Identification of key pathways and hub genes in the myogenic process of pluripotent stem cell: A bioinformatics study

Wenyong Fei
Yangzhou University Medical Academy

Mingsheng Liu
Dalian Medical University

Yao Zhang
Dalian Medical University

Shichao Cao
Dalian Medical University

Xuanqi Wang
Dalian Medical University

Bin Xie
Dalian Medical University

Jingcheng Wang (jingchengwang123@163.com)
Yangzhou University Medical Academy

Research article

Keywords: stem cells, muscular regeneration, MyoD, actin, bioinformatics analysis

Posted Date: July 14th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-41291/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published on January 4th, 2021. See the published version at https://doi.org/10.1186/s13018-020-01979-x.
Abstract

**Purpose:** The study aims to determine the process of myogenic differentiation in human pluripotent stem cells and to figure out that the key pathways and hub genes in the process, which do helpful for the further research of muscular regeneration.

**Method:** Three gene expression profiles, GSE131125, GSE148994, GSE149055, about the comparisons of pluripotent stem cells and myogenic stem cells from the Gene Expression Omnibus (GEO) database. Common differentially expressed genes (DEGs) were obtained and for the further analysis as Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and GSEA analysis and protein-protein interaction (PPI) network. In vitro cell research to verify the hub genes and key pathways.

**Result:** 824 DEGs were co-expressed in the three GSEs. 19 hub genes were identified from the PPI network. The GO and KEGG pathway analysis were performed to determine the functions of DEGs. GSEA analysis indicated the differentiated cells were enriched in muscle cell development and myogenesis.

**Conclusion:** Our research revealed the main hub genes and modules in the myogenic process of stem cells which contribute to further study about the molecular mechanism of myogenesis regeneration. Paving a way for more accurate treatment for muscle dysfunction.

Introduction

As the aging of human population, muscle dysfunction has been an interrupting issue in clinical research.[1-4] A series of diseases were correlated to the atrophy of skeletal muscles and leading to dysfunction of muscular organs as Duchenne muscular dystrophy (DMD), degenerative rotator cuff tear etc [5-7]. Stem cells are promising cells that have the potency of multi-directional differentiation and proliferation and are widely expected to be used in the field of tissue repair and regeneration. [8, 9] In the muscle regeneration field, stem cells also showed vigorous potency. [10]

Various researches have devoted to verifying the mechanism of myoblast differentiation. Myogenic differentiation as a multistage process, there stays several regulating factors as Myf5, Myf6, myoD and myog.[11-13] Meanwhile, couples of pathways were verified to be correlated to myogenic differentiation as PI3K-MAPK, p38, p53 and actin pathway.[14-17] To up-regulate the differentiation efficacy and contribute to the repair of degenerated muscular tissues, it is especially important to clarify the differentiation mechanism at genetic level.

With the wide spread use and development of high-throughput sequencing, bioinformatics analysis showed great advantage for determining the myogenic differentiation mechanism of stem cell at genetic level. However, no study was designed to integrate the myogenic differentiation datasets in GEO. In the present study, we integrated 3 datasets in GEO comparing human pluripotent stem cells and myogenic stem cells. Bioinformatics analysis was used to explore molecular mechanism of the pathogenesis in myogenic differentiation of stem cells.

Materials And Methods

**Microarray data obtained:**

Three human gene expression profiles, GSE131125(GPL 20844, SurePrint G3 Human GE v3 8x60K Microarray 039494), GSE148994 and GSE149055(GPL16686, Affymetrix Human Gene 2.0 ST Array) were obtained from the
GEO database. Both the GSE149055 and GSE148994 contained 6 human samples, of which 3 were undifferentiated stem cells and 3 were differentiated stem cells. GSE133125 contained 24 samples which include different time-point of the differentiation. We choose the 3 undifferentiated stem cells and 3 differentiated for 25 days into our analysis.

**Identification of differently expressed genes (DEGs)**

The downloaded platform files were matched to the gene expression profiles by the “VLOOKUP” function of Excel 2010. Gene differential analysis was determined to summarized the differentially expressed genes (DEGs). The DEGs threshold of our study was \(|\log_{2}\text{FC}|>1\) and adj.P-value<0.01. Heatmaps of DEGs from 3 groups were generated by graphpad 8.0.2. Online tool Venn, version 2.1(bioinformatics.psb.ugent.be/webtools/Venn;version 2.1) was used to determine the common DEGs among the three profiles.

**(Protein-protein interaction) PPI network construction and module selection**

Search Tool for the Retrieval Interacting Genes (STRING) database was used to construct the network of differentially expressed genes and proteins and Molecular Complex Detection (MCODE; version 1.31) in the Cytoscape (version 3.8.0) was used to analysis modules in the network.

**GO and pathway enrichment analysis construction**

Both the GO and KEGG analysis was applied under the online program Database for Annotation, Visualization and Integrated Discovery (DAVID, version 6.8) whose subgroup of functional annotation tools can help the researchers to understand the biological meanings about the selected genes. Gene Set Enrichment Analysis (version 4.0.3) was used to verify whether DEGs Showed statistical significance in one phenotype or pathway based on the expression profiles.

**Cells sacrificed**

The rabbit adipose derived stem cells were sacrificed in our laboratory from an eight-week-old Newzealand white male rabbit. The subpatellar fat pad was harvested for the further step. The remaining tissues were minced finely into the shape of chyle, mixed with collagenase type (Tianjin Haoyang Biotechnology Comparison, China) and moved to a 50 ml centrifuge tube. At the end of digestion, the tissues were transferred into a centrifuge machine (Sigma Centrifuge, Germany), and the rotator speed was set at 1700 r/min, which lasted 10 minutes. The top content and supernatant were discarded, leaving the lower white block mass content in the tube. After centrifuging the liquid for 3 minutes at 800 r/min and discarding the supernatant, the lower content containing the stem cells was obtained finally. Subsequently, we inoculated the stem cells into the flask filled with 5 ml DMEM of fetal bovine serum (FBS, Gibco Company, St Louis, MO, USA) and observed the morphological characteristics under the invert microscope (Leica, Germany). The cells were cultured with the DMEM of fetal bovine serum subjected to culture and change every three days.

**Flow cytometry**

The cells were trypsinized, washed, and re-suspended in phosphate-buffered saline (PBS) and blocked with 3% fetal bovine serum for 15 minutes before the flow cytometry analysis. We divided the cells into 3 ADSCs tubes, labeling the three ADSCs tubes as blank, CD90 testing, and CD45 testing (Invitrogen, USA, and Gibco, USA). To every tube was added the corresponding fluorescein-labeled antibody whose concentration was set at a paralleled level of
5μl/ml, and the tubes were incubated at 37℃ in a 5% CO2-saturated humidity float tank for 30 minutes. We centrifuged the tubes and its contents at 1200 r/min for 5 minutes following the last incubation. We then discarded the supernatant, washed the remaining products twice with PBS, added appropriate amounts of 4% paraformaldehyde for 30 minutes, and assessed the mixture on the flow cytometry machine (BD FACSVerse, USA).

**Tetrazolium method [MTT]**

The passage-3 ADSCs were digested and diluted, and the mixture was transferred to a 96-well culture plate (Thermo Scientific, USA) at 100μl cell suspension per well. 5-Aza was then added to each well containing ADSCs and the BMSCs at concentrations of 0, 10, 20, 30, 40 μmol/l. The absorbance of each well was measured at 550 nm (OD) by a microplate reader (Biotek, USA) to detect the cell viability under the induction of different concentrations of 5-Aza.

**Immunohistochemical staining and Semi-quantitative analysis**

We determined the KEGG pathway of actin cytoskeleton about expression of actin by immunohistochemical staining. The cells were digested and diluted 9 days after induction and added 150μl 4% paraformaldehyde fixative to every slide and left them undisturbed for 30 minutes before adding 150μl 0.1% Triton x-100 microplate reader (Biotek, USA). Primary antibody α-SMA (1:200) (Proteintech, USA), secondary antibody (1:200) (Proteintech, USA) and Hoechst 33258 stain (C1011 Beyotine, China) was added to each slide in a dark environment at room temperature. Finally, we observed the cells under a fluorescence microscope (Leica, Germany), photographed, and stored them. ImageJ (Rawak Software, Germany) software was used for photography and Prism Demo software for data statistics (GraphPad Software, USA).

**Quantitative real-time PCR**

Total RNA was extracted from the ADSCs after induction of 9 days using Trizol lysate (Invitrogen). The schizolytic cells were then transferred into another tube without RNA enzymes, and 200 μl pre-cooling chloroform (Sigma Centrifuge, Germany) was added per milliliter of Trizol. The ccentrifugation yielded RNA sediments that were preserved in a -20°C surrounding for 30 minutes. The sediments were washed with 75% ethyl alcohol and centrifuged for 5 minutes, and the supernatant was discarded after washing and centrifuging the sediments twice. The reverse transcription system was prepared using a reverse transcription kit (Thermo Scientific, USA) according to instructions provided in the protocol of the kit.

**Statistical analysis**

Statistical analysis was performed on Graphpad 8.0.2. Expressed data were shown as mean±SD. Student’s t test was used to evaluate the statistical significance of different 3 groups. P value less than 0.05 was considered as significant.

**Results**

**Identification of DEGs**

The three datasets were standardized and the results are shown in Figure 1. The threshold of DEGs determination was that |LOG(FC)| lower than 1 and adj.P.Value lower than 0.01. From the GSE133125 database, there were 5051 up-regulated and 5199 down-regulated DEGs. Meanwhile, 864 up-regulated and 1038 down-regulated DEGs were calculated from GSE149055. As for GSE148994. As for GSE149055, there were 1068 up-regulated and 3913 down-regulated DEGs. Heat
map of DEGs in each dataset was shown in Figure 1C, D and E. The DEGs in each group were mixed by the Venn plot. From the Venn plot shown in Figure 1A and B, there were 824 common DEGs among the three subgroups, of which 350 were up-regulated DEGs and 474 were down-regulated.

**Protein-protein interaction (PPI) network construction and sub-modules**

824 notes and 3200 edges consist the full network shown in Figure 2A. Meanwhile, with the aid of the MCOD app, top 3 modules were selected and shown in Figure 2B,C,D with 28 notes and 349 edges in module 1, 36 notes and 237 edges in module 2 and 47 note notes and 176 edges in modules 3. From the MCOD function, 19 hub genes were selected: ASXL1, BOC, CENPH, DIMT1, ESRP1, GLDC, H0XD3, IGFBP5, JUN, MGST1, MRPS34, MSTN, MYOD1, MYOG, NBAS, PLS1, POLR3G, RNF144B, UST.

**GO and pathway enrichment analysis from the DEGs**

The GO analysis was processed to determine the function distributions of common DEGs from three aspect. Figure 3 showed up-regulated DEGs enrichment including KEGG pathways, molecular function (MF), biological processes (BP) and cell composition (CC). In KEGG analysis, the top 3 enriched pathways were pathway in cancer, PI3K pathwat, actin cytoskeleton regulation. DEGs were enrich in transcription functions in BP, extracellular communications in CC and DNA binding in MF. Meanwhile, the down-regulated DEGs showed in Figure 4A mainly distributed in metabolic pathway, biosynthesis of antibodies and cell cycle. In the up-regulated function analysis, from BP to CC and MF, MYOD1 showed significantly differentially expressed. According to the KEGG analysis, the enriched pathway “actin cytoskeleton regulation” was on the way of myogenisis differentiation. The GSEA analysis of DEGs was processed and the results were shown in Figure 5B and C. The DEGs were enrich in “myogenisis” and “muscle cell development”. The GSEA results were shown in Figure 5 C and D, MYOD1 was “core enrichment” gene in both enriched pathways. Therefore, we set the MYOD1 as the hub gene and “actin cytoskeleton regulation” pathway as the mainly enriched functional pathway.

**The expression of ADSCs and BMSCs surface markers**

Specific membrane markers confirmed the identity of ADSCs via flow cytometry. According to the results, the ADSCs results are presented in figure 6C, with a strong expression of CD90 at 82.8% positive and weak expression of CD45 at 4.58% positive and the results were shown in Figure 6C.

**Cell viability authenticated by MTT**

The data were converted into figure 6B to show cell viabilities when different concentrations of 5-Aza induced the cells. As is shown in the figure 6B, 5-Aza do have does dependent and time dependent toxic effects on ADSCs. It can be calculated that the IC50 in ADSCs groups were 9.178μmol/l at 24 hours, 6.469μmol/l at 48 hours and 8.664μmol/l at 72 hours after induction. The further induce concentration of 5-Aza was set as 0, 10 and 20μmol/l and named as group A, B and C, respectively.

**Actin expression determined by immunohistochemistry**

The results of the expression of actin were shown in Figure 6D. Actin was labeled and stained red by α-SAM, and the nucleus was stained blue by hochest, with the composed pictures showing that there were just parts of the cells expressing actin. The differentiated rate was calculated by graphpad 8.0.2. And the rate of each group was 0.019,
0.074 and 0.116 for groups A, B, and C, respectively. Of which the differentiate rate in group B and C was significantly up-regulated when compared to group A (P<0.01).

The content of myoD mRNA measured by RT-PCR

We further used the RT-PCR technology to detect the content of the mRNA of myoD in each group under the induction of 5-Aza. The results were recorded at 1.009, 2.391, and 4.876, respectively in each group. Of which the content in group C was significantly up-regulated than group B (p<0.05) whose content was also up-regulated compare to group A with significance (P<0.05).

Discussion

Mountainous efforts have been devoted to the research of pluripotent stem cells in our nowadays research for their regenerating and repairing damaged tissues effects.[18-20] In the musculoskeletal field, the degenerated and decreased of muscle tissue have confused the clinical effects of various diseases.[21, 22] The regeneration and remobilization of degenerated and damaged muscle tissues have been a hot issue in the research.[23] Stem cells, owning the myogenic differentiation, provide a possibility for current issue. However, the specific key pathways and genes in the myogenesis of stem cells is still under mystic.

There stands various signaling pathways which were count in the myogenic differentiate process of stem cells. Fu, S reported that PI3K pathway related genes and proteins were up-regulated expressed in the myogenic differentiate courses of mouse stem cells[24]. Meanwhile, up-regulated p53 and actin signaling pathways were also proved to be responsible for the myogenesis of stem cells which were certified by Liu, L.[25] and Petschnik, A. E, E [26]. Except that, p38 signaling pathway and wnt pathway were both proved to be responsible for the process[27, 28]. As for the myogenic genes, MRFs, myoD, myoG, etc[29, 30] were all reported as myogenic related genes.

In the present study, a bioinformatics analysis was used to analysis the key pathways and hub genes in myogenesis of stem cells based on 3 GEO databases. According to the analysis, a total of 824 DEGs were hunted out and applied for the further GO and KEGG analysis to certify potential biological functions and pathways in myogenic differentiation. Except that, 111 genes from the top 3 clusters and 19 hub genes analyzed from the MCODE method were identified from the PPI network.

MyoD has been described as the decisive gene and component of diverting undifferentiated cells into myoblasts[31]. In Rudnicki[32] research, knock out of myoD and myf5 results in the prevention of formation of skeletal muscle in the embryo period. The study results revealed that myoD and myf5 were determined genes in the origination of muscle cells. In the present bioinformatic analysis, myoD showed key effects in the myogensis. In GO analysis of up-regulated genes, myoD showed significant in positive regulation of myoblast differentiation in BP, transcription factor complex in CC and chromatin binding, transcription factor binding, transcription factor activity in MF. Meanwhile, the GSEA analysis revealed that moyD were both core enriched elements in myogenesis and muscle cell development in the three GEO databases. The unit results from our analysis revealed that myoD can be one of the hub genes in the myogenic differentiating process. From the laboratory experiments, RT-PCR results revealed that myoD were exactly significantly up-regulated in myogenic induced stem cells.

Also, we performed KEGG analysis to trace out the exact relevant pathways in the myogenic differentiation not only in the DEGs, but also based on the intensive module analysis from the PPI network. From the DEGs, the activation of PI3K, actin cytoskeleton regulation, p53 signaling pathway were proved to be tightly associated with myogenesis
process. Meanwhile, the intensive analysis showed that actin cytoskeleton regulation pathway was also enriched. From our laboratory experiments, the expression of actin was exactly significantly up-regulated in myogenic induced stem cells.

The study still has several limitations. Firstly, the included GEO profiles were still not rich enough. Secondly, the specific genes regulations in different time point of differentiation were omitted in our study. We still need to conduct further validate experiment to proof our speculation in the future.

**Conclusion**

Our study identified a series of DEGs in the myogenic differentiation process compared to undifferentiated stem cells. The 19 hub genes ASXL1, BOC, CENPH, DIMT1,ESRP1,GLDC,H0XD3,IGFBP5,JUN,MGST1,MRPS34,MSTN,MYOD1,MYOG,NBAS,PLS1,POLR3G,RNF144B,UST were selected from the series bioinformatics analysis. From the further GO and KEGG analysis, the pathways own enriched genes were selected. Our analysis revealed the hub genes and key pathways in the myogenic differentiation process of stem cells.

**Declarations**

**Ethics approval and consent to participate**

This study was approved by ethics of committee of Northern Jiangsu People's Hospital.

**Consent for publish**

Not applicable.

**Availability of data and materials**

The data used and analyzed during the current study are available from the corresponding author on reasonable request.

**Competing Interest**

The authors declare that they have no competing interests.

**Funding:**

This work was supported by the funds from the young medical key talent project of Jiangsu province (contract number QNRC2016458). Jiangsu Provincial Medical Innovation Team (Grant#CXTDB2017004).

**Authors’ contributions**

Wenyong Fei, Mingsheng Liu and Jingcheng Wang: conception and design, financial support, experiment, manuscript writing, final approval of manuscript.

Mingsheng Liu made the equal contribution to the article and should be considered co-first author. Correspondence: Jingcheng Wang.
Yao Zhang, Shichao Cao, Xuanqi Wang, Bin Xie: analysis and interpretation of data and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Wuhan Geek Translate Ltd for its linguistic assistance during the preparation of this manuscript.

References

13. Huang, W., et al., The Inhibition on MDFIC and PI3K/AKT Pathway Caused by miR-146b-3p Triggers Suppression of Myoblast Proliferation and Differentiation and Promotion of Apoptosis. Cells, 2019. 8(7).


**Figures**
Figure 1

Figure legends not provided with this version of the manuscript.
Figure 2

Figure legends not provided with this version of the manuscript.
Figure 3

Figure legends not provided with this version of the manuscript.
Figure 4

Figure legends not provided with this version of the manuscript.
Figure 5

Figure legends not provided with this version of the manuscript.
Figure 6

Figure legends not provided with this version of the manuscript.