

外显子跳跃模式中组蛋白修饰的组合模式分析



田园芳¹, 陈伟^{1,2*}

(1. 华北理工大学生命科学院 河北 唐山 063000; 2. 成都中医药大学中医药创新研究院 成都 611137)

【摘要】可变剪接是基因表达调控的重要过程。组蛋白修饰参与了可变剪接,但具体调控机制尚不清楚。为了研究组蛋白修饰与可变剪接的关系,首先基于人胚肺成纤维细胞系的 RNA-Seq 数据,获得了该细胞系外显子跳跃模式的可变剪接数据。在此基础上,分析了外显子跳跃模式中排除和包含外显子中组蛋白修饰间的相关性。通过构建组蛋白修饰间的贝叶斯网络,推断了人胚肺成纤维细胞系外显子跳跃可变剪接事件中组蛋白修饰之间的因果关系。

关键词 可变剪接; 贝叶斯网络; 外显子跳跃; 组蛋白修饰

中图分类号 TP391; Q811.4 **文献标志码** A **doi:**10.12178/1001-0548.2022138

Analysis of Combinatorial Pattern of Histone Modifications in Exon Skipping Event

TIAN Yuanfang¹ and CHEN Wei^{1,2*}

(1. School of Life Sciences, North China University of Science and Technology Tangshan Hebei 063000;

2. Innovative Institute of Chinese Medicine and Pharmacy, Chengdu University of Traditional Chinese Medicine Chengdu 611137)

Abstract Alternative splicing is a key process of gene expression regulation. The histone modifications are closely correlated with alternative splicing, however, its regulatory mechanisms remain unclear. In order to reveal the relationships between histone modifications and alternative splicing, by analyzing the RNA-Seq data, the exon skipping events from human embryonic lung fibroblast cell line were obtained. Subsequently, the relationships among different kinds of histone modifications were analyzed for both excluded and included exons in exon skipping event. Finally, the Bayesian network was built to deduce the casual relationships among histone modifications in exon skipping event.

Key words alternative splicing; Bayesian network; exon skipping event; histone modifications

RNA 可变剪接 (alternative splicing, AS) 是真核生物中剪接体选择性剪切外显子形成不同 RNA 异构体的过程^[1], 是调节基因表达、产生蛋白质分子多样性的关键环节^[2]。文献 [3-4] 发现, 90% 以上的人类基因都会经历可变剪接。可变剪接不仅增加了生物分子的复杂性、多样性, 还与疾病的发生有关^[5]。如癌基因 ETS 中外显子 7b 的剪接与细胞增殖的减少有关^[6]。由于剪接因子 MBNL3 的调节, lncRNA PXN-AS1 的外显子 4 被保留在转录本中, 促进了肝癌的发生^[7]。丙酮酸激酶前体 mRNA 在剪接过程中保留了外显子 10, 产生的亚型 PKM2 过度表达, 导致了肿瘤的发生^[8]。可变剪接产生的雌激素受体 α 和 β 的变体 ER α 46 和 ER β 1 则与乳腺癌密切相关^[9]。因此, 对可变剪接调控机制的研究尤为重要。

RNA 可变剪接并非独立的生物过程, 而是与转录过程存在着时空上的偶联^[10]。除了位于外显子和内含子中的顺式和反式元件^[11-12], 可变剪接还受到组蛋白修饰、DNA 甲基化等表观遗传因素的调节。随着 DNA 元件百科全书 (encyclopedia of DNA elements, ENCODE) 计划的深入开展, 组蛋白修饰参与可变剪接调控的现象也被逐渐发现, 尤其是出现在内含子/外显子区的组蛋白甲基化和组蛋白乙酰化修饰与细胞系特异性可变剪接密切相关^[13]。组蛋白乙酰化修饰对 RNA 可变剪接的调控在神经细胞粘附分子 (neural cell adhesion molecule, NCAM) 基因中被发现。NCAM 基因中盒式外显子的切除与外显子中高含量的 H3K9ac 密切相关^[14]。NCAM 基因盒式外显子中 H3K9ac 改变了该区域的染色质

收稿日期: 2022-05-11; 修回日期: 2022-07-05

基金项目: 国家自然科学基金 (31771471)

作者简介: 田园芳 (1996-), 女, 主要从事生物信息学方面的研究。

*通信作者: 陈伟, E-mail: greatchen@ncst.edu.cn

结构, 造成转录过程中 RNA Pol II 的移动速率加快, 从而导致 NCAM 基因中盒式外显子发生了可变剪接。在小鼠胚胎干细胞分化为神经元的过程中, 组蛋白乙酰化修饰还调控了 Nf1 基因可变外显子 23a 和 Fas 基因可变外显子 6 的可变剪接^[15]。另外, 成纤维原细胞生长因子受体基因 (fibroblast growth factor receptor2, FGFR2) 中外显子区的 H3K36me3 是可变剪接调控蛋白的识别标记。FGFR2 基因中存在一对互斥外显子 IIIb 和 IIIc, FGFR2-IIIb 只在上皮细胞中表达, 而 FGFR2-IIIc 却在间质细胞中表达^[16]。通过分析间质细胞和上皮细胞中 FGFR2 基因的组蛋白修饰后发现, 与上皮细胞相比, 间质细胞的 FGFR2 基因外显子中富含 H3K36me3。因此, 染色质重塑复合物 MRG15 通过与 FGFR2 基因中的 H3K36me3 相互作用, 能够招募多聚嘧啶结合蛋白 (polypyrimidine tract-binding protein, PTB) 与 FGFR2-IIIb 外显子侧翼的内含子剪接抑制子结合, 从而使得 FGFR2-IIIb 在间质细胞中被切除。

组蛋白修饰间还存在因果关联, 多种组蛋白修饰组合在一起形成级联, 共同调控基因表达^[17-18]。在可变剪接过程中, 不同类型的组蛋白修饰可以通过协同或拮抗方式调控剪接复合因子的招募, 从而实现 RNA 剪接过程的调控。文献^[19]发现, 人胚肺成纤维细胞系 (IMR90 cell line) 中 BIN1 基因的可变剪接就是多种组蛋白修饰 (H3K36me3、H3K4me3、H2BK12ac、H4K5ac) 协同作用的结果。这些工作既为研究组蛋白甲基化和组蛋白乙酰化修饰调控可变剪接提供了理论依据, 又显示出从组蛋白修饰等表观遗传因素中挖掘新信息是认识可变剪接调控机制的新途径。

外显子跳跃模式是哺乳动物最常见的可变剪接模式^[20-21]。文献^[22-23]发现了 CD4⁺ T 细胞外显子跳跃模式中多种组蛋白修饰在外显子和内含子中富集程度的差异性, 并利用组蛋白修饰差异信息对包含和排除外显子进行了识别。通过构建组蛋白修饰间的相互作用网络, 文献^[24]还分析了 CD4⁺ T 细胞的外显子跳跃模式中组蛋白修饰之间的因果关系。组蛋白修饰不仅可以通过改变 RNA Pol II 的延伸率或招募剪接因子参与可变剪接的调节^[25], 还能通过彼此之间的相互作用调控可变剪接。通过对人胚胎干细胞系 (H1 cell line) 的转录组学和表观遗传组学数据进行关联分析, 文献^[26]发现组蛋白修饰的动态变化与细胞特异性剪接机制相关。最近,

文献^[27]发现了组蛋白修饰在 H1 和 IMR90 细胞系的外显子跳跃模式中的协同分布规律, 并在 MCF10a、K562 和 HeLa 等细胞系中进行了验证, 遗憾的是其并未阐明组蛋白修饰间的因果关系。

鉴于此, 本文以 IMR90 细胞系中外显子跳跃剪接事件为研究对象, 分析了 28 种组蛋白修饰在外显子跳跃模式排除和包含外显子上的相关性, 通过构建贝叶斯网络推断了组蛋白修饰间的因果关系。

1 材料与方法

1.1 数据来源

人类基因组 (GRCh37 版本) cDNA 序列和基因注释文件来自 Ensembl 数据库 (<https://asia.ensembl.org/index.html>)。IMR90 细胞系转录组测序数据 (RNA-seq) 来自 GEO 数据库 (<https://www.ncbi.nlm.nih.gov/geo/>, GSM2400222)。使用高通量数据过滤工具 Trimmomatic^[28] (版本 0.39-2) 对 RNA-seq 进行处理, 参数设置为 “PEILLUMINACLIP: TruSeq3-PE-2.fa:2:30:10:1:true LEADING:3 TRAILING:3 SLIDING WINDOW:4:20 MINLEN:40 TOPHRED33”。

从 GEO 数据库 (登录号 GSE16256) 获取 IMR90 细胞系的 28 种组蛋白修饰数据的 bed 文件, 相应的 GEO 样本号如表 1 所示。

表 1 组蛋白修饰的 GEO 样本号

组蛋白修饰	GEO 样本号
H2A.Z	GSM818007
H2AK5ac	GSM521866
H2AK9ac	GSM818012
H2BK120ac	GSM521869
H2BK12ac	GSM521871
H2BK15ac	GSM521875
H2BK20ac	GSM521879
H2BK5ac	GSM818017
H3K14ac	GSM521881
H3K18ac	GSM469965
H3K23ac	GSM521885
H3K27ac	GSM469966
H3K27me3	GSM469968
H3K36me3	GSM521890
H3K4ac	GSM521893
H3K4me1	GSM521895
H3K4me2	GSM521899
H3K4me3	GSM469970
H3K56ac	GSM521902
H3K79me1	GSM521904
H3K79me2	GSM521909
H3K9ac	GSM469973
H3K9me1	GSM752986
H3K9me3	GSM469974
H4K20me1	GSM521915
H4K5ac	GSM469975
H4K8ac	GSM521919
H4K91ac	GSM521924

1.2 转录本定量

统计 RNA-seq 比对到基因组上的短读序列 (reads) 的数量, 即每个转录本的表达量。由于测序深度、转录本长度等因素的影响, 表达量并不能直接代表转录本的相对丰度。因此对这些表达量进行 TPM^[29] 标准化:

$$\text{TPM} = \frac{N_i/L_i \times 10^6}{\text{sum}(N_1/L_1 + N_2/L_2 + \dots + N_n/L_n)} \quad (1)$$

式中, N_i 表示映射到第 i 个转录本的 reads 数; L_i 表示第 i 个转录本的长度。

使用 Salmon^[30] (版本 1.5.1) 计算转录本的 TPM 值, 首先在参考 cDNA 序列上建立索引, 参数使用默认值。生成的索引与处理后的 RNA-seq 进行定量, 参数设置为 “-l A-validateMappings-gcBias--seqBias”。

1.3 外显子跳跃事件分析

外显子跳跃事件中被保留在成熟转录本中的外显子为包含外显子, 被剪接的外显子为排除外显子。根据发生外显子跳跃事件的基因上所有转录本的 TPM, 计算每个外显子的包含率 (percent spliced in, PSI):

$$\text{PSI} = \frac{\sum_{i \in S_1} \text{TPM}_i}{\sum_{j \in S_1 \cup S_2} \text{TPM}_j} \quad (2)$$

式中, TPM_i 表示基因的第 i 个包含跳跃外显子的转录本的 TPM 值; S_1 表示包含跳跃外显子的转录本集合; TPM_j 表示基因的第 j 个转录本的 TPM 值; S_2 表示不包含跳跃外显子的转录本集合。PSI 的值越大, 表示外显子被包含在最终转录本的概率越大, PSI 的值越小, 外显子被排除的概率越大。

使用 Suppa 软件^[31] (版本 2.3), 根据基因注释信息生成外显子跳跃事件, 并利用 TPM 计算每个跳跃外显子的 PSI, 所有参数使用默认值。最终获得 36 468 个外显子跳跃事件, 其外显子 PSI 值的范围为 0~1。PSI 为 1 或 0 时, 表示外显子被包含或被排除在基因的所有转录本中。为了适当扩充数据量, 定义 $\text{PSI} > 0.85$ 的外显子为包含外显子, $\text{PSI} < 0.15$ 的外显子为排除外显子。由于外显子过短无法匹配到组蛋白修饰数据, 因此只保留长度在 150~300 bp 之间的外显子, 最后获得 5 122 个包含外显子和 4 638 个排除外显子。

1.4 相关性计算

根据 bed 文件中组蛋白修饰在基因组上的位点信息, 使用 BEDTools^[32] (版本 2.30.0) 计算其在排除和包含外显子上的富集程度 (reads 数)。然后根据 reads 数, 使用 R 语言 (版本 4.1.2) Hmisc 包中的 rcorr 函数分别计算排除和包含外显子上组蛋白修饰之间的皮尔逊相关系数。由此获得 28 种组蛋白修饰间的相关性系数矩阵, 用 corrplot 绘制相关性热图, 并设置统计显著性阈值 $p=0.05$ 。

1.5 贝叶斯网络构建

贝叶斯网络是一种描述变量间因果关系的统计推理模型, 其网络拓扑结构是有向无环图 (directed acyclic graph, DAG)^[33]。网络中包括节点和边, 节点表示随机变量, 边表示变量之间的条件依赖关系。所有边为单向箭头, 箭头指向的节点为子节点, 箭头另一端的节点为父节点。在贝叶斯网络中, 如果两个节点之间不存在有向边, 则说明这两个节点彼此条件独立; 如果两个节点间存在有向边, 则说明这两个节点间存在因果关系。

根据 1.4 节的 reads 数, 对排除和包含外显子上组蛋白修饰进行离散化, “1” 表示外显子上存在组蛋白修饰 (reads>1), “0” 表示外显子上不存在组蛋白修饰 (reads=0)。为了得到稳定的组蛋白修饰相互作用网络, 采用 10 交叉检验法对所得网络进行验证。首先将组蛋白修饰数据平均分成 10 份, 其中 9 份被用作训练集, 用于构建基本的贝叶斯网络, 1 份被用作测试集对网络的稳定性进行验证。使用 WinMine (<https://www.microsoft.com/en-us/research/project/winmine-toolkit/>) 构建组蛋白修饰的贝叶斯网络, 最终得到 10 个不同的网络。如果连接节点的有向边在 10 个网络中均存在, 则将这些节点和边保留, 用以构建最终网络。利用 Cytoscape^[34] (版本 3.8.2) 软件展示所得的网络拓扑结构。网络中节点表示组蛋白修饰, 并根据 1.4 节得到的相关系数对边赋值。

2 结果

2.1 组蛋白修饰间的相关性分析

计算组蛋白修饰之间的相关系数, 通过绘制相关性热图, 分析组蛋白修饰间的相关性, 如图 1 和图 2 所示。图中 “×” 表示组蛋白修饰间的相关性不显著, 色块大小表示相关性强弱。结果表明, IMR90 细胞系中排除和包含外显子上大部分组蛋白修饰之间存在协同 ($r > 0$) 或拮抗作用 ($r < 0$)。

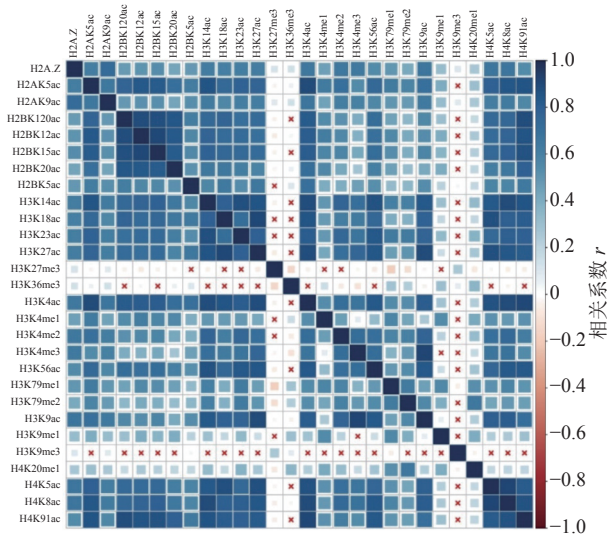


图 1 排除外显子上组蛋白修饰之间的相关性

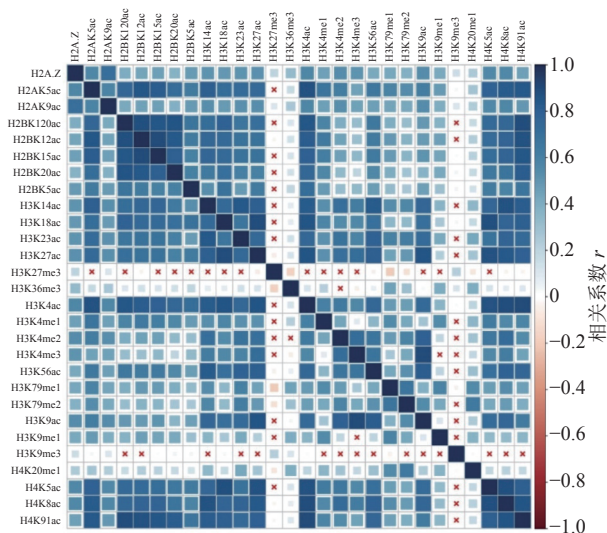


图 2 包含外显子上的组蛋白修饰之间的相关性

具体而言, 组蛋白乙酰化之间表现出了正相关性, 并且具有强正相关的组蛋白修饰存在于同一组蛋白上(H2B: H2BK120ac、H2BK12ac、H2BK15ac、H2BK20ac; H3: H3K14ac、H3K18ac、H3K23ac、H3K27ac; H4: H4K5ac、H4H8ac、H4K91ac)。在 CD4⁺T 细胞中也发现了组蛋白 H3 和 H4 上相同的组合模式, 但其正相关性弱于 IMR90 细胞系^[24]。

组蛋白甲基化之间同时表现出正相关性和负相关性, 如 H4K20me1 和 H3K79me1、H3K79me2 之间正相关, 而 H3K27me3 和 H3K36me3、H3K79me1、H3K79me2 之间负相关。在间质细胞中存在相同的现象, H3K27me3 和 H3K36me3 对基因 FGFR2 的外显子 IIIb 的保留和剪切表现出了拮抗作用^[35]。

排除和包含外显子上组蛋白修饰间的相关性也

存在差异。如 H3K36me3 在包含外显子上与 10 种组蛋白修饰 (H2BK120ac、H2BK15ac、H3K14ac、H3K18ac、H3K23ac、H3K27ac、H3K4ac、H3K56ac、H4K5ac、H4K91ac) 之间正相关, 如图 2 所示, 而在排除外显子中则未发现显著的相关性, 如图 1 所示。文献 [19, 36] 研究发现, H3K36me3 富集在包含外显子上, 推测其可能与不同的组蛋白修饰形成组合模式, 参与了 RNA 可变剪接的调控。

2.2 组蛋白修饰之间的因果关系分析

通过构建 IMR90 细胞中组蛋白修饰间的贝叶斯网络, 如图 3 和 4 所示, 对外显子跳跃剪接事件中组蛋白修饰间的因果关系进行推断, 发现贝叶斯网络中涉及的组蛋白修饰不仅包括同一组蛋白中同一氨基酸不同程度的修饰, 还包括不同组蛋白中同一类型的组蛋白修饰, 以及不同组蛋白中不同类型的组蛋白修饰。

在网络拓扑结构图中, 只有子节点的组蛋白修饰用紫色标出; 只有父节点的组蛋白修饰用绿色标出; 既有父节点也有子节点的组蛋白修饰用蓝色标出。

比较两组网络拓扑结构发现, 排除和包含外显子上组蛋白修饰之间存在 18 种相同的因果关系, 分别为: H2A.Z→H2AK9ac→H2BK5ac、H2AK9ac→H3K9me1、H2A.Z→H3K27me3、H2A.Z→H3K9me1、H2A.Z→H3K9me3、H2AK5ac→H2BK12ac、H2AK5ac→H3K14ac、H2AK5ac→H2BK15ac、H2AK5ac→H4K91ac、H3K18ac→H3K14ac、H3K18ac→H2BK120ac、H3K18ac→H3K56ac、H3K4me2→H3K56ac、H4K8ac→H3K79me1→H3K36me3、H4K8ac→H3K36me3、H4K8ac→H2BK15ac。

在排除外显子中, 27 种组蛋白修饰构成了 71 种因果关系, 如图 3 所示。H3K4me3 直接或间接调控了剩余的 26 种组蛋白修饰。8 种组蛋白修饰 (H2BK120ac、H2BK15ac、H2BK5ac、H3K14ac、H3K27me3、H3K4me1、H3K56ac、H3K9me1) 可能直接与剪接因子相互作用, 参与了可变剪接调控。在包含外显子中, 26 种组蛋白修饰构成了 35 种因果关系, 如图 4 所示, 其中 7 种组蛋白修饰 (H2A.Z、H2AK5ac、H3K18ac、H3K4ac、H3K4me2、H4K5ac、H4K8ac) 直接或间接调节剩余的 19 种组蛋白修饰。16 种组蛋白修饰 (H2BK120ac、H2BK12ac、H2BK15ac、H2BK20ac、H2BK5ac、H3K14ac、

H3K27ac、H3K27me3、H3K36me3、H3K4me3、H3K56ac、H3K9ac、H3K9me1、H3K9me3、

H4K20me1、H4K91ac) 可能直接与剪接因子相互作用, 参与了可变剪接的调控。

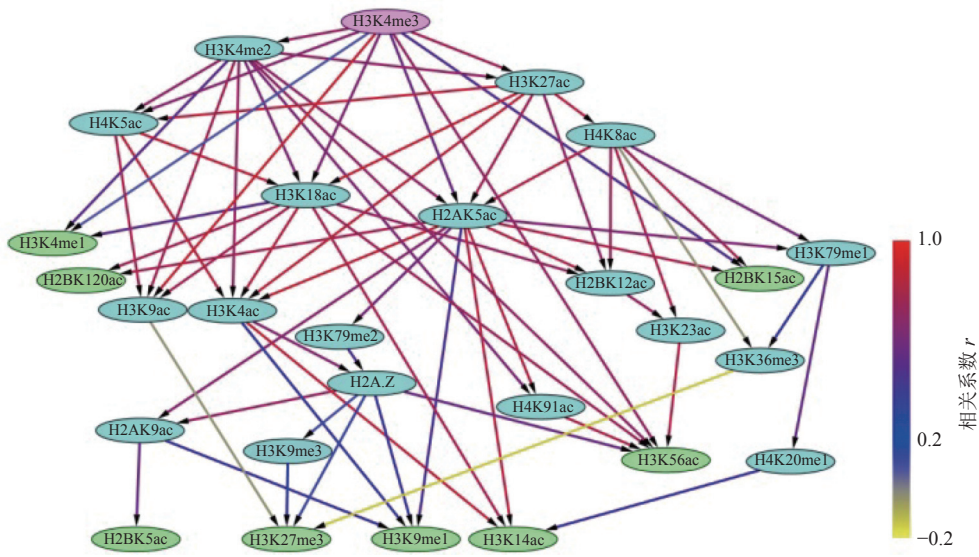


图 3 排除外显子上组蛋白修饰之间的贝叶斯网络

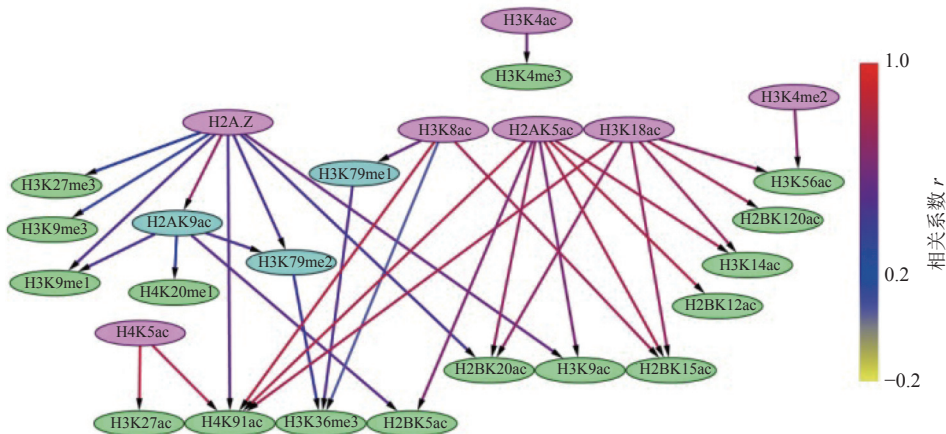


图 4 包含外显子上组蛋白修饰之间的贝叶斯网络

此外, 排除和包含外显子对应的网络复杂程度也存在明显差异。排除外显子对应的网络拓扑结构复杂, 组蛋白修饰组合的调控路径长, 并且部分组蛋白修饰间表现出了拮抗关系, 如 H4K8ac→H3K36me3→H3K27me3、H3K9ac→H3K27me3, 如图 3 所示。包含外显子对应的网络拓扑结构则相对简单, 组蛋白修饰组合的调控路径较短, 如图 4 所示。另外 IMR90 细胞中排除和包含外显子上的 H3K27ac 和 H4K5ac 之间因果关系相反。

组蛋白修饰在可变剪接过程中的调控作用已被发现, 如 H3K4me3 与 U2 snRNP 结合可调节剪接速率^[37-38]。IMR90 细胞系包含外显子上的调控网络

显示 H3K4me3 可能受到了 H3K4ac 的影响, 如图 4 所示。而排除外显子上 H3K4me3 不仅与 9 种组蛋白修饰 (H3K4me2、H4K5ac、H3K4me1、H3K9ac、H3K18ac、H2AK5ac、H3K56ac、H2BK15ac、H3K27ac) 之间存在直接因果关系, 还存在于所有的调控路径中, 如图 3 所示。在间质细胞中, H3K36 的去甲基化酶 KDM2a 被募集到富含 H3K27me3 的区域, 保持了低 H3K36me3 水平, 从而促进了外显子 IIIb 的包含^[35]。这一结果表明, H3K36me3 与 H3K27me3 之间的拮抗作用调控了可变剪接。有意思的是, 在 IMR90 细胞系的排除外显子中也发现了 H3K36me3 和 H3K27me3 之间的拮抗关系, 如

图3所示。由此推测, IMR90细胞系中的可变剪接受到了组蛋白修饰间相互作用的调控, 相关结果还需进一步实验验证。

3 结束语

本文通过对IMR90细胞系中组蛋白修饰间的相关性进行分析, 发现了外显子跳跃剪接事件中组蛋白修饰间存在明显的组合模式。通过构建贝叶斯网络, 分析了排除和包含外显子中组蛋白修饰间的因果关系。由此推测, IMR90细胞系排除和包含外显子的组蛋白修饰可能通过组合方式与剪接因子相互作用, 直接或间接地参与了可变剪接的调控。

参 考 文 献

- [1] LEE Y, RIO D C. Mechanisms and regulation of alternative pre-mRNA splicing[J]. *Annu Rev Biochem*, 2015, 84: 291-323.
- [2] KELEMEN O, CONVERTINI P, ZHANG Z, et al. Function of alternative splicing[J]. *Gene*, 2013, 514(1): 1-30.
- [3] WANG E T, SANDBERG R, LUO S, et al. Alternative isoform regulation in human tissue transcriptomes[J]. *Nature*, 2008, 456(7221): 470-476.
- [4] PAN Q, SHAI O, LEE L J, et al. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing[J]. *Nat Genet*, 2008, 40(12): 1413-1415.
- [5] SCOTTI M M, SWANSON M S. RNA mis-splicing in disease[J]. *Nat Rev Genet*, 2016, 17(1): 19-32.
- [6] JUMBE S L, PORAZINSKI S R, OLTEAN S, et al. The evolutionarily conserved cassette exon 7b drives ERG's oncogenic properties[J]. *Transl Oncol*, 2019, 12(1): 134-142.
- [7] YUAN J H, LIU X N, WANG T T, et al. The MBNL3 splicing factor promotes hepatocellular carcinoma by increasing PXN expression through the alternative splicing of lncRNA-PXN-AS1[J]. *Nat Cell Biol*, 2017, 19(7): 820-832.
- [8] DAVID C J, CHEN M, ASSANAHA M, et al. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer[J]. *Nature*, 2010, 463(7279): 364-368.
- [9] YANG Q, ZHAO J, ZHANG W, et al. Aberrant alternative splicing in breast cancer[J]. *J Mol Cell Biol*, 2019, 11(10): 920-929.
- [10] TILGNER H, KNOWLES D G, JOHNSON R, et al. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs[J]. *Genome Res*, 2012, 22(9): 1616-1625.
- [11] SALDI T, CORTAZAR M A, SHERIDAN R M, et al. Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing[J]. *J Mol Biol*, 2016, 428(12): 2623-2635.
- [12] LOERCH S, LEACH J R, HORNER S W, et al. The pre-mRNA splicing and transcription factor Tat-SF1 is a functional partner of the spliceosome SF3b1 subunit via a U2AF homology motif interface[J]. *J Biol Chem*, 2019, 294(8): 2892-2902.
- [13] ANDERSSON R, ENROTH S, RADA-IGLESIAS A, et al. Nucleosomes are well positioned in exons and carry characteristic histone modifications[J]. *Genome Res*, 2009, 19(10): 1732-1741.
- [14] SCHOR I E, RASCOVAN N, PELISCH F, et al. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing[J]. *Proc Natl Acad Sci USA*, 2009, 106(11): 4325-4330.
- [15] ZHU H, HINMAN M N, HASMAN R A, et al. Regulation of neuron-specific alternative splicing of neurofibromatosis type 1 pre-mRNA[J]. *Mol Cell Biol*, 2008, 28(4): 1240-1251.
- [16] LUCO R F, PAN Q, TOMINAGA K, et al. Regulation of alternative splicing by histone modifications[J]. *Science*, 2010, 327(5968): 996-1000.
- [17] CIEŚLIK M, BEKIRANOV S. Combinatorial epigenetic patterns as quantitative predictors of chromatin biology[J]. *BMC Genomics*, 2014, 15(1): 1-18.
- [18] ZHU Y, SUN L, CHEN Z, et al. Predicting enhancer transcription and activity from chromatin modifications[J]. *Nucleic Acids Res*, 2013, 41(22): 10032-10043.
- [19] SHINDO Y, NOZAKI T, SAITO R, et al. Computational analysis of associations between alternative splicing and histone modifications[J]. *FEBS Lett*, 2013, 587(5): 516-521.
- [20] SAMMETH M, FOISSAC S, GUIGÓ R. A general definition and nomenclature for alternative splicing events[J]. *PLoS Comput Biol*, 2008, 4(8): e1000147.
- [21] CHAUDHARY S, KHOKHAR W, JABRE I, et al. Alternative splicing and protein diversity: Plants versus animals[J]. *Front Plant Sci*, 2019, 10: 708.
- [22] ENROTH S, BORNELÖV S, WADELIUS C, et al. Combinations of histone modifications mark exon inclusion levels[J]. *PLoS One*, 2012, 7(1): e29911.
- [23] CHEN W, FENG P, DING H, et al. Classifying included and excluded exons in exon skipping event using histone modifications[J]. *Front Genet*, 2018, 9: 433.
- [24] CHEN W, SONG X, LIN H. Combinatorial pattern of histone modifications in exon skipping event[J]. *Front Genet*, 2019, 10: 122.
- [25] LUCO R F, ALLO M, SCHOR I E, et al. Epigenetics in alternative pre-mRNA splicing[J]. *Cell*, 2011, 144(1): 16-26.
- [26] XU Y, ZHAO W, OLSON S D, et al. Alternative splicing links histone modifications to stem cell fate decision[J]. *Genome Biol*, 2018, 19(1): 133.
- [27] AGIRRE E, OLDFIELD A J, BELLORA N, et al. Splicing-associated chromatin signatures: A combinatorial and position-dependent role for histone marks in splicing definition[J]. *Nat Commun*, 2021, 12(1): 682.
- [28] BOLGER A M, LOHSE M, USADEL B. Trimmomatic: A

- flexible trimmer for Illumina sequence data[J]. *Bioinformatics*, 2014, 30(15): 2114-2120.
- [29] WAGNER G P, KIN K, LYNCH V J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples[J]. *Theory Biosci*, 2012, 131(4): 281-285.
- [30] PATRO R, DUGGAL G, LOVE M I, et al. Salmon provides fast and bias-aware quantification of transcript expression[J]. *Nat Methods*, 2017, 14(4): 417-419.
- [31] TRINCADO J L, ENTIZNE J C, HYSENAJ G, et al. SUPPA2: Fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions[J]. *Genome Biol*, 2018, 19(1): 1-11.
- [32] QUINLAN A R, HALL I M. BEDTools: A flexible suite of utilities for comparing genomic features[J]. *Bioinformatics*, 2010, 26(6): 841-842.
- [33] YU H, ZHU S, ZHOU B, et al. Inferring causal relationships among different histone modifications and gene expression[J]. *Genome Res*, 2008, 18(8): 1314-1324.
- [34] FRANZ M, LOPES C T, HUCK G, et al. Cytoscape.js: A graph theory library for visualisation and analysis[J]. *Bioinformatics*, 2016, 32(2): 309-311.
- [35] GONZALEZ I, MUNITA R, AGIRRE E, et al. A lncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature[J]. *Nat Struct Mol Biol*, 2015, 22(5): 370-376.
- [36] KOLASINSKA-ZWIERZ P, DOWN T, LATORRE I, et al. Differential chromatin marking of introns and expressed exons by H3K36me3[J]. *Nat Genet*, 2009, 41(3): 376-381.
- [37] SIMS R J III, MILLHOUSE S, CHEN C F, et al. Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing[J]. *Mol Cell*, 2007, 28(4): 665-676.
- [38] DAVIE J R, XU W, DELCUVE G P. Histone H3K4 trimethylation: Dynamic interplay with pre-mRNA splicing[J]. *Biochem Cell Biol*, 2016, 94(1): 1-11.

编辑 刘飞阳