Invitrogen), 1× penicillin and streptomycin (Sigma), 0.01 mM β-mercaptoethanol

(Invitrogen) and 1000 units/ml LIF(Millipore). mESCs were differentiated by forming embryoid bodies(EB) in the mESC medium without LIF.

Plasmid construction and cell transfection

The coding sequence of Mesogenin1(*Msgn1*) was inserted into a PiggyBac (PB) vector. The recombinant plasmid was then transduced into mESCs combined

with 2 μg transposase using LTX (Invitrogen). After transfection, mESCs were selected with 2 $\mu\text{g}/\text{ml}$ puromycin for 24 hours to obtain positive clones.

Western blotting

mESCs were lysed with RIPA buffer. Bradford (Bio-Rad) assay was used to determine the protein concentration. The proteins were separated with a 10% PAGE gel and electrotransferred onto a PVDF membrane. Probing was performed with specific primary antibodies and HRP-conjugated secondary antibodies. The primary antibodies used were FLAG (Sigma) and β -actin (Santa Cruz Biotechnology).

Quantitative real-time PCR (qRT-PCR)

Total RNA was extracted from the cells using TRIzol. cDNA was synthesized from 2 μg total RNA using reverse transcriptase (Takara) according to the manufacturer's instructions. Primers were designed using Primer 3.0 input software and are listed in Table 3. qRT-PCR was performed with TransStart Tip Green qPCR SuperMix (Takara). The expression level of each transcript was normalized with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and analyzed using the $2^{-\Delta\Delta\text{Ct}}$ method.

AUTHOR CONTRIBUTIONS

KH and DL designed the research. QZ, HZ, WL, DZ, YW, GW, SY, QB performed the experiments and/or data analysis. LT helped to perform the validation experiments. QZ, HZ, LT and DL wrote and revised the paper with input from all authors.

CONFLICTS OF INTEREST

The authors of this manuscript have no conflict of interest to declare.

FUNDING

We acknowledge financial support by the Scientific Research Foundation and Academic & Technology Leaders Introduction Project, and 211 Project of Anhui University (10117700023, 02303203-32030081), and The Student Research Training Program of Anhui University (J18520131), and Natural Science Foundation Project of Anhui Province (1508085MH189, 1508085QC63) as well as The Education Revitalization Project of Anhui Province: Stem Cell and Translational Medicine (Y05201374). We must also acknowledge the mESCs donation from the lab of Qi-Long Ying (University of Southern California).

REFERENCES

1. Kageyama R, Niwa Y, Isomura A, González A, Harima Y. Oscillatory gene expression and somitogenesis. *Wiley Interdiscip Rev Dev Biol.* 2012; 1:629–41. doi:10.1002/wdev.46
2. Fior R, Maxwell AA, Ma TP, Vezaro A, Moens CB, Amacher SL, Lewis J, Saúde L. The differentiation and movement of presomitic mesoderm progenitor cells are controlled by Mesogenin 1. *Development.* 2012; 139:4656–65. doi:10.1242/dev.078923
3. Dequéant ML, Pourquié O. Segmental patterning of the vertebrate embryonic axis. *Nat Rev Genet.* 2008; 9:370–82. doi:10.1038/nrg2320
4. Dubrulle J, McGrew MJ, Pourquié O. FGF signaling controls somite boundary position and regulates segmentation clock control of spatiotemporal Hox gene activation. *Cell.* 2001; 106:219–32. doi:10.1016/S0092-8674(01)00437-8
5. Dale KJ, Pourquié O. A clock-work somite. *BioEssays.* 2000; 22:72–83. doi:10.1002/(SICI)1521-1878(200001)22:1<72::AID-BIES12>3.0.CO;2-S
6. Wittler L, Shin EH, Grote P, Kispert A, Beckers A, Gossler A, Werber M, Herrmann BG. Expression of *Msgn1* in the presomitic mesoderm is controlled by synergism of WNT signalling and *Tbx6*. *EMBO Rep.* 2007; 8:784–89. doi:10.1038/sj.embor.7401030
7. Nowotschin S, Ferrer-Vaquer A, Concepcion D, Papaioannou VE, Hadjantonakis AK. Interaction of *Wnt3a*, *Msgn1* and *Tbx6* in neural versus paraxial mesoderm lineage commitment and paraxial mesoderm differentiation in the mouse embryo. *Dev Biol.* 2012; 367:1–14. doi:10.1016/j.ydbio.2012.04.012
8. Yabe T, Takada S. Mesogenin causes embryonic mesoderm progenitors to differentiate during development of zebrafish tail somites. *Dev Biol.* 2012; 370:213–22. doi:10.1016/j.ydbio.2012.07.029
9. Chalamalasetty RB, Dunty WC Jr, Biris KK, Ajima R, Iacovino M, Beisaw A, Feigenbaum L, Chapman DL, Yoon JK, Kyba M, Yamaguchi TP. The *Wnt3a*/ β -catenin target gene *Mesogenin1* controls the segmentation clock by activating a Notch signalling program. *Nat Commun.* 2011; 2:390. doi:10.1038/ncomms1381
10. Artavanis-Tsakonas S, Rand MD, Lake RJ. Notch signaling: cell fate control and signal integration in development. *Science.* 1999; 284:770–76. doi:10.1126/science.284.5415.770
11. Oswald F, Täuber B, Dobner T, Bourteele S, Kostezka

- U, Adler G, Liptay S, Schmid RM. p300 acts as a transcriptional coactivator for mammalian Notch-1. *Mol Cell Biol.* 2001; 21:7761–74. doi:10.1128/MCB.21.22.7761-7774.2001
12. Conlon RA, Reaume AG, Rossant J. Notch1 is required for the coordinate segmentation of somites. *Development.* 1995; 121:1533–45.
 13. van Eeden FJ, Granato M, Schach U, Brand M, Furutani-Seiki M, Haffter P, Hammerschmidt M, Heisenberg CP, Jiang YJ, Kane DA, Kelsh RN, Mullins MC, Odenthal J, et al. Mutations affecting somite formation and patterning in the zebrafish, *Danio rerio*. *Development.* 1996; 123:153–64.
 14. Huppert SS, Ilagan MX, De Strooper B, Kopan R. Analysis of Notch function in presomitic mesoderm suggests a gamma-secretase-independent role for presenilins in somite differentiation. *Dev Cell.* 2005; 8:677–88. doi:10.1016/j.devcel.2005.02.019
 15. Rida PC, Le Minh N, Jiang YJ. A Notch feeling of somite segmentation and beyond. *Dev Biol.* 2004; 265:2–22. doi:10.1016/j.ydbio.2003.07.003
 16. Jiang YJ, Aerne BL, Smithers L, Haddon C, Ish-Horowicz D, Lewis J. Notch signalling and the synchronization of the somite segmentation clock. *Nature.* 2000; 408:475–79. doi:10.1038/35044091
 17. Gao SC, Yin HB, Liu HX, Sui YH. [Research progress on MAPK signal pathway in the pathogenesis of osteoarthritis]. *Zhongguo Gu Shang.* 2014; 27:441–44.
 18. Yao E, Chuang PT. Hedgehog signaling: from basic research to clinical applications. *J Formos Med Assoc.* 2015; 114:569–76. doi:10.1016/j.jfma.2015.01.005
 19. Campbell V, Copland M. Hedgehog signaling in cancer stem cells: a focus on hematological cancers. *Stem Cells Cloning.* 2015; 8:27–38.
 20. Hwang M, Perez CA, Moretti L, Lu B. The mTOR signaling network: insights from its role during embryonic development. *Curr Med Chem.* 2008; 15:1192–208. doi:10.2174/092986708784310459
 21. Bruns AM, Horvath CM. Activation of RIG-I-like receptor signal transduction. *Crit Rev Biochem Mol Biol.* 2012; 47:194–206. doi:10.3109/10409238.2011.630974
 22. Huang X, Xu J, Huang M, Li J, Dai L, Dai K, Zhang X. Histone deacetylase1 promotes TGF- β 1-mediated early chondrogenesis through down-regulating canonical Wnt signaling. *Biochem Biophys Res Commun.* 2014; 453:810–16. doi:10.1016/j.bbrc.2014.10.021
 23. Xie G, Karaca G, Swiderska-Syn M, Michelotti GA, Krüger L, Chen Y, Premont RT, Choi SS, Diehl AM. Cross-talk between Notch and Hedgehog regulates hepatic stellate cell fate in mice. *Hepatology.* 2013; 58:1801–13. doi:10.1002/hep.26511
 24. Bedogni B, Warneke JA, Nickoloff BJ, Giaccia AJ, Powell MB. Notch1 is an effector of Akt and hypoxia in melanoma development. *J Clin Invest.* 2008; 118:3660–70. doi:10.1172/JCI36157
 25. Yamashita AS, Geraldo MV, Fuziwara CS, Kulcsar MA, Friguglietti CU, da Costa RB, Baia GS, Kimura ET. Notch pathway is activated by MAPK signaling and influences papillary thyroid cancer proliferation. *Transl Oncol.* 2013; 6:197–205. doi:10.1593/tlo.12442
 26. Zhang K, Zhang YQ, Ai WB, Hu QT, Zhang QJ, Wan LY, Wang XL, Liu CB, Wu JF. Hes1, an important gene for activation of hepatic stellate cells, is regulated by Notch1 and TGF- β /BMP signaling. *World J Gastroenterol.* 2015; 21:878–87.
 27. Chopra S, Kumar N, Rangarajan A, Kondaiah P. Context dependent non canonical WNT signaling mediates activation of fibroblasts by transforming growth factor- β . *Exp Cell Res.* 2015; 334:246–59. doi:10.1016/j.yexcr.2015.03.001
 28. Castilho RM, Squarize CH, Chodosh LA, Williams BO, Gutkind JS. mTOR mediates Wnt-induced epidermal stem cell exhaustion and aging. *Cell Stem Cell.* 2009; 5:279–89. doi:10.1016/j.stem.2009.06.017
 29. Beyer TA, Narimatsu M, Weiss A, David L, Wrana JL. The TGF β superfamily in stem cell biology and early mammalian embryonic development. *Biochim Biophys Acta.* 2013; 1830:2268–79. doi:10.1016/j.bbagen.2012.08.025
 30. Hidalgo M, Le Bouffant R, Bello V, Buisson N, Cormier P, Beaudry M, Darribère T. The translational repressor 4E-BP mediates hypoxia-induced defects in myotome cells. *J Cell Sci.* 2012; 125:3989–4000. doi:10.1242/jcs.097998
 31. Shang FF, Zhao W, Zhao Q, Liu J, Li DW, Zhang H, Zhou XF, Li CY, Wang TH. Upregulation of eIF-5A1 in the paralyzed muscle after spinal cord transection associates with spontaneous hindlimb locomotor recovery in rats by upregulation of the ErbB, MAPK and neurotrophin signal pathways. *J Proteomics.* 2013; 91:188–99. doi:10.1016/j.jprot.2012.12.002
 32. Olayioye MA, Badache A, Daly JM, Hynes NE. An essential role for Src kinase in ErbB receptor signaling through the MAPK pathway. *Exp Cell Res.* 2001; 267:81–87. doi:10.1006/excr.2001.5242
 33. Yamauchi J, Chan JR, Shooter EM. Neurotrophin 3 activation of TrkC induces Schwann cell migration through the c-Jun N-terminal kinase pathway. *Proc Natl Acad Sci USA.* 2003; 100:14421–26.

doi:10.1073/pnas.2336152100

34. Chalamalasetty RB, Dunty WC Jr, Biris KK, Ajima R, Iacovino M, Beisaw A, Feigenbaum L, Chapman DL, Yoon JK, Kyba M, Yamaguchi TP. The Wnt3a/ β -catenin target gene Mesogenin1 controls the segmentation clock by activating a Notch signalling program. *Nat Commun.* 2011; 2:390.
doi:10.1038/ncomms1381
35. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics.* 2004; 20:307–15.
doi:10.1093/bioinformatics/btg405
36. Chiaretti S, Li X, Gentleman R, Vitale A, Vignetti M, Mandelli F, Ritz J, Foa R. Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. *Blood.* 2004; 103:2771–78.
doi:10.1182/blood-2003-09-3243
37. Gotea V, Ovcharenko I. DiRE: identifying distant regulatory elements of co-expressed genes. *Nucleic Acids Res.* 2008; 36:W133-9. doi:10.1093/nar/gkn300

SUPPLEMENTARY MATERIAL

Please browse the links in Full Text versions to see the Supplementary Tables:

Table S-1. The significantly regulated pathways during somitogenesis under Msgn1 overexpression.

The details of all identified pathways according to GSEA of the data sets of 18 samples used in the microarray experiment by the comparison of Msgn1 overexpression to wild-type mice over a 48 hour timecourse were showed, which respectively includes pathways, functions, the number of genes, pvals, the number of TFs. There were 100 significant pathways identified with $p < 0.01$ at 12h, including 83 upregulated and 17 downregulated pathways. Similarly, 113 significant pathways, including 29 upregulated and 84 downregulated ones at 24h, and 183 significant pathways, including 50 upregulated and 113 downregulated ones at 48h, were identified.

Table S-2. The significantly regulated TFs during somitogenesis under Msgn1 overexpression.

TFs potentially involved in embryonic stem cells differentiating in the case of Msgn1 overexpression over a 48 hour timecourse. The table showed associated TFs with potential target genes which are co-regulated in each of above-mentioned 376 pathways and cutoff value of TF importance. Abbreviations: TFs, transcription factors.

Table S-3. The overlapping genes between microarray and ChIP-seq results in 12h, 24h and 48h.

Table S-4 The distribution of signaling pathways at three points.

Signaling pathways during somitogenesis under Msgn1 overexpression and KEGG pathway maps of signaling pathways at three points were showed. There were 1 signaling pathway only significantly regulated in 12h, 2 signaling pathways only significantly regulated in 24h and 2 signaling pathways only significantly regulated in 48h. 2 signaling pathways played roles at both 12h and 24h. 5 signaling pathways regulated significantly at 24h and 48h, 3 signaling pathways regulated significantly at 12h and 48h, 9 signaling pathways regulated at three different time points.

Table S-5. The associated genes in each significant pathways during somitogenesis under Msgn1 overexpression.

The associated genes in each significant pathway during somitogenesis under Msgn1 overexpression were showed. Data was performed using software packages developed in Bioconductor ver. 2.6.0 and R ver. 2.10.1. ID and gene symbol of 83

upregulated and 17 downregulated pathways at 24h, 29 upregulated and 84 downregulated ones at 24h, 50 upregulated and 113 downregulated ones at 48h were listed.