

胰岛素分泌细胞诱导过程中 miRNA-mRNA 调控网络的生物信息学分析

王涛, 潘鑫

(锦州医科大学医疗学院医学系, 辽宁 锦州 121013)

摘要:目的 探讨人胚胎干细胞(hESCs)向胰岛素分泌细胞(IPCs)诱导分化过程中基因表达的差异,并构建 miRNA-mRNA 调控网络。**方法** 选择 GEO 数据库内的 GSE42094 数据集作为研究对象,应用 GEO2R 分析差异表达基因,对其进行 GO 功能和 KEGG 路径以及蛋白互作网络分析,选用 miRTarBase 数据库预测其靶 miRNA,对 miRNA-mRNA 调控网络进行可视化,并采用文献数据验证。**结果** 共得到差异表达基因 188 个,选取诱导后显著下调基因共 21 个,分析发现 POU5F1、DNMT3B 和 NANOG 3 个基因集中在“基因表达调控”,且下调基因在“胞核内”比较集中;通过查找靶 miRNA,发现 miR-335-5p 能够调控 POU5F1、DNMT3B、NANOG、FAM124B 和 LECT1 5 个基因的表达;验证发现有 24 个 miRNA 能够被文献数据验证。**结论** 基于显著下调基因构建的 miRNA-mRNA 调控网络参与干细胞诱导分化进程并发挥重要作用。

关键词:微小 RNA;胚胎干细胞;胰岛素分泌细胞

中图分类号:R342.2

文献标识码:A

DOI:10.3969/j.issn.1006-1959.2023.18.002

文章编号:1006-1959(2023)18-0008-07

Bioinformatic Analyses of miRNA-mRNA Regulatory Network During Differentiation of Insulin-producing Cells

WANG Tao, PAN Xin

(Medicine Department of Medical Treatment College, Jinzhou Medical University, Jinzhou 121013, Liaoning, China)

Abstract: **Objective** To identify differentially expressed genes during differentiation of human embryonic stem cells (hESCs) into insulin-producing cells, and to construct the miRNA-mRNA regulatory network. **Methods** The datasets GSE42094 from GEO were employed in this study, differentially expressed genes were analyzed using GEO2R, and the GO function, KEGG pathway and protein interaction network were analyzed. The miRTarBase database was used to predict its target miRNA, and the miRNA-mRNA regulatory network was visualized and verified by literature data. **Results** A total of 188 differentially expressed genes were identified and 21 down-regulated genes after the induction were selected as candidate genes. POU5F1, DNMT3B and NANOG were "regulation of gene expression" and the significant term for all down-regulated genes were "nucleus". By searching for target miRNAs, it was found that miR-335-5p could regulate the expression of POU5F1, DNMT3B, NANOG, FAM124B and LECT1; validation found that 24 miRNAs could be verified by literature data. **Conclusion** The miRNA-mRNA regulatory network based on down-regulated genes may play a key role during induction of insulin-producing cells.

Key words: microRNA; Embryonic stem cells; Insulin-secreting cells

miRNA 作为一类小的非编码 RNA,其长度大约在 22 个核苷酸左右,能够与靶 mRNA 3' 端非翻译区互补,进而降解或阻止该 mRNA 翻译出蛋白质,实现转录后水平的基因表达调控^[1]。糖尿病作为一种慢性疾病,部分是由胰岛 β 细胞受损而引起,胰腺或胰岛移植对于该类疾病治疗效果较好,但是由于供体缺乏以及免疫排斥导致临床应用受限^[2-3],而干细胞分化为胰岛素分泌细胞 (insulin-producing cells, IPCs) 再行移植,有望彻底根治^[4-5]。生物信息学

通过整合多个学科领域信息和知识,能够从整体层面揭示疾病或生物过程的复杂分子机制^[6-8]。本研究运用生物信息学方法分析人胚胎干细胞 (human embryonic stem cells, hESCs) 向 IPCs 诱导分化过程中 mRNA 的表达谱,预测靶向作用于该 mRNA 的 miRNA,构建 miRNA-mRNA 调控网络。通过对调控网络的分析,拟找到 IPCs 诱导分化的关键节点,为进一步开展分子靶向干预提供新思路。

1 资料与方法

1.1 资料来源 选择 GEO (<https://www.ncbi.nlm.nih.gov/geo/>) 数据库内的 GSE42094 数据集作为研究对象,该数据集包含 23 个样本数据 (GSM1032319-41),涉及未分化 hESCs、IPCs 分化的 5 个时期、胰腺内胰岛和胎胰。选择未分化 hESCs (hESCs 组) 和 IPCs 分

基金项目:辽宁省教育厅科学研究经费面上项目(编号:LJKZ1374)

作者简介:王涛(1984.10-),男,山东潍坊人,硕士,讲师,主要从事干细胞的基础研究

化的 5 个时期 (Diff1、Diff2、Diff3、Diff4 和 IPCs 组) 研究 mRNA 的表达谱。应用数据库自带软件 GEO2R 对 6 组数据进行分析,获取差异表达的 mRNA。

1.2 GO 功能和 KEGG 路径分析 选用 DAVID(<https://david.ncifcrf.gov/tools.jsp>) 数据库进行功能富集分析,主要分析 GO 功能以及 KEGG 路径,以 $P<0.05$ 为差异有统计学意义。

1.3 蛋白互作网络分析 选用 STRING(<https://string-db.org/>) 数据库进行蛋白互作网络 (protein-protein interaction, PPI) 分析,选择 Homo sapiens 作为“Organism”的条件,其余参数均为系统默认设置。

1.4 miRNA 预测 针对差异表达基因预测其靶 miRNA,选用 miRTarBase (<https://mirtarbase.cuhk.edu.cn/>) 数据库,选择 reporter assay、western blot、qPCR 和 microarray4 种实验验证的 miRNA-mRNA 靶向匹配作为预测 miRNA 的入选标准,并应用 Cytoscape 3.9.1 软件对 miRNA-mRNA 的调控网络进行可视化。

1.5 预测 miRNA 的验证 将预测的 miRNA 与 PMID:20735361^[9] 的 Table S4 数据取交集,分析 miRNA 在人胚胎 T3 干细胞向胰腺胰岛样细胞团分化过程中的表达,其 Table S4 数据显示出胚胎 T3 干细胞(hES-T3 cells grown on mouse embryonic fibroblast feeder, T3ES)、胚胎内胚层(embryoid bodies differentiated from T3 cells, T3EB)和胰腺胰岛样细胞团(pancreatic islet-like cell clusters derived from T3 cells, T3pi)3 组标准化的 miRNA 表达量。

1.6 转录因子-靶基因调控网络的构建 针对验证的 miRNA 以及显著下调基因查找调控其表达的转录因子,选用 FunRich 3.1.3 软件分析转录因子,以 $P<0.05$ 为差异有统计学意义,并作为转录因子的入选标准,同时应用 Cytoscape 3.9.1 插件 iRegulon 预测转录因子,将二者分别与 Mahmoud H 等^[10]报道的 β 细胞分化相关的重要转录因子取交集,对转录因子-靶基因调控网络可视化。

2 结果

2.1 差异及因及其 GO 功能和 KEGG 路径分析 共得到差异表达的基因 188 个,选取诱导后显著下调的基因共 21 个(表 1),对其进行 GO 功能和 KEGG 路径分析,发现下调基因中 POU5F1、DNMT3B、NANOG 和 UCA1 集中在基因表达调控,POU5F1 和 NANOG 与内胚层的决定以及成体干细胞种群维持有关,PPP1R16B 和 PPP2R2C 与蛋白磷酸化调节活性有关。DAVID 数据库进行 KEGG 路径显著性分析并没有找到有统计学意义的信号通路。

2.2 蛋白互作网络分析 通过 PPI 分析 21 个诱导后显著下调的基因,发现 LOC729860、UCA1 和 HLA-DPB2 未被 STRING 数据库识别,因此该网络有 18 个节点,14 条边,其 PPI 富集 P 值为 $1.0\text{e-}09$,差异有统计学意义。进一步分析发现 POU5F1、UTF1、DNMT3B、NANOG、GABRB3、DPPA2 和 TCL1B 形成互作网络,剩余 11 个蛋白相互之间没有关联,见图 1。

表 1 表达下调的差异基因

ID	Gene.symbol	F	P	ID	Gene.symbol	F	P
ILMN_1705546	POU5F1	325.3	4.24E-14	ILMN_2401344	PPP2R2C	99	1.65E-10
ILMN_1851610	LOC729860	243.9	3.19E-13	ILMN_1749878	FAM124B	97.9	1.78E-10
ILMN_1768793	UTF1	204.2	1.10E-12	ILMN_1722527	LECT1	94.1	2.34E-10
ILMN_3239254	UCA1	184.9	2.20E-12	ILMN_1790197	VSIG10	86	4.32E-10
ILMN_2328972	DNMT3B	149.9	9.46E-12	ILMN_1780790	DPPA2	86	4.34E-10
ILMN_1727098	PPP1R16B	134.9	1.96E-11	ILMN_2382309	TCL1B	85.8	4.39E-10
ILMN_2112811	RPL36A	129.8	2.55E-11	ILMN_1785406	NAP1L3	82.7	5.64E-10
ILMN_1659013	NANOG	124.2	3.48E-11	ILMN_1793770	DNAJB6	81	6.50E-10
ILMN_1800468	KLKB1	123.7	3.58E-11	ILMN_2205405	HLA-DPB2	79.6	7.30E-10
ILMN_1687090	GABRB3	115.3	5.79E-11	ILMN_1654606	ZDHHC22	78.7	7.88E-10
ILMN_1755047	LRRC2	109.5	8.26E-11				

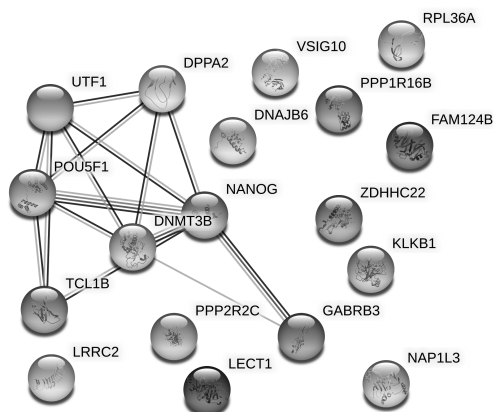


图 1 蛋白互作网络分析

2.3 miRNA-mRNA 调控网络 通过 miRTarBase 数据库查找靶向作用于显著下调基因的 miRNA, 选择 reporter assay、western blot、qPCR 和 microarray 4 种实验验证的 miRNA-mRNA 靶向匹配作为预测 miRNA 的入选标准, 共得到 25 个预测的 miRNA, 应用 Cytoscape 3.9.1 软件对 miRNA-mRNA 调控网络进行可视化(图 2), 发现 PPI 网络的 7 个关联蛋白也处于该调控网络的核心区域, 预测的大部分 miRNA 围绕其展开, 其中, miR-335-5p 能够调控 POU5F1、DNMT3B、NANOG、FAM124B 和 LECT1 的表达, 进一步在 GenomeNet 数据库(<https://www.genome.jp/pathway/hsa04550+5460>)中对这 5 个基因进行 KEGG 路径分析, 发现 POU5F1 和 NANOG 处于干细胞多能性调控信号通路(hsa04550)之中, 调控干细胞的自我更新。

2.4 预测 miRNA 的验证 将预测的 miRNA 与 PMID: 20735361 的 Table S4 数据取交集, 得到 24 个验证的 miRNA, 只有 1 个预测的 miRNA(miR-200b-3p)未

被验证, 大部分预测的 miRNA 得到验证, 即存在于干细胞及其诱导分化的过程中, 表明这种基于 4 种实验预测 miRNA 的方法具备一定的可信度(表 2)。表 2 显示 miR-26b-5p、miR-106b-5p、miR-200c-3p、miR-148a-3p、miR-375、miR-204-5p、miR-7-5p、miR-302a-3p 和 miR-124-3p 表达量均在胚胎内胚层阶段(T3EB)显著升高而又在胰腺胰岛样细胞团阶段(T3pi)显著降低, miR-26a-5p、miR-335-5p 和 miR-134-5p 表达量随着诱导进程逐渐升高。这两类 miRNA 的表达特点与显著下调基因表达量在 T3EB 阶段后显著降低形成对照, 提示二者之间存在负性调控。结合图 2 分析发现 miR-26b-5p 和 miR-335-5p 将 miRNA-mRNA 调控网络核心区域的 7 个基因往外延伸, 其中, LRR2 通过 miR-26b-5p、FAM124B 和 LECT1 通过 miR-335-5p 与核心区域关联。进一步将仅通过 microarray 实验预测的 miRNA 剔除后作调控网络图(图 3), 发现大部分 miRNA 主要调控 POU5F1、DNMT3B 和 NANOG 的表达, 该部分 miRNA 是通过 reporter assay、western blot 和 qPCR 实验验证了其与其靶基因的匹配关系, 其中, miR-200c-3p、miR-148a-3p、miR-204-5p、miR-302a-3p、miR-26a-5p、miR-134-5p 和 miR-335-5p 与其靶基因表达模式存在负性调控, 可得到进一步验证, miR-369-5p 也被上述 3 种实验所验证, 但是其在整个诱导过程中表达量未发生变化。结合表 2 发现, 仅通过 microarray 实验预测的 miRNA, 如 miR-26b-5p、miR-106b-5p、miR-375、miR-7-5p 和 miR-124-3p 与其靶基因的表达模式存在负性关系, 或可佐证该部分 miRNA 与其靶基因的靶向调控。

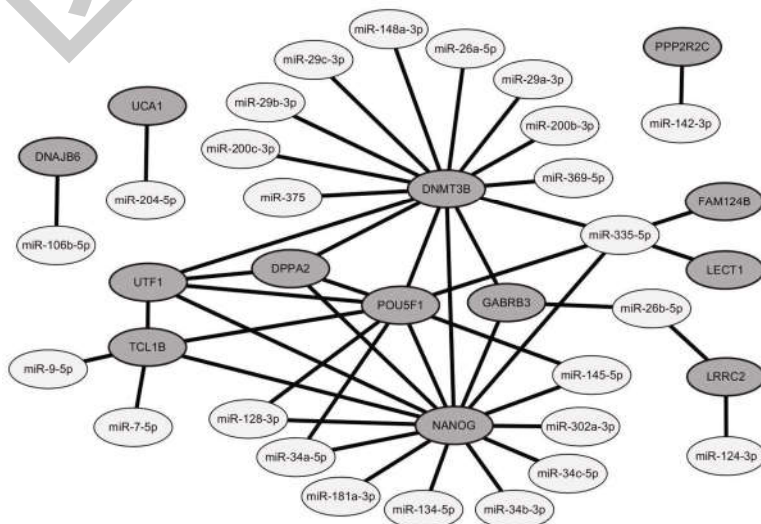


图 2 miRNA-mRNA 调控网络

包括其自身亦即存在顺式作用调控,POU5F1 调控 6 个基因表达,也存在顺式作用调控,且二者之间互为对方转录因子,存在相互转录调控现象,而 TFAP2A 调控 6 个基因表达,其中包含 POU5F1 和 NANOG 基因。ZDHHC22、PPP1R16B 和 GABRB3 3 个基因同时受上述 3 个转录因子的调控,结合 POU5F1 和 NANOG 基因表达量显著下降的变化,可以理解该部分显著下调基因主要存在转录水平的调控。进一步将 iRegulon 分析得到的 49 个转录因子与 Mahmoud H 等^[10]报道的 β 细胞分化相关的重要转录因子取交集,得到 ONECUT1(HNF6)、NEUROD1 和 NKX2-2 3 个转录因子(图 4D),发现 ONECUT1 能够调控 8 个基因表达,其中包括 NANOG 基因,NEUROD1 与 ONECUT1 共同调控 PPP2R2C、ZDHHC22 和 LECT1 3 个基因,而 LECT1 基因也受 NKX2-2 转录因子调控,可见该部分基因除受转录水平调控外还存在较强的转录后水平调控,以至于其表达量在分化后期显著下降。

2.6 验证 miRNA 的转录因子 针对 24 个验证的 miRNA 查找调控其表达的转录因子,FunRich 软件得到 158 个有统计学意义的转录因子,将其与 Mahmoud H 等^[10]报道的 β 细胞分化相关的重要转

录因子取交集,得到 NKX6-1、ONECUT1、PAX6、MAFB、HNF4A、PDX1、PAX4 和 MNX1(HB9)8 个转录因子(图 5A)。iRegulon 插件分析得到 52 个转录因子,将其与 188 个差异表达基因取交集,得到 TFAP2A、NR2F2 和 FOXA2 3 个转录因子,进一步与 Mahmoud H 等报道的 β 细胞分化相关的重要转录因子取交集,得到 ONECUT1、FOXA2 和 PAX4 3 个转录因子,将上述两部分转录因子融合分析,发现 FOXA2 既是 β 细胞分化相关的重要转录因子也是 188 个差异表达基因之一,其表达量见图 5B,可见其在整个诱导分化过程中一直处于高表达状态,对整个诱导分化进程发挥转录水平的调控。应用 Cytoscape 对转录因子-miRNA 可视化(图 5C),发现转录因子 TFAP2A 能够调控 7 个 miRNA 表达,其中 miR-375、miR-148a-3p、miR-124-3p 和 miR-200c-3p 表达量均在 T3EB 阶段显著升高而又在 T3pi 阶段显著降低,这与 TFAP2A 的表达模式一致(图 5D),表明存在较强转录水平的调控。而 miR-142-3p、miR-9-5p 和 miR-34a-5p 表达量在整个诱导分化过程中一直下降,提示这 3 个 miRNA 还受其他转录因子的调控或者越到分化后期其转录后水平的调控越强以至于检测到的 miRNA 表达量越低。

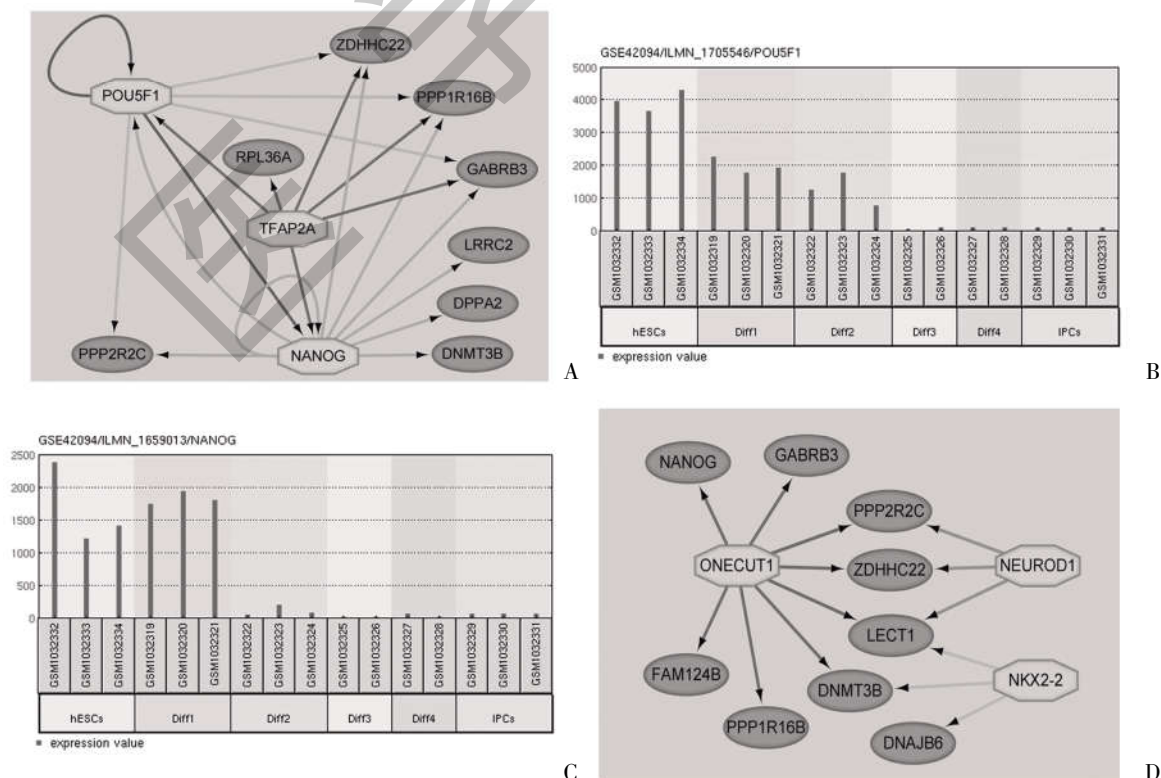


图 4 转录因子-靶基因调控网络

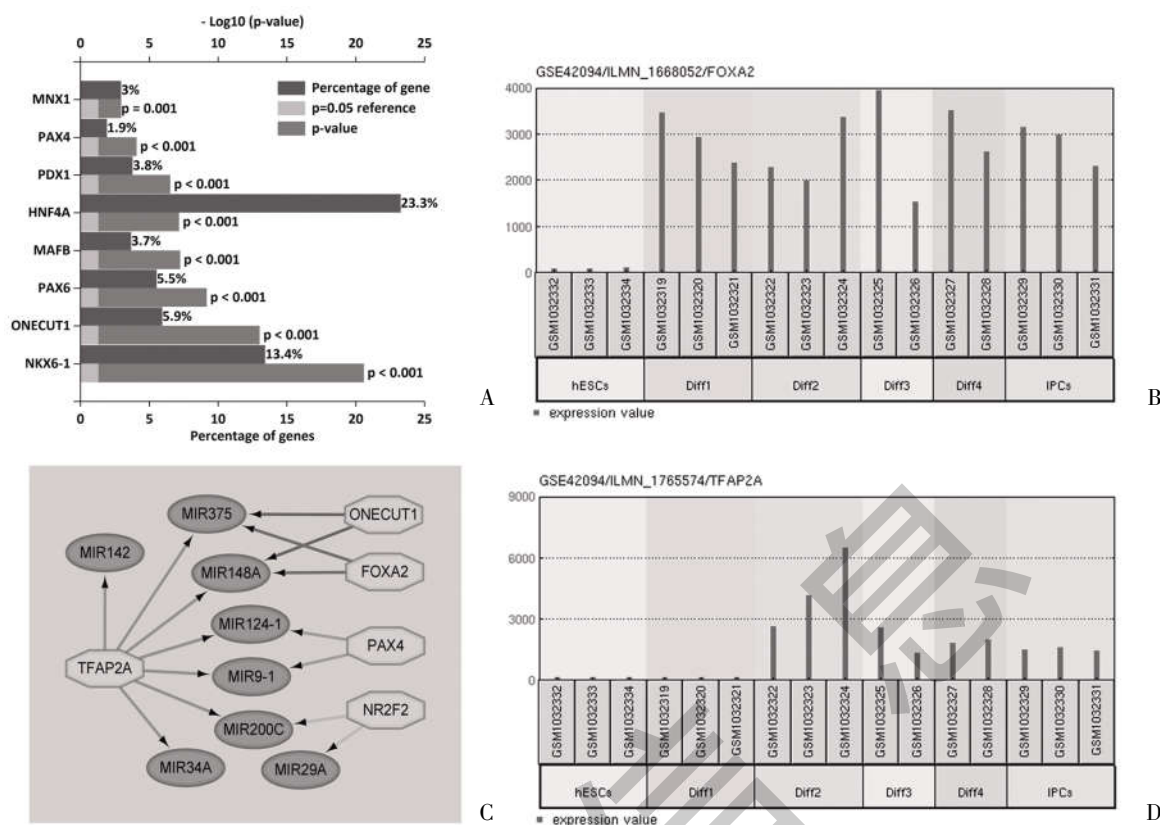


图 5 miRNA 的转录因子

3 讨论

本研究通过对 GSE42094 的 6 组样本数据分析,得到 188 个差异表达的基因,选取诱导后显著下调的 21 个基因进行后续研究。选用 DAVID 数据库对其进行 GO 功能和 KEGG 路径显著性分析,发现其中 POU5F1 和 NANOG 基因与基因表达调控、内胚层的决定和成体干细胞种群维持 3 个功能均有相关性,但没有找到有统计学意义的 KEGG 路径。进一步在 GenomeNet 数据库中检索 POU5F1 和 NANOG 基因的 KEGG 路径,发现二者同处于干细胞多能性调控信号通路,能够调控干细胞的自我更新,这与其 GO 功能相呼应。另外,该通路 TGF- β 、MAPK、PI3K-Akt 以及 Wnt 4 条信号通路相关联,这与 Mahmoud H 等^[10]报道的体外 β 细胞分化的关键信号通路相一致。

通过构建 miRNA-mRNA 调控网络发现, LOC729860、UTF1、KLKB1、DPPA2、NAP1L3、HLA-DPB2 等 6 个基因没有对应 miRNA 调控,而 PPP1R16B、RPL36A、VSIG10、ZDHHC22 等 4 个基因虽有相应 miRNA 调控,但不满足设定的人选标准,因此,这 10 个基因在应用 Cytoscape 3.9.1 软件进行 miRNA-mRNA 调控网络可视化过程中未有

miRNA 呈现。网络显示 miR-335-5p 能够调控 POU5F1、NANOG、FAM124B、DNMT3B 和 LECT1 5 个基因的表达。其中,POU5F1 和 NANOG 基因较为熟知^[11-13],而 FAM124B 在人 HeLa 细胞内能够与染色质解旋酶 DNA 结合蛋白 7 (chromodomain helicase DNA binding protein 7,CHD7)相互作用,而 CHD7 基因在小鼠 ESCs 内与 OCT4、NANOG 以及 SOX2 同处于一个增强子元件,提示 FAM124B 作为一个关联因子参与增强子介导的转录^[14,15]。DNA 甲基转移酶 3B (DNA methyltransferases 3B,DNMT3B)是 DNA 甲基转移酶蛋白家族成员之一,DNMT1 主要负责维持 DNA 甲基化,DNMT3A 和 DNMT3B 负责从头合成甲基转移酶^[16,17]。在 hESCs 内,DNMT3B 的表达量是已分化细胞的近 30 倍,其缺失能够影响线粒体的生成以及提高细胞内 α -酮戊二酸的水平,进而引起 hESCs 的多向分化^[18]。白细胞源性趋化蛋白 1 (leukocyte cell-derived chemotaxin 1,LECT1) 又称软骨源性抑制因子-1 (chondromodulin-1),是一个 25 kDa 的糖蛋白,由跨膜蛋白经转录后修饰生成,其有助于蛋白聚糖的合成以及体外软骨细胞的生长^[19]。在 hESCs 内,LECT1 表达量是已分化细胞的近 40 倍,其和 POU5F1、NANOG 一样被作为干性维持

基因^[20]。因此,miR-335-5p所靶向调控的这5个基因均与干性维持有关,提示miR-335-5p通过负性调控该类基因的表达,抑制hESCs的干细胞多能性调控信号通路,促使其转向分化。另外,miR-335-5p表达量在诱导分化过程中逐渐升高已由文献数据所验证,这与上述干性维持基因的显著下调形成对照。

2021年,Sabouri E等^[21]报道miR-375、miR-7、miR-186、miR-26a、miR-690以及miR-302a 6个IPCs诱导分化过程中的重要miRNA。本研究同样发现miR-375和miR-26a能够靶向调控DNMT3B基因表达,miR-7能够调控TCL1B以及miR-302a能够调控NANOG基因表达。可见该miRNA-mRNA调控网络具备一定可信度,通过进一步查找双方的转录因子也可部分地佐证,因此,基于显著下调基因构建的该miRNA-mRNA调控网络在干细胞诱导分化过程中发挥重要作用。

综上所述,本研究基于显著下调基因构建的miRNA-mRNA调控网络参与干细胞诱导分化进程并发挥重要作用。

参考文献:

- [1]Rahman MH,Peng S,Hu X,et al.A Network-Based Bioinformatics Approach to Identify Molecular Biomarkers for Type 2 Diabetes that Are Linked to the Progression of Neurological Diseases[J].Int J Environ Res Public Health,2020,17:1035.
- [2]Ma X,Jain NM,Hitscherich P,et al.Stem Cell-Derived Insulin-Producing Cells in 3D Engineered Tissue in a Perfusion Flow Bioreactor[J].Tissue Eng Part A,2021,27(17-18):1182-1191.
- [3]Refaie AF,Elbassoumy BL,Kloc M,et al.From Mesenchymal Stromal/Stem Cells to Insulin-Producing Cells: Immunological Considerations[J].Front Immunol,2021,12:690623.
- [4]Câmara BOS,Bertassoli BM,Ocarino NM,et al.Differentiation of Mesenchymal Stem Cells from Humans and Animals into Insulin-producing Cells: An Overview In Vitro Induction Forms [J].Curr Stem Cell Res Ther,2021,16(6):695-709.
- [5]Abazari MF,Zare Karizi S,Hajati-Birgani N,et al.PHBV nanofibers promotes insulin-producing cells differentiation of human induced pluripotent stem cells[J].Gene,2021,768:145333.
- [6]袁长深,容伟明,卢智贤,等.基于生物信息学构建骨肉瘤miRNA-mRNA的调控网络[J].中国组织工程研究,2021,25(17):2740-2746.
- [7]Lu Y,Li Y,Li G,et al.Identification of potential markers for type 2 diabetes mellitus via bioinformatics analysis [J].Molecular Medicine Reports,2020,22:1868-1882.
- [8]梁学振,谢国鑫,李嘉程,等.基于miRNA-mRNA调控网络股骨头坏死的关键基因筛选与分析 [J].中国组织工程研究,2022,26(11):1800-1807.
- [9]Chen BZ,Yu SL,Singh S,et al.Identification of microRNAs expressed highly in pancreatic islet-like cell clusters differentiated from human embryonic stem cells[J].Cell Biol Int,2011,35(1):29-37.
- [10]Mahmoud H,Elham H.Redefining the signaling pathways from pluripotency to pancreas development: In vitro β -cell differentiation[J].J Cell Physiol,2019,234(6):7811-7827.
- [11]Verdelli C,Morotti A,Tavanti GS,et al.The Core Stem Genes SOX2, POU5F1/OCT4, and NANOG Are Expressed in Human Parathyroid Tumors and Modulated by MEN1, YAP1, and β -catenin Pathways Activation [J].Biomedicines, 2021,9:637.
- [12]Choi SH,Kim JK,Jeon HY,et al.OCT4B Isoform Promotes Anchorage-Independent Growth of Glioblastoma Cells [J].Mol Cells,2019,42(2):135-142.
- [13]Fogarty NME,McCarthy A,Snijders KE,et al.Genome editing reveals a role for OCT4 in human embryogenesis[J].Nature, 2017,550(7674):67-73.
- [14]Kuang Y,Wang Y,Cao X,et al.New prognostic factors and scoring system for patients with acute myeloid leukemia [J].Oncol Lett,2021,22(6):823.
- [15]Yu X,Han Y,Liu S,et al.Analysis of Genetic Alterations Related to DNA Methylation in Testicular Germ Cell Tumors Based on Data Mining[J].Cytogenet Genome Res,2021,161(6-7):382-394.
- [16]Lai SC,Su YT,Chi CC,et al.DNMT3b/OCT4 expression confers sorafenib resistance and poor prognosis of hepatocellular carcinoma through IL-6/STAT3 regulation[J].J Exp Clin Cancer Res,2019,38(1):474.
- [17]Gagliardi M,Strazzullo M,Matarazzo MR.DNMT3B Functions: Novel Insights From Human Disease [J].Front Cell Dev Biol,2018,6:140.
- [18]Cieslar-Pobuda A,Ahrens TD,Caglayan S,et al.DNMT3B deficiency alters mitochondrial biogenesis and α -ketoglutarate levels in human embryonic stem cells [J].Stem Cells,2020,38:1409-1422.
- [19]Tsuyuguchi Y,Nakasa T,Ishikawa M,et al.The Benefit of Minced Cartilage Over Isolated Chondrocytes in Atelocollagen Gel on Chondrocyte Proliferation and Migration [J].Cartilage, 2021,12(1):93-101.
- [20]Zhu S,Qiu H,Bennett S,et al.Chondromodulin-1 in health, osteoarthritis, cancer, and heart disease [J].Cell Mol Life Sci, 2019,76(22):4493-4502.
- [21]Sabouri E,Rajabzadeh A,Enderami SE,et al.The Role of MicroRNAs in the Induction of Pancreatic Differentiation[J].Current Stem Cell Research & Therapy,2021,16(2):145-154.

收稿日期:2022-11-28;修回日期:2022-12-09

编辑/成森