

doi: 10.13241/j.cnki.pmb.2021.12.009

多组学大数据整合分析揭示早期胚胎发育过程中基因表达调控信息的变化规律*

江琦¹ 汤小晗² 陶欢³ 王军婷² 白雪梅² 李昊^{3Δ} 卢美松^{2Δ}

(1 同方股份有限公司 北京 100083; 2 哈尔滨医科大学附属第一医院 黑龙江 哈尔滨 150081;

3 军事科学院军事医学研究院辐射医学研究所 北京 100850)

摘要 目的:探究哺乳动物早期胚胎发育过程中基因表达调控信息的变化规律。**方法:**收集小鼠早期胚胎发育各时期的 RNA-seq, ATAC-seq, MethylC-Seq 和 H3K4me3 ChIP-seq 数据进行整合分析,观察小鼠早期胚胎发育各时期转录因子表达量的变化,计算各时期基因表达量与转录因子结合位点数量及染色质可及性的相关性,筛选各时期表达量前 10% 的基因,统计其表达量和转录因子占比,并进行启动子可及性分析。根据前期报道的转录因子三节点调控网络,对早期胚胎各时期转录因子调控网络的富集模式进行分析。根据多组学数据分析结果,推测早期胚胎发育调控过程中转录因子和表观遗传修饰信息的共调控模型。**结果:**转录因子数量和调控关系变化以及染色质可及性、DNA 甲基化修饰、组蛋白修饰等表观遗传修饰共同调控早期胚胎发育各时期的基因表达,这些因素在不同时期发挥不同程度的调控作用。**结论:**转录因子和表观遗传修饰在早期胚胎发育过程中动态调控基因表达。

关键词:多组学整合分析;早期胚胎发育;转录因子;表观遗传修饰;基因表达调控

中图分类号:R-33;Q132.7;Q344 **文献标识码:**A **文章编号:**1673-6273(2021)12-2241-04

Integrative Multi-omics Analysis of Gene Regulation Mechanism during Early Embryonic Development*

JIANG Qi¹, TANG Xiao-han², TAO Huan³, WANG Jun-ting², BAI Xue-me², LI Hao^{3Δ}, LU Mei-song^{2Δ}

(1 Tongfang Co., Ltd. Beijing, 100083, China; 2 The First Clinical Hospital affiliated to Harbin Medical University, Harbin,

Heilongjiang, 150081, China; 3 Institute of Radiation Medicine, Academy of Military Medical Sciences, Academy of Military Sciences, Beijing, 100850, China)

ABSTRACT Objective: Exploration of gene regulation mechanism during early mammalian embryonic development. **Methods:** RNA-seq, ATAC-seq, MethylC-Seq and H3K4me3 ChIP-seq data of mouse early embryonic development were collected for integrative analysis. Expression change of transcription factors during mouse early embryonic development were counted. Correlation between gene expression and number of transcription factor binding sites and chromatin accessibility were calculated. Highly expressed (top 10%) genes in different stages of early embryonic development were filtered in which proportion of transcription factors and chromatin accessibility in promoter regions were calculated. According to three-node regulatory network of transcription factors reported previously, the enrichment mode of transcription factors regulatory network in different stages of early embryonic development were analyzed. Besides, co-regulation model of transcription factors and epigenetic modification of early embryonic development was simulated based on integrative analysis of multi-omics data. **Results:** Number and regulatory mode of transcription factors and epigenetic modification, such as chromatin accessibility, DNA methylation, histone modification, co-regulate gene expression during early embryonic development. These factors show different regulatory mode in different stages of early embryonic development. **Conclusions:** Transcription factors and epigenetic modification play highly dynamic regulatory roles in gene expression during early embryonic development.

Key words: Multi-omics analysis; Early embryonic development; Transcription factor; Epigenetic modification; Gene regulation

Chinese Library Classification(CLC): R-33; Q132.7; Q344 **Document code:** A

Article ID: 1673-6273(2021)12-2241-04

* 基金项目:北京市自然科学基金项目(5204040)

作者简介:江琦(1988-),男,主要研究方向:信息工程,E-mail: leojiang@unitedstack.com

Δ 通讯作者:卢美松(1964-),女,妇产科教研室主任,生殖医学科主任,妇科腔镜中心主任,主要研究方向:妇科内镜、妇科肿瘤与生育力保护,E-mail: lumsong0417@163.com,电话:13703605008;

李昊(1989-),男,助理研究员,主要研究方向:生物组学数据的人工智能算法开发,E-mail: lihao_thu@163.com,电话:18689600527

(收稿日期:2021-03-08 接受日期:2021-03-31)

前言

精子与卵子结合,将两个终末分化的生殖细胞转换为全能的受精卵,伴随细胞分裂和卵母细胞激活(oocyte activation),母源-胚胎转换(maternal-to-zygotic transition,MZT),合子基因组激活(zygotic gene activation,ZGA),细胞命运决定和谱系分化等重要生物过程,早期胚胎发育逐渐完成^[1-6]。早期胚胎发育过程中基因表达的调控是高度复杂且动态有序的,支撑着早期胚胎发育各时期重要生物过程的过度度和转换,其中转录因子(transcription factor,TF)和表观遗传信息(如染色质可及性,DNA 甲基化,组蛋白修饰等)是早期胚胎发育基因表达调控的重要因素^[7-19]。本文整合分析了多组学数据以探究转录因子和表观遗传修饰在早期胚胎发育各时期调控基因表达的变化规律,并模拟了转录因子及表观遗传修饰共同调控早期胚胎发育基因表达的模式图。

1 材料与方法

1.1 材料

小鼠早期胚胎发育各时期的 RNA-seq 数据来自 GSE66582;小鼠早期胚胎发育各时期的 ATAC-seq 数据来自 GSE66390;小鼠早期胚胎发育各时期的 DNA 甲基化数据来自 GSE56697;小鼠早期胚胎发育各时期的组蛋白修饰数据来自 GSE71434;小鼠基因组参考文件为 mm9;转录因子注释文件来自 HOCOMOCO (v10)数据库。

1.2 方法

1.2.1 ATAC-seq 数据分析 ATAC-seq 数据的质控通过 FastQC 完成,比对通过 bowtie2 完成,比对后的质控通过 samtools 完成,转录因子的扫描通过 FIMO 完成,求 peak 通过 MACS2 完成。

1.2.2 RNA-seq 数据分析 RNA-seq 数据的质控通过 FastQC 完成,比对通过 bowtie2 完成,比对后通过 HTSeq-count 进行转

录本定量并获得表达矩阵。

1.2.3 ChIP-seq 数据分析 ChIP-seq 数据的质控通过 FastQC 完成,比对通过 bowtie2 完成,比对后的质控通过 samtools 完成,求 peak 通过 MACS2 完成。

1.2.4 转录因子调控网络模式分析 参考 Neph 等人使用的转录因子调控网络模式分析方法,我们通过 FIMO 扫描了转录因子结合位点(transcription number binding site,TFBS),并在此基础上通过 mfinder 软件求得了 13 种转录因子调控网络的富集情况(z-score),分别获得了早期胚胎发育各时期转录因子调控网络模式图。

1.3 TFBS 统计学方法

基因表达量与 TFBS 数量及染色质可及性的相关性分析:首先通过 ATAC-seq 数据分析得到各基因上游 2kb 区域内的 TFBS 个数以及染色质可及性信息(fragments per kilobase of exon model per million mapped read,FPKM),并计算它们与基因表达量的皮尔森相关性(R^2)。

2 结果

2.1 早期胚胎发育过程中转录因子编码基因的表达量变化

早期胚胎发育各时期转录因子编码基因的表达量变化分析显示各时期转录因子的表达量经历降低升高再下降的过程。早期胚胎发育至 2 细胞时期,成熟卵母细胞中遗留的转录因子逐渐被消耗,2 细胞时期之后胚胎自身转录因子生成。见图 1。

2.2 早期胚胎发育各时期基因表达量与转录因子结合位点数量及染色质可及性的相关性

对早期胚胎发育各时期基因表达量与 TFBS 数量及染色质可及程度的相关性进行分析,结果显示各时期的基因表达量与 TFBS 数量及染色质可及性的相关性呈现先上升后下降的趋势,相关性在 4 细胞时期和 8 细胞时期达到峰值($R^2=0.9$),而在 2 细胞早期和 mESC 时期相关性很低($R^2<0.1$)。见图 2。

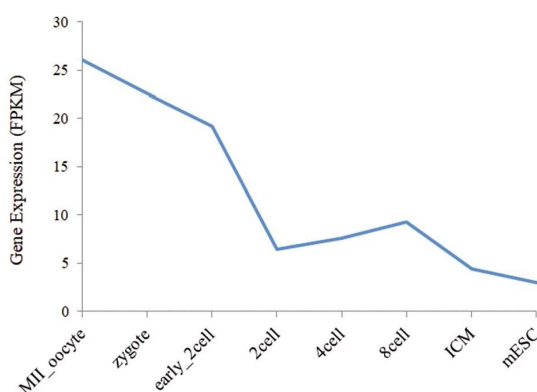


图 1 早期胚胎发育各时期转录因子表达量

Fig.1 TF expression during early embryonic development

2.3 转录因子和染色质可及性对早期胚胎发育各时期基因表达的调控规律

选取各时期中高表达的基因(表达量在当前时期中处于前 10%的基因)为研究对象,由基因表达热图可见,4 细胞时期前高表达基因中转录因子编码基因的表达量比例明显高于 4 细胞及之后(绿色部分热图),而 4 细胞时期前的高表达基因,其

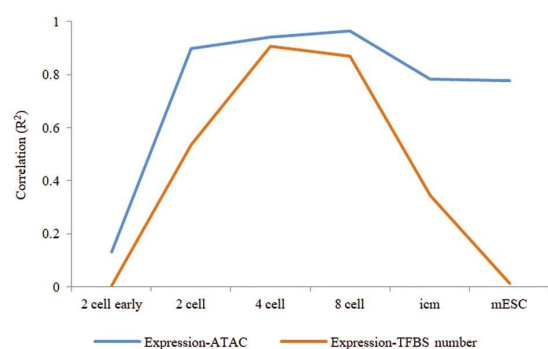


图 2 早期胚胎发育各时期基因表达量与转录因子结合位点数量(橘色)及染色质可及性(蓝色)的相关性

Fig.2 Correlation between gene expression and TFBS number (orange) and chromatin accessibility (blue) during early embryonic development

Note: The number of transcription factor binding sites and chromatin accessibility (FPKM) were calculated in the 2kb region upstream of each gene, separately.

表达量从卵子时期开始就较高。此外,卵母细胞、合子及 2 细胞早期,高表达基因的染色质启动子区的可及性低于其他时期。见图 3。

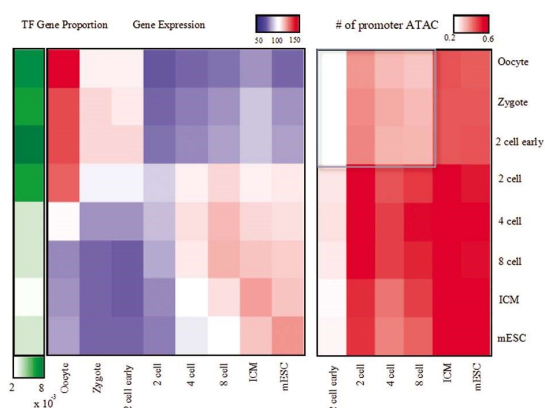


图3 早期胚胎发育各时期高表达基因(前 10%)TF 编码基因比例、表达量及启动子区可及性热图

Fig.3 Heatmaps of gene expression (top 10%), proportion of TF expression and promoter accessibility of early embryonic development

2.4 早期胚胎发育各时期转录因子调控网络模式

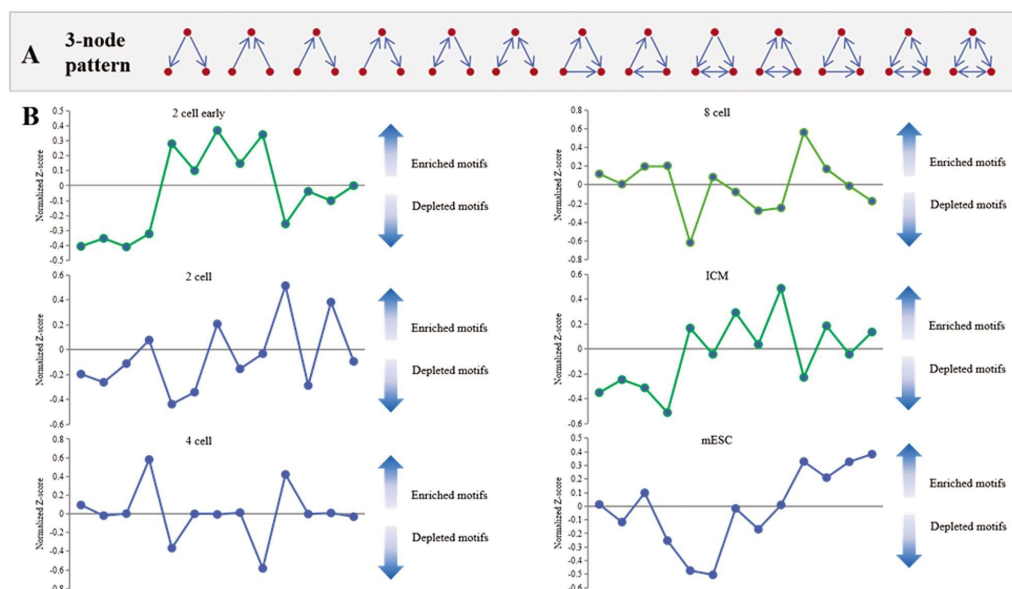


图4 早期胚胎发育各时期转录因子调控模式

Fig.4 Regulatory mode of transcription factors in different stage of early embryonic development

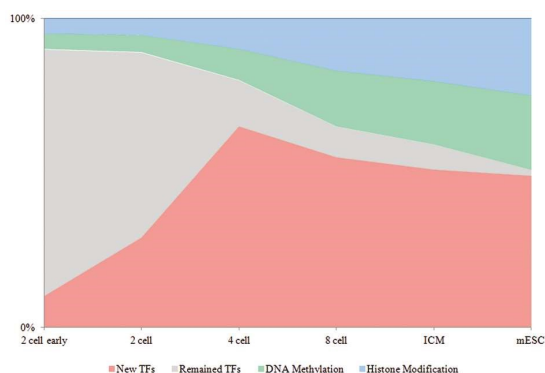


图5 早期胚胎发育基因表达调控模式图

Fig.5 Regulation mode of gene expression during early embryonic development

3 讨论

早期胚胎发育各时期转录因子表达量的变化趋势显示,早期胚胎发育到 2cell 时期,成熟卵母细胞中预存的促进早期胚

根据前期报道的 13 种三节点转录因子调控网络^[20],见图 4A,我们将小鼠早期胚胎各时期中发现的三节点转录因子调控网络的相对富集情况进行分析,结果显示小鼠早期胚胎发育各时期转录因子调控网络富集模式差异较大,2 细胞早期与 ICM 时期的转录因子调控模式具有很高的相似性,而 8 细胞时期与 ICM 时期的转录因子调控模式相反。见图 4B。

2.5 转录因子及表观遗传修饰对早期胚胎各时期基因表达的调控模式

通过计算 DNA 甲基化和活性组蛋白修饰 H3K4me3 的信号水平,整合多组学数据关联分析,早期胚胎发育各时期基因表达的调控因素展现出复杂有序的调控模式。2 细胞早期后母源转录因子所占比例逐渐减少,新生转录因子比例逐渐增多,在 2 细胞后逐渐取代母源转录因子在基因表达中发挥主要调控作用,DNA 甲基化及组蛋白修饰对基因表达的调控逐渐增强。见图 5。

胎快速发育的转录因子编码基因逐渐被消耗,伴随着胚胎基因组激活,新生的转录因子逐渐累积参与早期胚胎发育的基因表达调控。MZT 和 ZGA 是哺乳动物胚胎发育的重要事件,他们决定了胚胎能否在母源性 mRNA 消耗后承接后续的基因表达,完成母子间两套遗传信息的衔接^[4]。早期胚胎发育各时期转录因子的表达量变化说明,转录因子对早期胚胎基因表达的调控同样存在母源转录因子到胚胎自身新生转录因子调控的转换。

基因表达量与 TFBS 数量的相关性分析表明 2 细胞早期,基因表达量和 TFBS 数量不相关,这可能是胚胎植入初期,母源性 mRNA 的遗留效应导致的,测序得到的基因表达水平并不足以体现当前时期基因的表达程度。结合早期胚胎发育各时期转录因子表达量的变化,说明早期胚胎发育至 2 细胞时期,母源性转录因子在 2 细胞时期前的基因表达调控中发挥主要作用;2 细胞至 4 细胞时期,染色质逐渐被打开,胚胎新生成的转录因子开始调控基因表达;4 cell 至 8 细胞时期,染色质可及性广泛增加,越来越多胚胎自身转录因子生成,对基因的表达起主要调控作用。4 细胞之后,DNA 甲基化、组蛋白修饰逐渐发

挥作用,基因表达的调控因素变得复杂,转录因子水平与基因表达的相关性逐渐减弱。

对早期胚胎发育各时期高表达基因(表达量前 10%)分析发现,早期胚胎发育至 2 细胞时期,转录因子编码基因的表达量占高表达基因的比例较高,说明成熟卵母细胞遗留的 mRNA 中有大量的转录因子 mRNA,参与早期胚胎的基因表达调控,并促进 ZGA 的发生。2 细胞后期的转录因子占高表达基因的比例下降,启动子区域的染色质可及性增加,说明 2 细胞后早期胚胎基因表达调控有其他因素参与,这些因素包括染色质可及性、DNA 甲基化、组蛋白修饰等表观遗传修饰^[22,25-29]。

对早期胚胎发育各时期转录因子的调控模式分析发现,各时期转录因子对基因表达调控的模式具有一定多态性,2 细胞时期与 ICM 时期转录因子具有相似的调控模式,而 8 细胞时期与 ICM 时期转录因子的调控模式呈现相反的趋势。以上结果表明,除了转录因子数量变化外,早期胚胎各阶段转录因子的相互调控关系也发生了变化,两种调控方式共同参与了早期胚胎发育过程的基因表达调控。

总之,早期胚胎发育的基因表达调控是一个复杂且动态变化的过程,由转录因子数量及其调控关系,染色质可及性、DNA 甲基化修饰、组蛋白修饰等多种因素共同调控,维持早期胚胎发育过程中 MZT、ZGA、细胞命运决定及谱系分化等重要生物过程的有序进行,对早期胚胎的正常发育具有十分重要的意义^[30-32]。

参考文献(References)

- [1] Bowman C E, Arany Z, Wolfgang M J. Regulation of maternal-fetal metabolic communication [J]. Cellular and molecular life sciences: CMLS, 2021, 78(4): 1455-1486
- [2] Li L, Lu X, Dean J. The maternal to zygotic transition in mammals [J]. Molecular aspects of medicine, 2013, 34(5): 919-938
- [3] Lanner F. Lineage specification in the early mouse embryo [J]. Experimental cell research, 2014, 321(1): 32-39
- [4] Sha Q Q, Zhu Y Z, Li S, et al. Characterization of zygotic genome activation-dependent maternal mRNA clearance in mouse [J]. Nucleic acids research, 2020, 48(2): 879-894
- [5] Baillie-Benson P, Moris N, Martinez Arias A. Pluripotent stem cell models of early mammalian development [J]. Current opinion in cell biology, 2020, 66: 89-96
- [6] Mihajlović A I, Bruce A W. The first cell-fate decision of mouse preimplantation embryo development: integrating cell position and polarity [J]. Open biology, 2017, 7(11)
- [7] Rossant J, Tam P P L. Exploring early human embryo development [J]. Science (New York, NY), 2018, 360(6393): 1075-1076
- [8] Canovas S, Ross P J. Epigenetics in preimplantation mammalian development[J]. Theriogenology, 2016, 86(1): 69-79
- [9] Eckersley-Maslin M A, Alda-Catalinas C, Reik W. Dynamics of the epigenetic landscape during the maternal-to-zygotic transition [J]. Nature reviews Molecular cell biology, 2018, 19(7): 436-450
- [10] Ishiuchi T, Abe S, Inoue K, et al. Reprogramming of the histone H3.3 landscape in the early mouse embryo[J]. Nature structural & molecular biology, 2021, 28(1): 38-49
- [11] Ladstätter S, Tachibana K. Genomic insights into chromatin reprogramming to totipotency in embryos [J]. The Journal of cell biology, 2019, 218(1): 70-82
- [12] Liu X, Wang C, Liu W, et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos [J]. Nature, 2016, 537(7621): 558-562
- [13] Sun L, Fu X, Ma G, et al. Chromatin and Epigenetic Rearrangements in Embryonic Stem Cell Fate Transitions [J]. Frontiers in cell and developmental biology, 2021, 9: 637309
- [14] Vastenhouw N L, Cao W X, Lipshitz H D. The maternal-to-zygotic transition revisited[J]. Development (Cambridge, England), 2019, 146(11)
- [15] Wang Y, Liu Q, Tang F, et al. Epigenetic Regulation and Risk Factors During the Development of Human Gametes and Early Embryos [J]. Annual review of genomics and human genetics, 2019, 20: 21-40
- [16] Xia W, Xie W. Rebooting the Epigenomes during Mammalian Early Embryogenesis[J]. Stem cell reports, 2020, 15(6): 1158-1175
- [17] Xu Q, Xie W. Epigenome in Early Mammalian Development: Inheritance, Reprogramming and Establishment [J]. Trends in cell biology, 2018, 28(3): 237-253
- [18] Xu R, Li C, Liu X, et al. Insights into epigenetic patterns in mammalian early embryos[J]. Protein & cell, 2021, 12(1): 7-28
- [19] Zeng Y, Chen T. DNA Methylation Reprogramming during Mammalian Development[J]. Genes, 2019, 10(4)
- [20] Zhang B, Zheng H, Huang B, et al. Allelic reprogramming of the histone modification H3K4me3 in early mammalian development [J]. Nature, 2016, 537(7621): 553-557
- [21] Tosenberger A, Gonze D, Chazaud C, et al. Computational models for the dynamics of early mouse embryogenesis [J]. The International journal of developmental biology, 2019, 63(3-4-5): 131-142
- [22] Wu J, Huang B, Chen H, et al. The landscape of accessible chromatin in mammalian preimplantation embryos [J]. Nature, 2016, 534(7609): 652-657
- [23] Wang L, Zhang J, Duan J, et al. Programming and inheritance of parental DNA methylomes in mammals [J]. Cell, 2014, 157(4): 979-991
- [24] Kulakovskiy I V, Vorontsov I E, Yevshin I S, et al. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis[J]. Nucleic acids research, 2018, 46(D1): D252-d259
- [25] Neph S, Stergachis A B, Reynolds A, et al. Circuitry and dynamics of human transcription factor regulatory networks[J]. Cell, 2012, 150(6): 1274-1286
- [26] Dahlet T, Argüeso Lleida A, Al Adhami H, et al. Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity [J]. Nature communications, 2020, 11(1): 3153
- [27] Tang S B, Yang L L, Zhang T T, et al. Multiple superovulations alter histone modifications in mouse early embryos [J]. Reproduction (Cambridge, England), 2019, 157(6): 511-523
- [28] Deng M, Chen B, Liu Z, et al. Exchanges of histone methylation and variants during mouse zygotic genome activation [J]. Zygote (Cambridge, England), 2020, 28(1): 51-58
- [29] Yu B, Smith T H, Battle S L, et al. Superovulation alters global DNA methylation in early mouse embryo development [J]. Epigenetics, 2019, 14(8): 780-790
- [30] Pijuan-Sala B, Wilson N K, Xia J, et al. Single-cell chromatin accessibility maps reveal regulatory programs driving early mouse organogenesis[J]. Nature cell biology, 2020, 22(4): 487-497
- [31] Zhang Y, Xiang Y, Yin Q, et al. Dynamic epigenomic landscapes during early lineage specification in mouse embryos[J]. Nature genetics, 2018, 50(1): 96-105
- [32] Wang C, Liu X, Gao Y, et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development [J]. Nature cell biology, 2018, 20(5): 620-631