

谢鑫峰, 钟梓奇, 王子轶, 等. 基于全基因组关联分析研究文昌鸡初生重性状相关的候选基因 [J]. 畜牧与兽医, 2024, 56 (7): 1-5.
XIE X F, ZHONG Z Q, WANG Z Y, et al. Genome-wide association analysis revealed candidate genes related with birthweight traits in Wenchang chicken [J]. Animal Husbandry & Veterinary Medicine, 2024, 56 (7): 1-5.

基于全基因组关联分析研究文昌鸡初生重性状相关的候选基因

谢鑫峰, 钟梓奇, 王子轶, 程文豪, 郎艺, 严涛, 覃尧*, 肖倩*

(海南大学热带农林学院, 海南 海口 570228)

摘要: 为筛选文昌鸡初生重相关联的候选基因, 本研究采集 235 只文昌鸡初生重表型值, 结合全基因组重测序获得的 12 085 689 个高质量单核苷酸多态性位点 (SNP), 利用 GEMMA 软件进行全基因组关联分析 (GWAS)。结果显示: 与初生重相关的潜在 SNP 位点共有 11 个; 通过基因功能注释, 确定了 6 个可能影响初生重的候选基因, 分别为 cAMP 响应元件结合蛋白 5 (*CREB5*)、锯齿状 Notch 配体 1 (*JAG1*)、蓝带 WH2 重复蛋白 (*COBL*)、WD 重复结构域 11 (*WDR11*)、FIG4 磷酸肌酸 5-磷酸酶 (*FIG4*) 和跨膜结构域蛋白 (*VSTM2A*); 进一步富集分析发现, ATP 和钙离子结合、钙离子信号通路和 MAPK 信号通路在初生重中发挥重要作用。本研究结果为文昌鸡初生重相关性状提供了重要的候选分析标记, 为进一步选育优良品种提供理论支持。

关键词: 全基因组关联分析; 单核苷酸多态性; 文昌鸡; 初生重; 生长性状

中图分类号: S831.2 文献标志码: A 文章编号: 0529-5130(2024)07-0001-05

Genome-wide association analysis revealed candidate genes related with birthweight traits in Wenchang chicken

XIE Xinfeng, ZHONG Ziqi, WANG Ziyi, CHENG Wenhao,

LANG Yi, YAN Tao, QIN Yao*, XIAO Qian*

(College of Tropical Agriculture and Forestry, Hainan University, Haikou 570228, China)

Abstract: To identify candidate genes associated with the birth weight of Wenchang chickens, this study collected phenotypic data on birth weight from 235 individuals of the Wenchang chicken breed. Then, blood DNA was collected from these Wenchang chickens, and 12 085 689 high quality single nucleotide polymorphisms (SNPs) were obtained by whole genome resequencing. Finally, GEMMA software was used to perform genome-wide association study (GWAS) on the SNPs. The results showed that a total of 11 SNPs were found to be associated with the birth weight of the chickens. Six candidate genes (*CREB5*, *JAG1*, *COBL*, *WDR11*, *FIG4*, and *VSTM2A*) affecting the birthweight were identified by gene function annotation. The functional enrichment analysis showed that ATP/calcium binding, the calcium signaling pathway and the MAPK signaling pathway played important roles in the birthweight of the birds. The results of this study provided important candidate analytical markers for growth traits in Wenchang chickens and offered theoretical support for further superior breeding selection.

Keywords: genome-wide association study; single nucleotide polymorphisms; Wenchang chicken; birthweight; growth traits

我国是一个禽类消费大国, 商品鸡的大规模养殖满足了人们对鸡肉的基本需求。地方鸡种肉质鲜美, 有独特风味, 深受消费者喜爱。文昌鸡产于海南省文昌市, 以皮薄骨酥、肉质香甜嫩滑、营养俱佳等特点

被消费者所喜爱^[1]。然而, 在相同的饲养条件下, 地方鸡种的生长速度低于商业品种, 导致其价格偏高^[2]。如何提高地方品种生长性状是畜禽遗传学研究的重点, 而初生重是生长性状的一个重要指标。因此, 研究文昌鸡初生重相关的候选基因, 可为文昌鸡后期的选育提供一定基础。高通量测序技术和单核苷酸多态性位点 (single nucleotide polymorphism, SNP) 芯片的发展, 为遗传育种提供了新的技术手段, 越来越多的研究利用全基因组关联分析 (genome-wide association study, GWAS) 进行动物遗传改良, 期望能够选出具有优良表型的个体, 加速育种进程。张笑科

收稿日期: 2023-06-26; 修回日期: 2024-05-04

基金项目: 海南省自然科学基金高层次人才项目 (2019RC034)

第一作者: 谢鑫峰, 男, 硕士研究生

* 通信作者: 覃尧, 博士, 讲师, 硕士生导师, 研究方向为物分子病毒学与免疫学, E-mail: qinyao@hainanu.edu.cn; 肖倩, 博士, 副教授, 硕士生导师, 研究方向为动物遗传育种, E-mail: xiaoqian@hainanu.edu.cn。

等^[3]通过 SNP 芯片鉴定与杜洛克猪生长性状相关的候选基因；朱迪等^[4]利用全基因组关联分析鉴定太行鸡产蛋性状功能基因等。因此，本研究选取海南省特色地方品种文昌鸡，并针对初生重进行全基因组关联分析，鉴定影响初生重性状的分子标记，为文昌鸡在生长性状上的选育提供理论基础。

1 材料与方法

1.1 样本及表型数据收集

共收集 235 只育雏阶段（1 日龄）文昌鸡，并测定每个个体的初生重。

1.2 基因型数据的获取

采取育雏阶段（1 日龄）文昌鸡血液样本 235 份（2 mL/份），EDTA 抗凝处理后低温 12 h 内运至实验室，-80 °C 冻存备用。采用 DNA 提取试剂盒 [天根生化科技（北京）有限公司] 并通过琼脂糖（北京博奥拓达科技有限公司）凝胶电泳检测和分光光度计验证 DNA 的质量和完整性。以 cDNA > 50 ng/mL， OD_{260}/OD_{280} 值在 1.7 ~ 2.0 之间，电泳条带清晰，无杂带和拖尾情况判断为合格。将检测合格的 DNA 送往上海派森诺生物科技有限公司构建重测序文库，并采用 Illumina NovaSeq（PE150，Illumina）平台测序。采用 fastp 软件（<https://github.com/OpenGene/fastp>）对所有个体的原始数据进行质控过滤，参数设置为默认参数，利用 BWA（v0.7.17）软件^[5]对过滤后的测序数据与参考基因组（bGalGal1.mat.broiler.GRCg7b）进行比对。分别利用 GATK（v4.1.6.0）软件^[6]和 STITCH 软件^[7]进行全基因组水平的 SNP 检测及缺失基因型的填补。为了确保本研究能够获得高可信度的 SNP，需要对获得的 SNP 再次进行过滤，过滤标准：最小等位基因频率（MAF） ≥ 0.05 ，缺失率 ≤ 0.5 ，平均测序深度 $\geq 5\times$ ，SNP 检出率 $\geq 90\%$ 。过滤后得到的 SNP 可用于下一步数据分析。

1.3 群体结构分析

GWAS 分析中群体分层和亲缘关系会引起结果的假阳性。本研究使用主成分分析（principal component analysis, PCA）对研究样本的群体结构进行分析，同时使用 GCTA 软件^[8]来评估群体内的亲缘关系，并用 R 语言对结果进行可视化。

1.4 全基因组关联分析

针对文昌鸡初生重性状，采用 GEMMA 软件中的

混合线性模型进行全基因组关联分析。模型如下：

$$y = W\beta + G\gamma + g + e,$$

式中： y 为表型向量，代表初生重； $W\beta$ 为固定效应，包括群体结构和性别； G 为基因型向量， γ 是基因型效应值， $g \sim N(0, K\Phi^2)$ 为多基因效应， $e \sim N(0, I\sigma^2)$ 为残差效应，多基因效应中的 K 为标记推断的亲缘关系矩阵。

全基因组关联分析的结果用控制错误发现率法（false discovery rate, FDR）进行多重检验，计算公式为：

$$P = FDR \times n/m,$$

其中： n 为分析结果中 P 值低于 FDR 的个数， m 为鉴定到的 SNP 个数。将显著性 P 值 $< 0.001 \times n/m$ 的位点认为是显著性关联位点，显著性 P 值 $< 0.05 \times n/m$ 的位点认为是潜在显著关联位点，同时将潜在显著关联位点用于后续的通路富集分析。

1.5 显著位点基因功能注释与通路分析

从 NCBI 网站上下载对应物种的参考基因组信息，使用 ANNOVAR 软件将显著 SNP 注释到其对应的基因上，搜集潜在关联位点上下游 50 kb 内的基因并结合文献注释候选基因的功能，将潜在显著性位点注释后，使用 David 软件对得到的所有基因进行 GO 和 KEGG 的通路富集分析。

2 结果

2.1 初生重性状统计

对雏鸡（1 日龄）初生重进行统计，平均重量为 (28.14 ± 2.40) g。

2.2 全基因组 SNP 位点标记鉴定

通过对全基因组水平的 SNP 检测及缺失基因型的填补，质控后得到 12 085 689 个有效 SNP。这些位点在基因组上各染色体上分布如图 1 所示。

2.3 群体结构和聚类分析

由图 2 得知，文昌鸡群体内存在一定的群体分层和亲缘关系。为了减少群体分层和亲缘关系给 GWAS 分析结果带来的假阳性，将前 3 列 PCA 值当作协变量加入到混合线性模型当中；亲缘关系视为随机效应，在分析过程中添加亲缘关系矩阵用来描述个体之间的相关性从而更准确地估计 SNP 与表型之间的关联，避免群体分层和亲缘关系对结果的影响。



图1 全基因组 SNP 密度

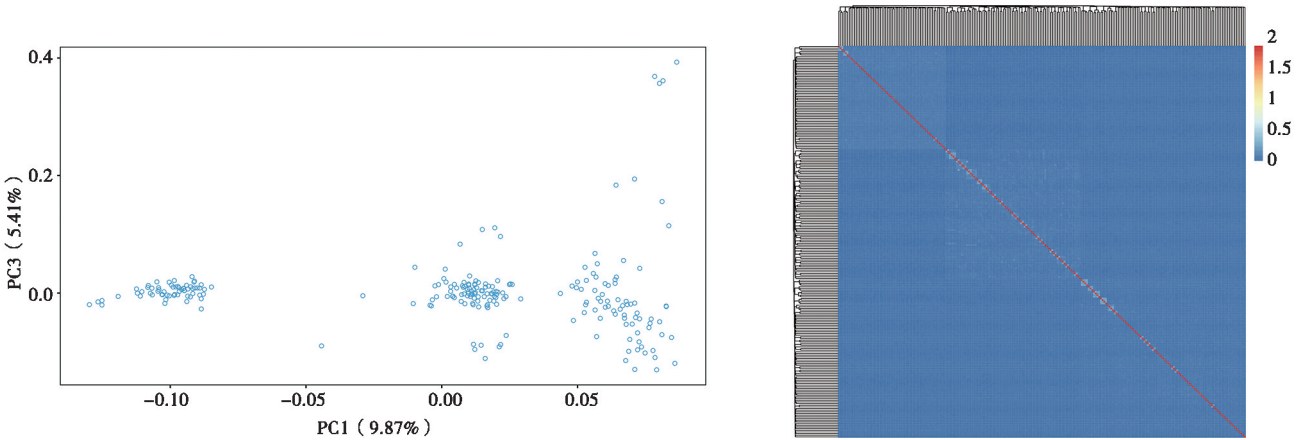


图2 文昌鸡群体结构 (左) 和聚类分析 (右)

2.4 GWAS 分析

使用 GEMMA 软件中的混合线性模型进行全基因组关联分析, 图 3 展示了生长性状的全基因组关联分析结果。通过 FDR 校正得到显著性阈值 ($P = 1 \times 10^{-6}$), 发现了 11 个与初生重相关联的显著性 SNP 位点, 这些位点分别邻近或坐落于 11 个基因, 分别是: cAMP 响应元件结合蛋白 5 (*CREB5*)、蓝带

WH2 重复蛋白 (*COBL*)、锯齿状 Notch 配体 1 (*JAG1*)、SLX4 相互作用蛋白 (*SLX4IP*)、FIG4 磷酸肌酸 5-磷酸酶 (*FIG4*)、VRK 丝氨酸/苏氨酸激酶 1 (*VPK1*)、含有 2A 的 V 型和跨膜结构域蛋白 (*VSTM2A*)、BCL11 转录因子 B (*BCL11B*)、6 型 SH2 结构域 (*SH2D6*)、DNA 核苷酸外转移酶 (*DNTT*) 以及 WD 重复结构域 11 (*WDR11*) (表 1)。

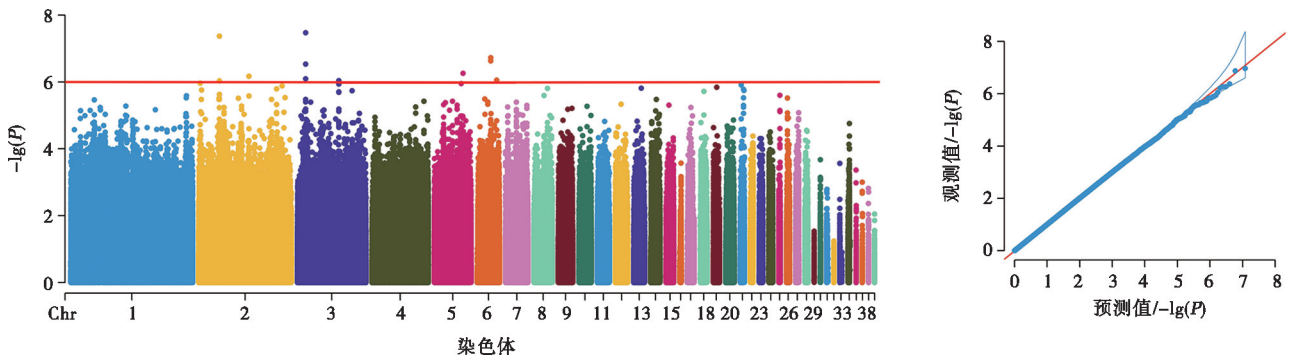


图3 文昌鸡初生重关联分析结果曼哈顿图 (左) 和 QQ 图 (右)

表 1 文昌鸡初生重显著关联 SNP 位点

染色体	位置/bp	候选基因	P 值
2	33 355 914	<i>CREB5</i>	4.26×10^{-8}
2	80 840 471	<i>COBL, VSTM2A</i>	6.75×10^{-7}
2	33 356 531	<i>CREB5</i>	9.24×10^{-7}
3	13 225 975	<i>JAG1</i>	3.39×10^{-8}
3	13 225 444	<i>JAG1</i>	2.93×10^{-7}
3	13 229 858	<i>JAG1, SLX4IP</i>	8.01×10^{-7}
3	66 505 947	<i>FIG4</i>	9.16×10^{-7}
5	46 277 690	<i>VRK1, BCL11B</i>	5.53×10^{-7}
6	21 920 057	<i>SH2D6, DNNT</i>	1.87×10^{-7}
6	21 968 402	<i>SH2D6, DNNT</i>	2.33×10^{-7}
6	31 464 170	<i>WDR11</i>	8.82×10^{-7}

2.5 候选 SNP 位点的通路分析

全基因组关联分析结果中,对显著性 P 值小于 2.64×10^{-3} 的 SNP 位点进行筛选注释后共得到 475 个潜在显著候选基因,使用 David 软件对这些基因进行 GO 和 KEGG 通路富集分析,结果如图 4 所示,共得到 45 个显著的 GO 富集项和 7 个显著的 KEGG 通路。GO 富集项中,分子功能 (MF) 占 29%,生物过程 (BP) 占 49%,细胞成分 (CC) 占 22%。在 GO 富集项中,初生重潜在关联基因主要富集在 ATP 结合 (GO: 0005524~ATP binding) 和钙离子结合 (GO: 0005509~calcium ion binding) 相关通路。在 KEGG 的通路富集中,初生重性状主要富集在 MAPK 信号通路 (GGA04010: MAPK signaling pathway) 和钙离子信号通路 (GGA04020: Calcium signaling pathway)。

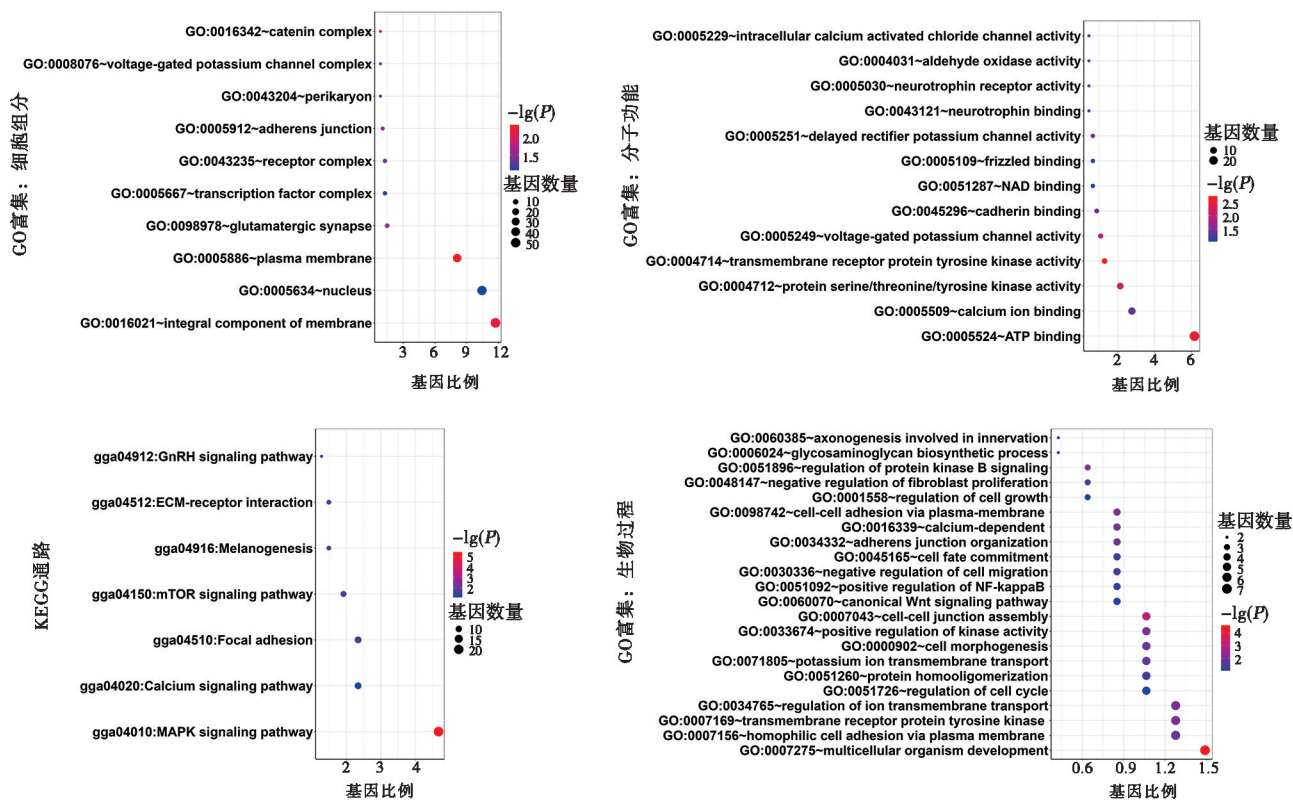


图 4 初生重 GO 和 KEGG 通路富集分析结果

3 讨论

文昌鸡是海南省当地的知名品种,但其生长速度低于商业肉鸡品种,影响文昌鸡产业的发展^[9]。因此如何提高文昌鸡的生长速度,对发展文昌鸡产业、提高市场竞争力有重要意义。

在鸡生产中,GWAS 常被用来鉴定与生产相关性状的候选基因^[10]。本研究对文昌鸡的初生重进行 GWAS 分析,对研究中发现的 11 个显著性基因进行

基因功能注释,发现了 6 个最有可能影响初生重的候选基因 (*CREB5*, *JAG1*, *COBL*, *WDR11*, *FIG4*, *VSTM2A*)。*CREB5* 是诱导关节软骨中 *prg4* (蛋白质润滑素) 表达的关键因子,同时通过驱动信号分子的表达来促进关节形成和关节软骨细胞的后续发育^[11-12]。*JAG1* 主要在高度保守的 Notch 信号通路中起作用的 5 种细胞表面配体之一。Notch 信号在细胞命运决定中起着关键作用,并且在整个发育过程中和许多器官系统中都很活跃。*JAG1* 的功能突变丧失将

引起动脉-肝脏发育不良综合征 (Alagille 综合征), 主要影响肝脏、心脏、骨骼、眼睛、面部、肾脏和脉管系统^[13]。*COBL* 最初被鉴定为在胚胎发育过程中调节脊椎动物轴形成的基因^[14]。*COBL* 编码的蛋白质参与肌动蛋白组装的调节^[15]。*VSTM2A* 编码一种新型蛋白, 能够促进脂肪生成。研究表明 *VSTM2A* 通过调节 BMP 信号传导和 *Ppar γ 2* 表达来控制脂肪生成, *VSTM2A* 过度表达可以增强脂肪生成, 而缺失则会影 响脂肪生成^[16]。*WDR11* 基因中的杂合错义变异与性腺功能减退症有关。*WDR11* 缺陷小鼠会出现垂体畸形发生和生长迟缓, 以及心脏缺陷, 生殖器官不育和发育不全的发育缺陷^[17]。*FIG4* 编码磷脂酰肌醇磷酸盐 (PIP) 的磷酸酶, 两者在骨骼肌中发挥重要的功能, *FIG4* 缺失将会损害肌肉功能^[18]。这些基因的综合作用决定了文昌鸡的初生重, 推测这些基因可能是决定文昌鸡生长性状的候选基因。另一方面, 通路富集分析发现, 文昌鸡初生重性状的潜在候选基因主要富集到 ATP 结合、钙离子结合、钙离子信号通路和 MAPK 信号通路。目前已经有大量研究证实 MAPK 信号通路在骨骼生长发育中发挥重要作用^[19-21], 因此推测这些通路与生长性状密切相关。

4 结论

本研究利用 235 个文昌鸡样本的重测序数据, 鉴定到 12 085 689 个全基因组高密度遗传标记; 利用上述标记对文昌鸡初生重进行全基因组关联分析, 发现影响初生重的 6 个候选基因 (*CREB5*, *JAG1*, *COBL*, *WDR11*, *FIG4*, *VSTM2A*); 通过 GO 和 KEGG 分析发现, ATP 结合、钙离子结合、钙离子信号通路和 MAPK 通路在影响文昌鸡初生重性状上发挥重要作用。上述结果为文昌鸡生长性状选育提供了重要的分子标记, 为下一步的保种育种工作奠定了理论基础。

参考文献:

- [1] 符安, 顾丽红, 陈益勇, 等. 文昌鸡概况及全产业链分析 [J]. 中国畜禽种业, 2022, 18 (12): 30-33.
- [2] ALAM M, ULLAH M O, MALIK S U F, et al. Broiler and indigenous chickens: a comparison through biochemical parameters [J]. International Journal of Sustainable Agricultural Research, 2020, 7 (4): 228-233.
- [3] 张笑科, 廖伟莉, 陈信佑, 等. 杜洛克猪生长性状全基因组关联分析及候选基因鉴定 [J]. 畜牧兽医学报, 2023, 54 (5): 1868-1876.
- [4] 朱迪, 谢艳华, 王宇哲, 等. 全基因组关联分析鉴定太行鸡产蛋性状功能基因 [J]. 中国畜牧杂志, 2023, 59 (9): 224-230.
- [5] LI H, DURBIN R. Fast and accurate short read alignment with Burrows-Wheeler transform [J]. Bioinformatics, 2009, 25 (14): 1754-1760.
- [6] MCKENNA A, HANNA M, BANKS E, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data [J]. Genome Res, 2010, 20 (9): 1297-1303.
- [7] DAVIES R W, FLINT J, MYERS S, et al. Rapid genotype imputation from sequence without reference panels [J]. Nat Genet, 2016, 48 (8): 965-969.
- [8] YANG J, LEE S H, GODDARD M E, et al. GCTA: a tool for genome-wide complex trait analysis [J]. Am J Hum Genet, 2011, 88 (1): 76-82.
- [9] TIAN S, LI W, ZHONG Z, et al. Genome-wide re-sequencing data reveals the genetic diversity and population structure of Wenchang chicken in China [J]. Anim Genet, 2023, 54 (3): 328-337.
- [10] 张文武, 邱峰芳, 刘影, 等. 白耳黄鸡冠齿数性状全基因组关联分析 [J]. 中国家禽, 2022, 44 (8): 16-20.
- [11] ZHANG C H, GAO Y, JADHAV U, et al. Creb5 establishes the competence for Prg4 expression in articular cartilage [J]. Commun Biol, 2021, 4 (1): 322.
- [12] ZHANG C H, GAO Y, HUNG H H, et al. Creb5 coordinates synovial joint formation with the genesis of articular cartilage [J]. Nat Commun, 2022, 13 (1): 7295.
- [13] GROCHOWSKI C M, LOOMES K M, SPINNER N B. Jagged1 (*JAG1*): structure, expression, and disease associations [J]. Gene, 2016, 576 (1 Pt 3): 381-384.
- [14] GASCA S, HILL D P, KLINGENSMITH J, et al. Characterization of a gene trap insertion into a novel gene, cordon-bleu, expressed in axial structures of the gastrulating mouse embryo [J]. Dev Genet, 1995, 17 (2): 141-154.
- [15] HUSSON C, RENAULT L, DIDRY D, et al. Cordon-Bleu uses WH2 domains as multifunctional dynamizers of actin filament assembly [J]. Mol Cell, 2011, 43 (3): 464-477.
- [16] SECCO B, CAMIRÉ É, BRIÈRE M A, et al. Amplification of adipogenic commitment by *VSTM2A* [J]. Cell Rep, 2017, 18 (1): 93-106.
- [17] HAAG N, TAN E C, BEGEMANN M, et al. Biallelic loss-of-function variants in *WDR11* are associated with microcephaly and intellectual disability [J]. Eur J Hum Genet, 2021, 29 (11): 1663-1668.
- [18] REIFLER A, LENK G M, LI X, et al. Murine Fig4 is dispensable for muscle development but required for muscle function [J]. Skelet Muscle, 2013, 3 (1): 21.
- [19] YU X, QUAN J, LONG W, et al. LL-37 inhibits LPS-induced inflammation and stimulates the osteogenic differentiation of BMSCs via P2X7 receptor and MAPK signaling pathway [J]. Exp Cell Res, 2018, 372 (2): 178-187.
- [20] ZHAO P, XIAO L, PENG J, et al. Exosomes derived from bone marrow mesenchymal stem cells improve osteoporosis through promoting osteoblast proliferation via MAPK pathway [J]. Eur Rev Med Pharmacol Sci, 2018, 22 (12): 3962-3970.
- [21] CHEN G, DENG C, LI Y P. TGF- β and BMP signaling in osteoblast differentiation and bone formation [J]. Int J Biol Sci, 2012, 8 (2): 272-288.