

· 论 著 ·

基于基因共表达权重网络分析骨质疏松相关的关键 LncRNA

黎永华* 谭仁霆 黄世福 邱金龙

海南省儋州市中医医院骨伤科,海南 儋州 571700

中图分类号: R34;R589. 5 文献标识码: A 文章编号: 1006-7108(2021) 04-0556-07

摘要: **目的** 通过基因共表达权重网络分析(WGCNA)探讨 LncRNA 与骨质疏松的关系。**方法** 收集 GEO 数据库中 GSE56815 芯片的 80 个样本,通过重注释获得了 127 LncRNA,用 WGCNA 对表达数据进行样本聚类分析,并与骨质疏松进行相关性分析。寻找相关性最高的模块的靶基因并绘制网络关系图,随后进行 GO 和 KEGG 富集分析。**结果** 共检测到 9 个 LncRNA 模块,关键模块为 turquoise 模块包含 29 个靶基因,并获得 204 个靶向基因,绘制网络关系图后获得枢纽 LncRNA 为 ACVR2B-AS1、LINC01278、SND1-IT1、C22orf24、ZNF213-AS1、WT1-AS、PLAC4、RHPN1-AS1 和 LINC01140。GO 功能主要集中在细胞间信息传递上;细胞功能集中在细胞核、细胞膜和线粒体中,涉及成骨细胞分化的调控。KEGG 主要富集于 GnRH、Notch、神经营养和钙离子信号通路上。**结论** 利用 WGCNA 方法,获得与骨质疏松性状相关模块,鉴定有生物学意义的基因模块,获得了核心 LncRNA,并通过寻找潜在的靶基因,进行了富集分析。这些结果为 LncRNA 在骨质疏松症病理生理机制的研究中提供新的见解。

关键词: 基因共表达;WGCNA;LncRNA;骨质疏松;靶基因

Analysis of key lncRNA related to osteoporosis based on gene co-expression weight network

LI Yonghua*, TAN Renting, HUANG Shifu, QIU Jinlong

Department of Orthopedics and Traumatology, Danzhou Traditional Chinese Medicine Hospital, Hainan Province, Danzhou 571700, China

* Corresponding author; LI Yonghua, Email: 245698556@qq.com

Abstract: **Objective** To explore the relationship between LncRNA and osteoporosis through gene co-expression weight network analysis (WGCNA). **Methods** Collected 80 samples of GSE56815 chip in GEO database, obtained 127 LncRNA by re-annotation, used WGCNA to analyze the clustering analysis of the expression data, and analyzed the correlation with osteoporosis. Find the target genes of the most relevant modules and draw a network relationship diagram, and then perform GO and KEGG enrichment analysis. **Results** A total of 9 LncRNA modules were detected. The key module was the turquoise module containing 29 target genes, and 204 target genes were obtained. After drawing the network diagram, the hub LncRNA was ACVR2B-AS1, LINC01278, SND1-IT1, C22orf24, ZNF213-AS1, WT1-AS, PLAC4, RHPN1-AS1 and LINC01140. The GO function was mainly focused on the transmission of information between cells; the cell function was concentrated in the nucleus, cell membrane and mitochondria, and was involved in the regulation of osteoblast differentiation. KEGG was mainly enriched in GnRH, Notch, neurotrophic and calcium ion signaling pathways. **Conclusion** The WGCNA method was used to obtain modules related to osteoporosis traits, to identify biologically meaningful gene modules, to obtain core LncRNA, and to perform enrichment analysis by searching for potential target genes. These result provide new insights into the study of LncRNA in the pathophysiology of osteoporosis.

Key words: gene co-expression weights; WGCNA; lncRNA; osteoporosis; target genes

基金项目: 广东省自然科学基金项目(2018A030313369)

* 通信作者: 黎永华, Email: 245698556@qq.com

骨质疏松症是常见的骨代谢疾病。骨质疏松症患者常伴随着脊柱、髋部和腕部骨折的发生及其他骨科围手术期的并发症^[1]。目前的研究重点聚焦在骨质疏松的发病机制,但却忽略了骨质疏松的治疗

手段^[2]。而循环单核细胞作为外周细胞的一种,可分化为巨噬细胞,树突状细胞和破骨细胞等,并作为破骨细胞的前体细胞用于破骨细胞分化^[3]。而破骨分化可以激活和凋亡某些细胞因子(如 IL-1, IL-6),这些相关因子对骨平衡和骨代谢至关重要^[4]。因此,循环单核细胞的功能与骨质疏松症的发病机制密切相关。绝经后卵巢功能下降,会导致雌激素分泌减少,随后抑制 RANKL 蛋白的表达,从而导致单核细胞的破骨分化能力下调,阻碍了骨重塑的进程^[5-6]。LncRNA 是一种长度大于 200 个单位的非编码 RNA,参与多种疾病的发生发展进程。有研究发现,LncRNA 可以通过负反馈的方式反向调节骨髓间充质干细胞(BMCs)的成骨分化,最终导致骨质疏松的形成^[7]。而机械应力和金属颗粒作为常见的外源性因素,可以诱导 LncRNA 的表达出现异常从而抑制成骨细胞功能并促进其凋亡,导致骨质疏松的产生^[8-9]。LncRNA 还能通过 Notch 通路介导转录和免疫因子的表达促进破骨细胞的产生和增加其活性^[10]。

骨质疏松症与单核细胞之间的分子作用机制受众多基因调控^[11],众多基因的影响增加了该领域的研究难度^[12]。而目前,加权基因共表达网络分析(WGCNA)^[13]被广泛报道用于生物基因研究中,而 WGCNA 与既往的差异基因表达分析不同的是,WGCNA 关注基因间的关联,并通过聚类的方式将它们归纳到各类型的模块部分中。建立起临床性状与模块基因的关联,并确定与感兴趣性状相关的模块,之后通过计算获得模块特征值用于选取模块基因。

在研究中,我们通过使用来自 GEO 数据库的 80 个单核细胞样品数据构建骨质疏松的共表达模块,将共表达模块与骨密度性状结合,获得相关模块,寻找与骨密度具有密切关联性的 LncRNA,通过数据库获得 LncRNA 的靶基因,建立靶基因-LncRNA 的互作网络,对这些潜在 LncRNA 的靶基因进行基因的富集分析,分析与模块基因相关的生物过程和相关通路。尝试为骨质疏松在 LncRNA 层面上的研究指明方向

1 材料和方法

1.1 数据获取

表达谱数据来自 GEO 数据库,数据集编号为 GSE56815^[14]。分别获得 40 例正常骨密度和 40 例

低骨密度的绝经前后妇女的基因数据,State 组用以区分绝经前女性和绝经后女性。所有样本均采用 GPL96 [HG-U133 A] Affymetrix Human Genome U133 A Array 平台检测。

1.2 数据处理和重注释

采用 R 软件中的 oligo 软件^[15]通过 RMA 算法对下载的基因数据进行基因校正、标准化和汇总。运用 Ensemble Gene 97 数据库^[16]进行重注释,得到基因类型和基因符号的对应关系,根据基因类型获得 LncRNA。当多个探针对应同一个 LncRNA 时,取其平均值作为 LncRNA 的最终表达值。

1.3 构建共表达模块

该分析用于对同一类基因进行聚类分群。样本聚类分析是通过 R 的 WGCNA^[17]软件对 LncRNA 数据进行处理后获得的。并对样本进行构建聚类树,寻找潜在的离群样本并将其删除。随后根据公式构建软阈值,合适的软阈值需要具有较好的平均连通性,同时符合扑拟合指数曲线并呈现线性关系。随后对 LncRNA 进行共表达模块的构建,模块中的最小基因数设为 3,相关模块的颜色可视化通过 colors 函数进行聚类树呈现。

1.4 识别模块和 BMD 之间的相关性

首先,我们对基因表达与临床信息的线性回归关系中的 P 值进行转换,转换为 $\log_{10}$ 并将其定义为基因显著性 GS(gene significance),将模块成员显著性(MM)定义为模块中所有基因与模块的关联,具有高构建性值的模型被认为与临床特征具有显著相关性。此外,对每个基因模块进行主成分分析。计算模块特征基因和 BMD 之间的皮尔逊相关性系数,以确定与骨质疏松相关的模块。采用 t 检验统计皮尔逊相关性系数,当 P 值 <0.05 时认为模块与 BMD 具有显著相关性。

1.5 网络可视化与 LncRNA 功能预测

选取与临床表型显著相关的基因模块,使用 starbase^[18]和 DIAND-LncBase 数据库预测 LncRNA 的靶向基因(<http://starbase.sysu.edu.cn/index.php>和 http://carolina.imis.athena-innovation.gr/diana_tools/web/)。使用 Cytoscape 网络关系软件进行共表达关系网络图的绘制,同时结合 LncRNA 的靶基因构建共表达网络。在 R 软件上采用 ClusterProfiler^[19]数据对模块基因进行 GO(基因本体论)和 KEGG(京都基因与基因组百科全书)分析。

2 结果

2.1 数据处理及共表达模块的构建

预处理基因数据及临床特征,对重复的基因和缺失值进行剔除,共获得 127 个 LncRNA 的表达矩阵。为确保结果的准确性检测离群样本,检测未见明显离群样本,随后进行样本聚类分析(图 1)。当软阈值等于 4 时,共表达网络接近为无尺度网络,此

时的数值是令曲线趋于平滑的最小阈值,这有利于网络的平均连接程度维持在稳定的状态从而包含足够多的信息(图 2)。选取软阈值为 4 并得到基因聚类树,并根据上述方法进行切割设置。最终得到 9 个 LncRNA 模块,模块中的基因个数从 4 到 29 不等(见图 3)。此外,gray 模块不属于本研究中所定义的模块(图 3)。

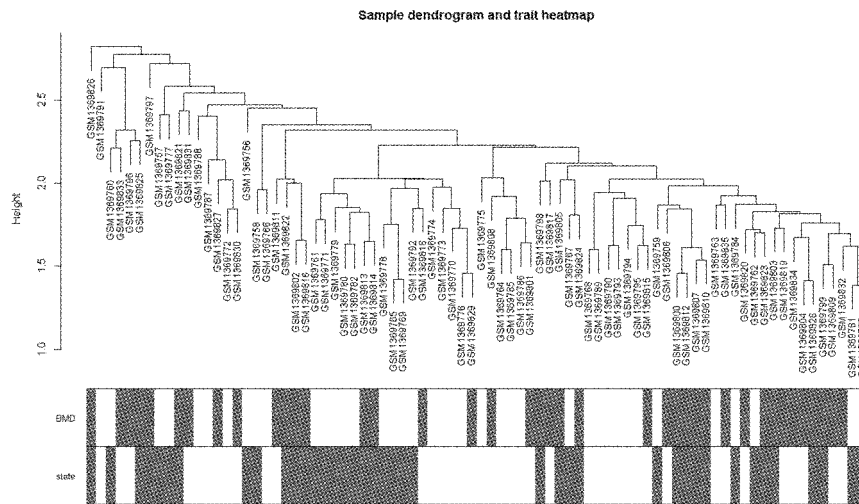


图 1 样品树状图和性状热图

Fig.1 Sample tree diagram and heat map of traits

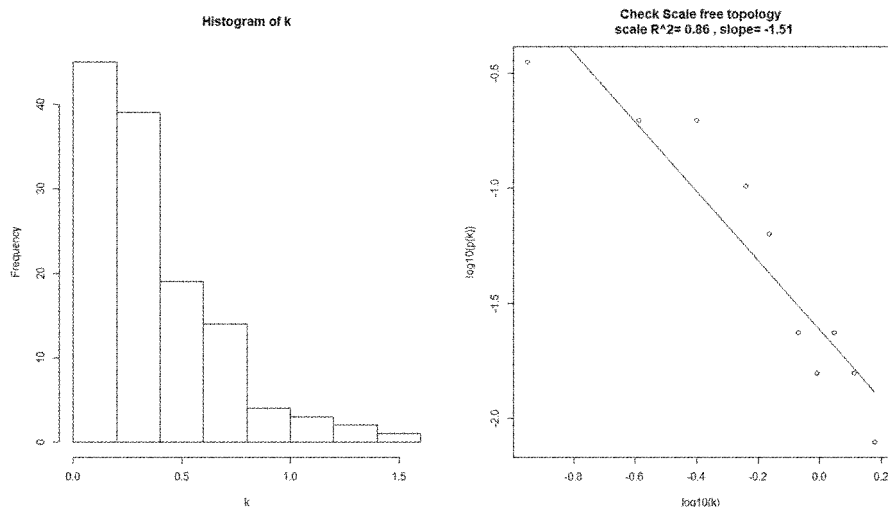


图 2 软阈值权重的网络拓扑矩阵

Fig.2 Network topology matrix with soft threshold weight

2.2 基因模块与临床性状关联分析

将患者的临床信息与模块数据合并后进行相关性分析,寻找特征相关性最高的模块(图 4)。我们发现多个模块与骨质疏松相关,进一步分析发现,

turquoise 模块与骨质疏松呈正相关性($r=0.31, P=0.005$)。针对这个模块,进行相关性分析并绘制散点图分析模块内基因是否符合线性(图 5)。我们根据皮尔逊相关性系数发现这些基因不仅是模块中的

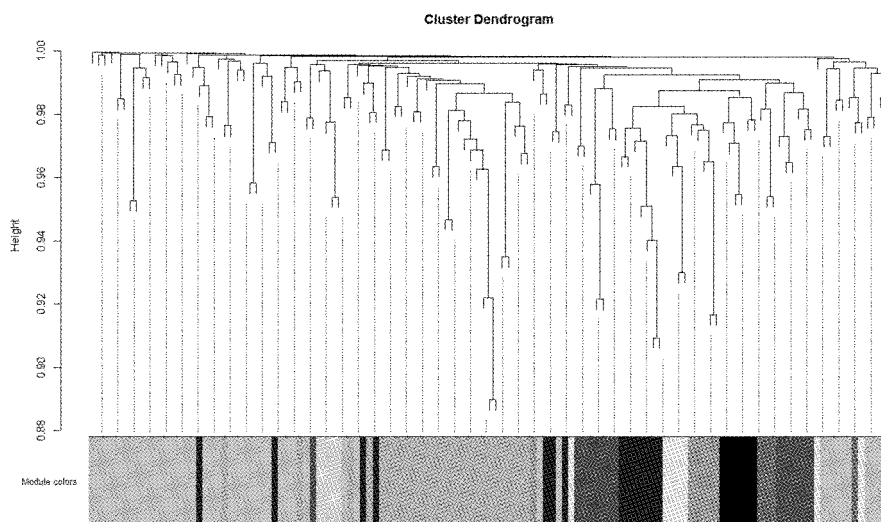
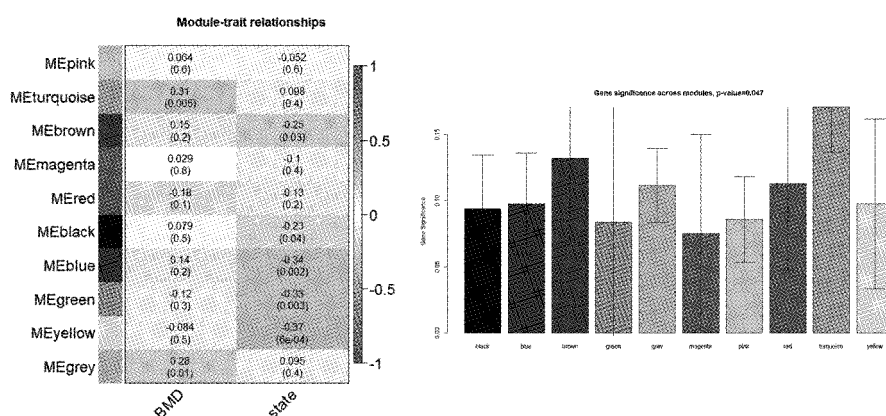


图3 基因共表达网络分层聚类树与共表达模块

Fig.3 Gene co-expression network hierarchical clustering tree and co-expression module



注:左侧为模块基因与性状的相关关系图,每个单元格包含相应的相关性和 P 值,红色为上调,蓝色为下调。右侧为模块的重要性柱状图,条形柱的高度带模块与性状的重要性越高代表越重要。

图4 性状模块相关关系图

Fig.4 Correlation diagram of trait modules

重要基因还与临床性状密切相关。

2.3 获得 LncRNA 靶基因并构建可视化网络

分析模块内基因的相关基因,模块内共有 29 个 LncRNA。登录 STARBASE 和 DIAND-LncBase 网站输入获得的 29 个 LncRNA,最终共获得 204 个靶向基因。将靶基因与 LncRNA 的关联数据导入 Cytoscape 构建靶基因-LncRNA 可视化网络 and 核心基因互作用网络(图 6)。turquoise 模块的枢纽 LncRNA 为 ACVR2B-AS1、LINC01278、SND1-IT1、C22orf24、ZNF213-AS1、WT1-AS、PLAC4、RHPN1-AS1 和 LINC01140。

2.4 GO 和 KEGG 富集分析

对 turquoise 模块中 LncRNA 的靶向的 204 个基

因进行 GO 功能分析和 KEGG 富集分析发现。分子功能主要集中在核糖核酸、RNA 和蛋白质的结合;生物功能主要集中在细胞和蛋白质运输、成骨细胞的分化和神经等调控上;细胞功能主要集中在细胞核、细胞膜和线粒体中。GO 功能主要与细胞物质的转运、蛋白和 RNA 的合成相关,涉及成骨细胞分化的调控。KEGG 主要富集于 GnRH、Notch、神经营养和钙离子信号通路上(图 7)。

3 讨论

骨质疏松症作为常见骨病,可被分为原发性和继发型两种类型,最常见的一种是绝经后骨质疏松症。由于患者成骨和破骨的平衡被打破导致成骨减

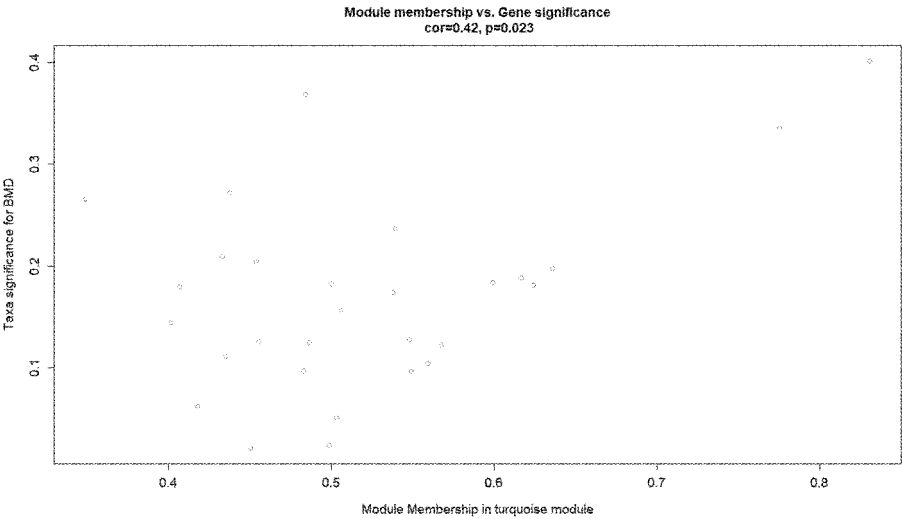


图 5 基因模块表达水平与基因重要性的散点图

Fig.5 Scatter diagram of gene module expression level and gene importance

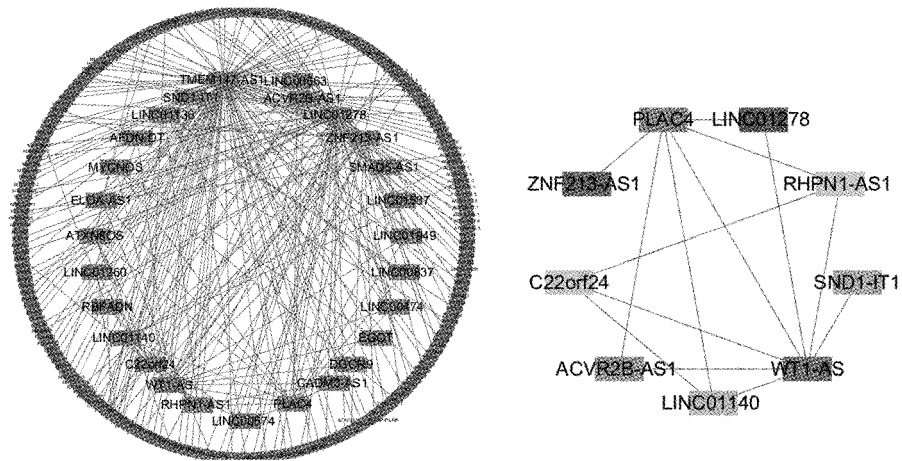


图 6 靶基因-LncRNA 相互作用网络和核心基因互作网络图

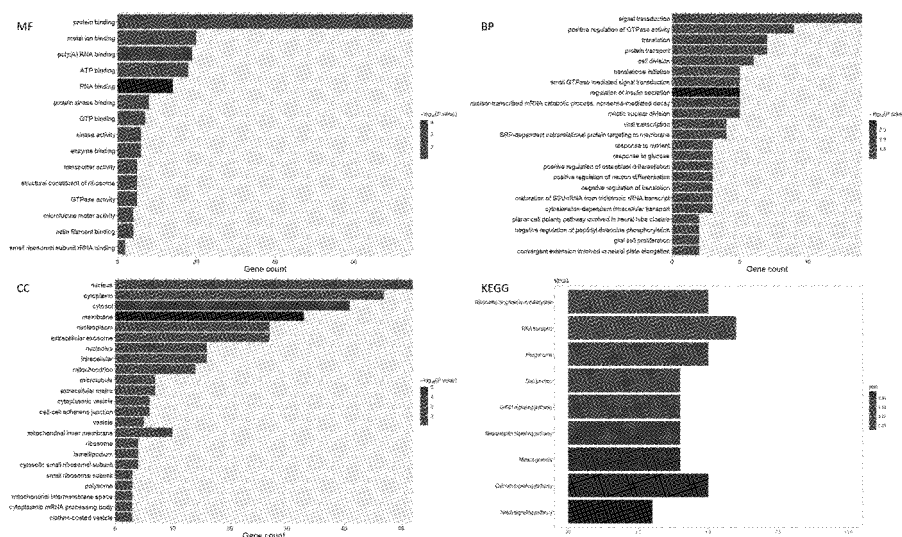
Fig.6 Target gene-LncRNA interaction network and core gene interaction network diagram

少和破骨增多,最终引起骨质疏松症的发生^[20]。临床上,骨质疏松常伴随着骨折的发生^[21]。因此,为了寻找靶向治疗方法,需要深入了解其相关机制及基因调控的运行。既往的研究中,Zhang 等^[22]基于基因组表达数据,通过整合多组学数据找到与患者 BMD 相关的 Mrna 和 Mirna。而本研究则是基于已有的测序数据,进行 WGCNA 分析寻找与骨质疏松相关的 LncRNA。

相较于既往的差异基因表达分类法,WGCNA 不仅对基因进行聚类分群,还将基因与临床性状信息进行关联,有利于发现某些在统计学上无差异但又在人体生物过程中具有重要意义的基因。结合临床特征,综合分析不同临床特征的重要性,有利于寻找不同临床特征在基因层面的相关性,并提高了芯

片数据的利用率,从而获得针对特定形状的关键聚类基因。获得关键模块的基因后,对聚类基因进行 GO 功能注释和 KEGG 富集分析帮助我们更好的了解基因的互作用及潜在通路和机制。

我们得到 9 个有效模块,而结果显示 turquoise 模块与骨质疏松显著相关。提取了 29 个 LncRNA 的相关靶点基因信息,并进行网络可视化和明确枢纽基因。LncRNA 在体内不能翻译为蛋白质,但参与多种细胞过程,由于具有组织特异性,被广泛用于分子靶点和标志物的研究中。许多 LncRNA 仅位于细胞核中,LncRNA 调节着基因的表达,它可以通过多种不同的机制对基因进行干预。虽然本研究在 turquoise 模块中只获得 29 个 LncRNA,但获得其潜在靶基因 204 个,可见 LncRNA 可通过多种机制调



注:MF 为分子功能,BP 为生物过程,CC 为细胞组成,KEGG 为富集信号通路,颜色的变化与 P 值相关,柱图的长度与关联基因的数量呈正相关。

图 7 turquoise 模块的 GO 和 KEGG 富集分析

Fig.7 GO and KEGG enrichment analysis of turquoise module

控基因的表达。

这些 LncRNA 主要参与细胞间信息交流途径,骨形成依赖于骨髓单核细胞供应代谢物,通过协调骨形成和骨吸收的平衡维持正常骨代谢,防止骨量丢失,这些过程都需要细胞之间的信息交换^[23]。例如,细胞外纤调蛋白和 BMP2 通路可以被 LncRNA 激活进而促进 BMCs 的成骨分化,最终抑制成脂分化^[24]。此外,我们的研究也发现 GnRH、Notch、神经营养因子和钙离子信号通路,也参与了 LncRNA 相关的骨形成途径。有研究指出脑垂体分泌的促性腺激素(Gn)刺激破骨细胞的形成和功能调控,其可抑制骨转换和骨量变化,与骨质疏松的发病密切相关^[25]。而钙离子通路的调控,与钙离子的吸收和骨代谢密切相关。神经营养因子通路参与调节骨髓干细胞的成骨分化。Nahlé 的研究发现了神经营养因子通过结合小鼠的 MSCs,激活 STAT1 和 STAT3 的磷酸化,抑制与成骨相关的主要基因的上调,最终阻止成骨细胞的生成和矿化^[26]。值得注意的是,已有研究表明 Notch 通路在骨质疏松中发挥重要作用^[27]。成骨细胞和破骨细胞的动态平衡受到 Notch 通路的调控,成骨细胞的表达会随着 Notch 的敲除而受到抑制^[28]。此外,Canalis 等^[29]的研究发现 Notch 可以抑制成骨细胞的分化。有实验证明了 Notch 与 Runx2 存在负反馈调节作用,Runx2 作为 OPG 的启动抑制剂,当 Notch 被沉默后,Runx2 表达增多,会引起 OPG 的转录被抑制,破骨细胞形成增

加^[30]。这些通路解释了 LncRNA 靶基因与骨质疏松的关联,为骨质疏松的未来的研究提供了新的研究思路。

据我们所知,这是首次将 WGCNA 与骨质疏松相关 LncRNA 相结合的研究。但是,本文仍存在着几点不足。首先,探针重注释的方法可以帮助我们既往的 Geo 芯片中获得可靠的 LncRNA 数据,但由于芯片测序平台的不同以及芯片的更新迭代不同,探针重注释法不能覆盖所有 LncRNA。其次,本研究的结果仍需要通过细胞和分子实验来验证我们的结论,以了解这些检测到的 LncRNA 是否与骨质疏松具有因果关系。再者,本研究的样本量相对较小,需要结合更大样本的数据进行外验证并通过相关的机器学习算法寻找重要性更高的靶标 LncRNA。再者,这些样本的临床数据不完整,无法获得更多相关数据,我们无法获得骨质疏松的血液指标数据和更多临床相关数据,如果有了这部分的数据,我们的结果会更有说服力,结果也会更有意义。

本研究发现这些 LncRNA 与骨质疏松的潜在发病具有密切关系。但尚未检索到关于本研究所发现的核心 LncRNA 与骨质疏松相关文献,因此这些潜在的 LncRNA 靶点既有对骨质疏松发病的预测意义,也为我们未来的骨质疏松研究提供了潜在的靶点。综上所述,我们通过加权基因共表达网络分析方法寻找了的与骨质疏松相关的 LncRNA,获得了

核心 LncRNA, 并通过寻找潜在的靶基因, 进行了富集分析。这些结果为 LncRNA 在骨质疏松症病理生理机制的研究中提供新的见解。

【参 考 文 献】

- [1] Cosman F, de Beur SJ, Leboff MS, et al. Clinician's guide to prevention and treatment of osteoporosis [J]. *Osteoporos Int*, 2014, 25(10): 2359-2381.
- [2] Chen X, Yang L, Ge D, et al. Long non-coding RNA XIST promotes osteoporosis through inhibiting bone marrow mesenchymal stem cell differentiation [J]. *Exp Ther Med*, 2019, 17(1): 803-811.
- [3] Boyle WJ, Simonet WS, Lacey DL. Osteoclast differentiation and activation [J]. *Nature*, 2003, 423(6937): 337-342.
- [4] Auffray C, Sieweke MH, Geissmann F. Blood monocytes: development, heterogeneity, and relationship with dendritic cells [J]. *Annu Rev Immunol*, 2009, 27: 669-692.
- [5] Faienza MF, Ventura A, Marzano F, et al. Postmenopausal osteoporosis: the role of immune system cells [J]. *Clin Dev Immunol*, 2013, 2013: 575936.
- [6] Perrien DS, Achenbach SJ, Bledsoe SE, et al. Bone turnover across the menopause transition: correlations with inhibins and follicle-stimulating hormone [J]. *J Clin Endocrinol Metab*, 2006, 91(5): 1848-1854.
- [7] Wang Q, Li Y, Zhang Y, et al. LncRNA MEG3 inhibited osteogenic differentiation of bone marrow mesenchymal stem cells from postmenopausal osteoporosis by targeting miR-133a-3p [J]. *Biomed Pharmacother*, 2017, 89: 1178-1186.
- [8] Li CJ, Xiao Y, Yang M, et al. Long noncoding RNA Bmncr regulates mesenchymal stem cell fate during skeletal aging [J]. *J Clin Invest*, 2018, 128(12): 5251-5266.
- [9] 刘莉菲, 周绪昌, 邹军. 长链非编码 RNA 对骨质疏松症调控作用的研究进展 [J]. *生命科学*, 2019, 31(11): 1158-1163.
- [10] Wang Y, Luo TB, Liu L, et al. LncRNA LINC00311 promotes the proliferation and differentiation of osteoclasts in osteoporotic rats through the notch signaling pathway by targeting DLL3 [J]. *Cell Physiol Biochem*, 2018, 47(6): 2291-2306.
- [11] Pietschmann P, Grisar J, Thien R, et al. Immune phenotype and intracellular cytokine production of peripheral blood mononuclear cells from postmenopausal patients with osteoporotic fractures [J]. *Exp Gerontol*, 2001, 36(10): 1749-1759.
- [12] 柴毅, 谭峰, 樊巧玲. 用生物信息学分析预测绝经后骨质疏松症核心基因与互作 miRNA 的研究 [J]. *中国骨质疏松杂志*, 2018, 24(10): 1267-1272.
- [13] Pei G, Chen L, Zhang W. WGCNA application to proteomic and metabolomic data analysis [J]. *Methods Enzymol*, 2017, 585: 135-158.
- [14] Zhou Y, Gao Y, Xu C, et al. A novel approach for correction of crosstalk effects in pathway analysis and its application in osteoporosis research [J]. *Sci Rep*, 2018, 8(1): 668.
- [15] Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing [J]. *Bioinformatics*, 2010, 26(19): 2363-2367.
- [16] Alhamdoosh M, Law CW, Tian L, et al. Easy and efficient ensemble gene set testing with EGSEA [J]. *F1000 Res*, 2017, 6: 2010.
- [17] Langfelder P, Horvath S. Fast R functions for robust correlations and hierarchical clustering [J]. *J Stat Softw*, 2012, 46(11): i11.
- [18] Li JH, Liu S, Zhou H, et al. starBase v2.0: decoding miRNA-ncRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data [J]. *Nucleic Acids Res*, 2014, 42(Database issue): D92-D97.
- [19] Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters [J]. *OMICS*, 2012, 16(5): 284-287.
- [20] Liu HY, Wu AT, Tsai CY, et al. The balance between adipogenesis and osteogenesis in bone regeneration by platelet-rich plasma for age-related osteoporosis [J]. *Biomaterials*, 2011, 32(28): 6773-6780.
- [21] Lau EM, Suriwongpaisal P, Lee JK, et al. Risk factors for hip fracture in Asian men and women: the Asian osteoporosis study [J]. *J Bone Miner Res*, 2001, 16(3): 572-580.
- [22] Zhang JG, Tan IJ, Xu C, et al. Integrative analysis of transcriptomic and epigenomic data to reveal regulation patterns for BMD variation [J]. *PLoS One*, 2015, 10(9): e138524.
- [23] Bidwell JP, Alvarez MB, Hood MJ, et al. Functional impairment of bone formation in the pathogenesis of osteoporosis: the bone marrow regenerative competence [J]. *Curr Osteoporos Rep*, 2013, 11(2): 117-125.
- [24] Li CJ, Xiao Y, Yang M, et al. Long noncoding RNA Bmncr regulates mesenchymal stem cell fate during skeletal aging [J]. *J Clin Invest*, 2018, 128(12): 5251-5266.
- [25] 伍西羽, 伍贤平, 廖二元. 促性腺激素与骨代谢 [J]. *中华内分泌代谢杂志*, 2011, 27(3): 269-273.
- [26] Nahle S, Pasquin S, Laplante V, et al. Cardiotrophin-like cytokine (CLCF1) modulates mesenchymal stem cell osteoblastic differentiation [J]. *J Biol Chem*, 2019, 294(32): 11952-11959.
- [27] 张一奇, 刘子云, 付勤. Notch 通路在骨疾病领域的研究进展 [J]. *解剖科学进展*, 2018, 24(4): 411-414.
- [28] Baldridge D, Shchelochkov O, Kelley B, et al. Signaling pathways in human skeletal dysplasias [J]. *Annu Rev Genomics Hum Genet*, 2010, 11: 189-217.
- [29] Canalis E, Parker K, Feng JQ, et al. Osteoblast lineage-specific effects of notch activation in the skeleton [J]. *Endocrinology*, 2013, 154(2): 623-634.
- [30] Liu W, Singh SR, Hou SX. JAK-STAT is restrained by Notch to control cell proliferation of the Drosophila intestinal stem cells [J]. *J Cell Biochem*, 2010, 109(5): 992-999.

(收稿日期: 2020-07-13; 修回日期: 2020-07-27)