



Combinatorial Pattern of Histone Modifications in Exon Skipping Event

Wei Chen^{1,2,3*}, Xiaoming Song² and Hao Lin^{3*}

¹ Innovative Institute of Chinese Medicine and Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu, China, ² Center for Genomics and Computational Biology, School of Life Sciences, North China University of Science and Technology, Tangshan, China, ³ Key Laboratory for Neuro-Information of Ministry of Education, Center for Informational Biology, School of Life Sciences and Technology, University of Electronic Science and Technology of China, Chengdu, China

OPEN ACCESS

Edited by:

Dariusz Mrozek,
Silesian University of Technology,
Poland

Reviewed by:

Leyi Wei,
Tianjin University, China
Balachandran Manavalan,
Ajou University, South Korea

*Correspondence:

Wei Chen
chenweimu@gmail.com
Hao Lin
hlin@uestc.edu.cn

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Genetics

Received: 19 December 2018

Accepted: 04 February 2019

Published: 18 February 2019

Citation:

Chen W, Song X and Lin H (2019)
Combinatorial Pattern of Histone
Modifications in Exon Skipping Event.
Front. Genet. 10:122.
doi: 10.3389/fgene.2019.00122

Histone modifications are associated with alternative splicing. It has been suggested that histone modifications act in combinatorial patterns in gene expression regulation. However, how they interact with each other and what is their casual relationships in the process of RNA splicing remain unclear. In this study, the combinatorial patterns of 38 kinds of histone modifications in the exon skipping event of the CD4⁺ T cell were analyzed by constructing Bayesian networks. Distinct combinatorial patterns of histone modifications that illustrating their casual relationships were observed in excluded/included exons and the surrounding intronic regions. The Bayesian networks also indicate that some histone modifications directly correlate with RNA splicing. We anticipate that this work could provide novel insights into the effects of histone modifications on RNA splicing regulation.

Keywords: histone modification, methylation, acetylation, RNA splicing, Bayesian network, casual relationship

INTRODUCTION

Alternative splicing is a process that can generate multiple mRNA isoforms from a single gene by splicing pre-mRNA molecules in different ways (Black, 2003). As an important process of gene expression, alternative splicing ensures the diversity of gene expression products. It has been estimated that alternative splicing occurs in approximately 90% human genes (Pan et al., 2008; Wang E. T. et al., 2008). Alternative splicing is reported to closely correlate with apoptosis, embryonic development and even a series of diseases (Garcia-Blanco et al., 2004; David et al., 2010; Lai and Greenberg, 2013; Scotti and Swanson, 2016). Although great efforts have been made on studying alternative splicing, the mechanisms of cell type-specific and stage-specific alternative splicing are still unclear (Nellore et al., 2016).

Recent studies have revealed that alternative splicing is regulated not only by *trans*-acting factors that can interact with *cis*-acting elements (Badr and Heath, 2015; Badr et al., 2016), but also by epigenetic factors, such as DNA methylation, nucleosome occupancy, and so on (Mele et al., 2017). Since RNA splicing is coupled to transcription, histone modifications were also found to be involved in alternative splicing regulation (Kornblihtt et al., 2013). Luco et al. (2010) found that the alternative splicing of the FGFR2 gene was correlated with the level of H3K36me3. Saint-Andre et al. (2011) demonstrated that the inclusion/exclusion of the alternative exons of the CD44 mRNA is affected by H3K9me3. The combinatorial effect of histone modifications on alternative splicing was also reported. Recently, Shindo et al. (2013) found that the alternative splicing of the BIN1 gene in IMR90 cell was regulated by the cooperation of H3K36me3, H3K4me3, H2BK12ac, and

H4K5ac. These results strongly indicate that histone modifications play important roles in RNA splicing regulation and are key clues for revealing the regulatory mechanism of alternative splicing.

Based on these experimental results, several computational methods have been proposed to predict the alternative exons in exon skipping event based on histone modifications. The pioneer work was proposed by Enroth et al. (2012), in which a rule-based model was developed to classify included and excluded exons based on histone modification combinations. Later on, based on Enroth et al.'s (2012) dataset, Chen et al. (2014) proposed a quadratic discriminant (QD) function method and obtained an accuracy of 68.5% for classifying the included and excluded exons in the exon skipping event. More recently, a random forest based method was developed for the same aim and obtained an accuracy of 72.91% in the 10-fold cross validation test (Chen et al., 2018b). These results strongly indicate that histone modifications play important roles in RNA splicing regulation and are key clues for revealing the regulatory mechanism of alternative splicing. These results indicate that we should find the novel splicing code from the epigenome information.

Inspired by recent works (Cui et al., 2011; Zhu et al., 2013), in this study, the Bayesian network of histone modifications were constructed in the excluded/included exons and their preceding and succeeding intronic regions of exon skipping event to investigate how histone modifications interact with each other and find their casual relationships in the process of RNA splicing. By analyzing the Bayesian networks, distinct combinational patterns and casual relationships of histone modifications were observed in different regions relative to exons.

MATERIALS AND METHODS

Dataset

Based on the exon expression data of the CD4⁺ T cell (Oberdoerffer et al., 2008), by calculating the ratio between exon expression and gene expression, Enroth et al. (2012) obtained 13,374 “included” and 11,587 “excluded” exons. All of these exons are longer than 50 bp with flanking introns longer than 360 bp, and none of them are the first or last exon in any transcripts (Enroth et al., 2012).

The ChIP-seq data for the 20 kinds of histone methylations (H3K27me2, H3K4me1, H3K79me2, H3K9me3, H4K20me3, H3K27me3, H3K4me2, H3K79me3, H3R2me1, H4R3me2, H2BK5me1, H3K36me1, H3K4me3, H3K9me1, H3R2me2, H3K27me1, H3K36me3, H3K79me1, H3K9me2, and H4K20me1) and 18 kinds of histone acetylation modifications (H2AK5ac, H2BK20ac, H3K23ac, H3K9ac, H4K8ac, H2AK9ac, H2BK5ac, H3K27ac, H4K12ac, H4K91ac, H2BK120ac, H3K14ac, H3K36ac, H4K16ac, H2BK12ac, H3K18ac, H3K4ac, and H4K5ac) of the CD4⁺ T cell were obtained from previous works (Barski et al., 2007; Wang Z. et al., 2008). By using the SICTIN tool, Enroth et al. (2010) discretized the histone modification signals to binary (present/absent) attributes over the three regions, namely excluded/included exons, the closest 180 bp flanking intronic regions proceeding and succeeding the exons.

After winnowing out exons with no modifications present, they finally obtained 12,692 “included” and 11,165 “excluded” exons. The present/absent of the 38 kinds of histone modifications in the excluded/included exon and the preceding and succeeding intronic regions was annotated by “1” (indicating the presence of histone modification) or “0” (indicating the absence of histone modification), which were used to construct the histone modification Bayesian network. All the data can be found in Enroth et al.'s (2010) work.

Bayesian Network

Bayesian network is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG) (Yu et al., 2008). The nodes in Bayesian network represent variables, and the edges represent conditional dependencies. A directed edge ($\rightarrow$) from node a_i to node a_j represents a statistical dependence or the causal relationships between the corresponding variables. The arrow indicates that the variable a_j depends on the variable a_i . If there is no edge between two nodes a_i and a_j , indicating that the variables i and j are independent of each other.

In this study, the WinMine package which is available at <https://www.microsoft.com/en-us/research/project/winmine-toolkit/#!downloads>, was used to construct the Bayesian network of histone modifications in the excluded/included exons and the preceding and succeeding intronic regions. The nodes in the potential networks will be the histone modifications.

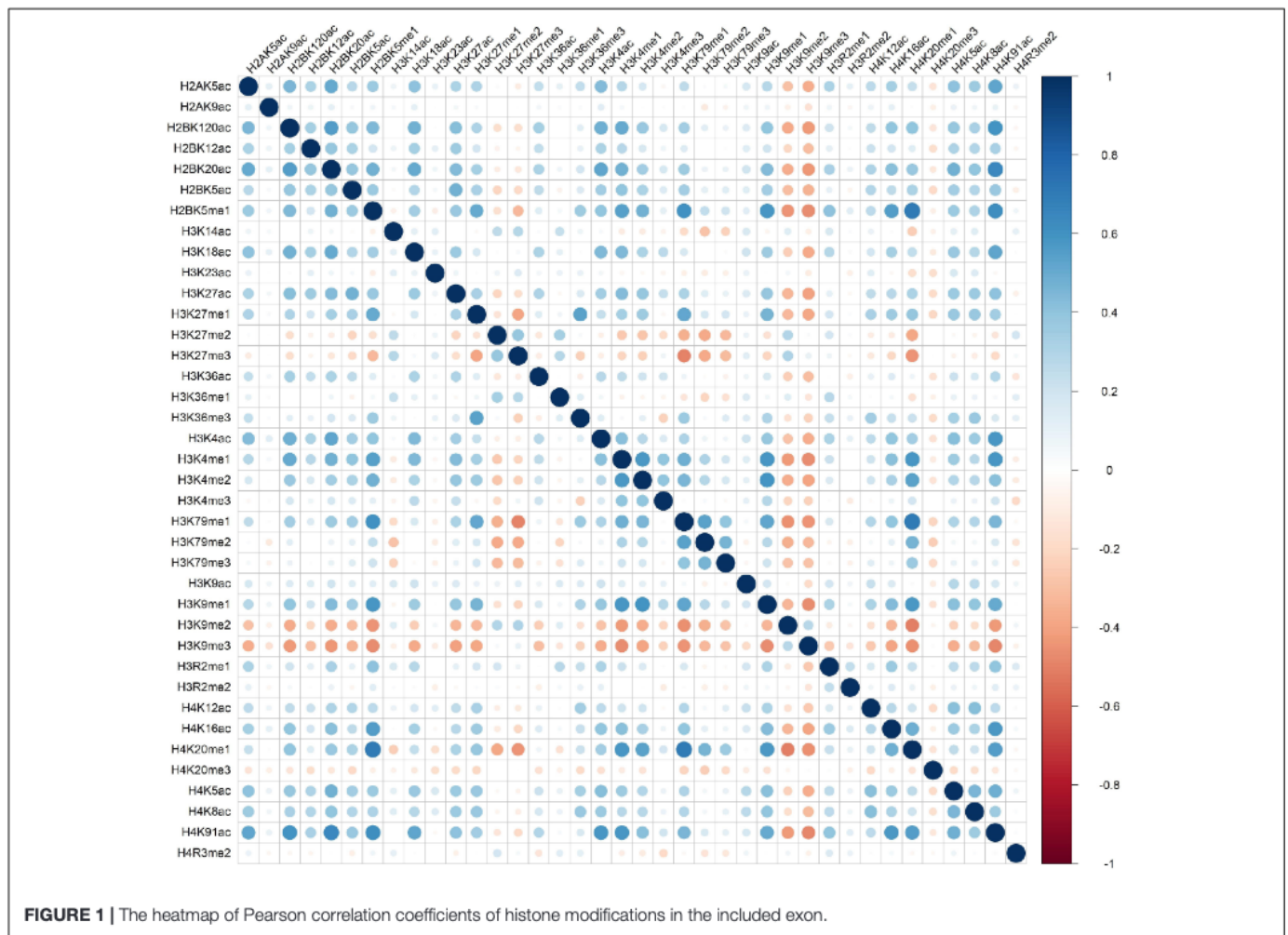
RESULTS AND DISCUSSION

Correlations Between Histone Modifications

Previous studies have reported that gene expression is in part regulated by histone modifications that act in a combinatorial fashion, i.e., the so-called “histone code” (Yu et al., 2008; Cui et al., 2011; Zhu et al., 2013). In order to find whether the combinatorial pattern of histone modifications exist in the process of RNA splicing, we first calculated the Pearson correlation coefficients between the 38 kinds of histone modifications in the excluded/included exons and the preceding and succeeding intronic regions, respectively.

Distinct combinational patterns of histone modifications were observed in the excluded/included exon and the surrounding regions. For example, H2BK5me1 was found to be positively correlated with H3K4me1, H3K4me2, H3K79me1, H3K9me1, H4K20me1, and H4K91ac in both included and excluded exons, **Figures 1** and **2**. The negative relationship were found between H3K9me3 with most of the remaining 37 kinds of histone modifications in both excluded and included exons, **Figures 1** and **2**. These results also hold for the preceding and succeeding intronic regions of the included and excluded exons (**Supplementary Figures S1–S4**).

Besides the common pattern, the excluded/included exon specific combinational patterns of histone modifications were also observed in excluded/included exon, **Figures 1** and **2**. For example, in the included exon, H3K4ac and



H2BK5me1, H3K79me1 and H3K23ac, H4K20ac and H3K4me1, H4K20ac and H3K4me2 exhibit negative correlations, which is absent in excluded exon; while the significantly negative correlation between H3K14ac and H2BK5me1, H4K120ac and H3K27me2, H4K120ac and H3K27me2 were only observed in the excluded exon. The excluded/included exon specific combinatorial patterns of histone modifications can also be found in the preceding and succeeding intronic regions of the included and excluded exons (Supplementary Figures S1–S4).

Interaction Network of Histone Modifications

In order to investigate how histone modifications interact with each other and how their combinatorial fashions regulate RNA splicing, the Bayesian networks were constructed to deduce the causal relationships among histone modifications in the excluded/included exons and the surrounding intronic regions, respectively. In the Bayesian network, the nodes are the histone modifications, and the edge from one node to another one is their Pearson correlation coefficient.

The 10-fold cross-validation test method was used to find the robust Bayesian networks (Yu et al., 2008). The detailed procedure is as following. In the 10-fold cross-validation test, the dataset (Materials and Methods) is randomly partitioned into ten subsets, and nine of them were used to generate a Bayesian network. Based on the Pearson correlation coefficient of histone modifications of the nine subsets, a fundamental Bayesian network demonstrating the casual relationship between histone modifications was built by using the WinMine package. The 10-fold cross-validation was repeated 10 times. Accordingly, 10 fundamental Bayesian networks will be obtained for the excluded/included exons and the surrounding intronic regions, respectively. According to previous work (Yu et al., 2008), the final Bayesian network was then constructed based on the 10 fundamental Bayesian networks, in which each edge should be appeared within seven of the 10 fundamental Bayesian networks. The edges in the networks were colored according to the Pearson correlations between the two nodes linked by the edge.

It was found that the Bayesian network for the excluded exon event contains 19 edges and 10 combinatorial patterns of histone modifications, including interactions between different levels of the same modification (e.g., H3K79me1, H3K79me2,

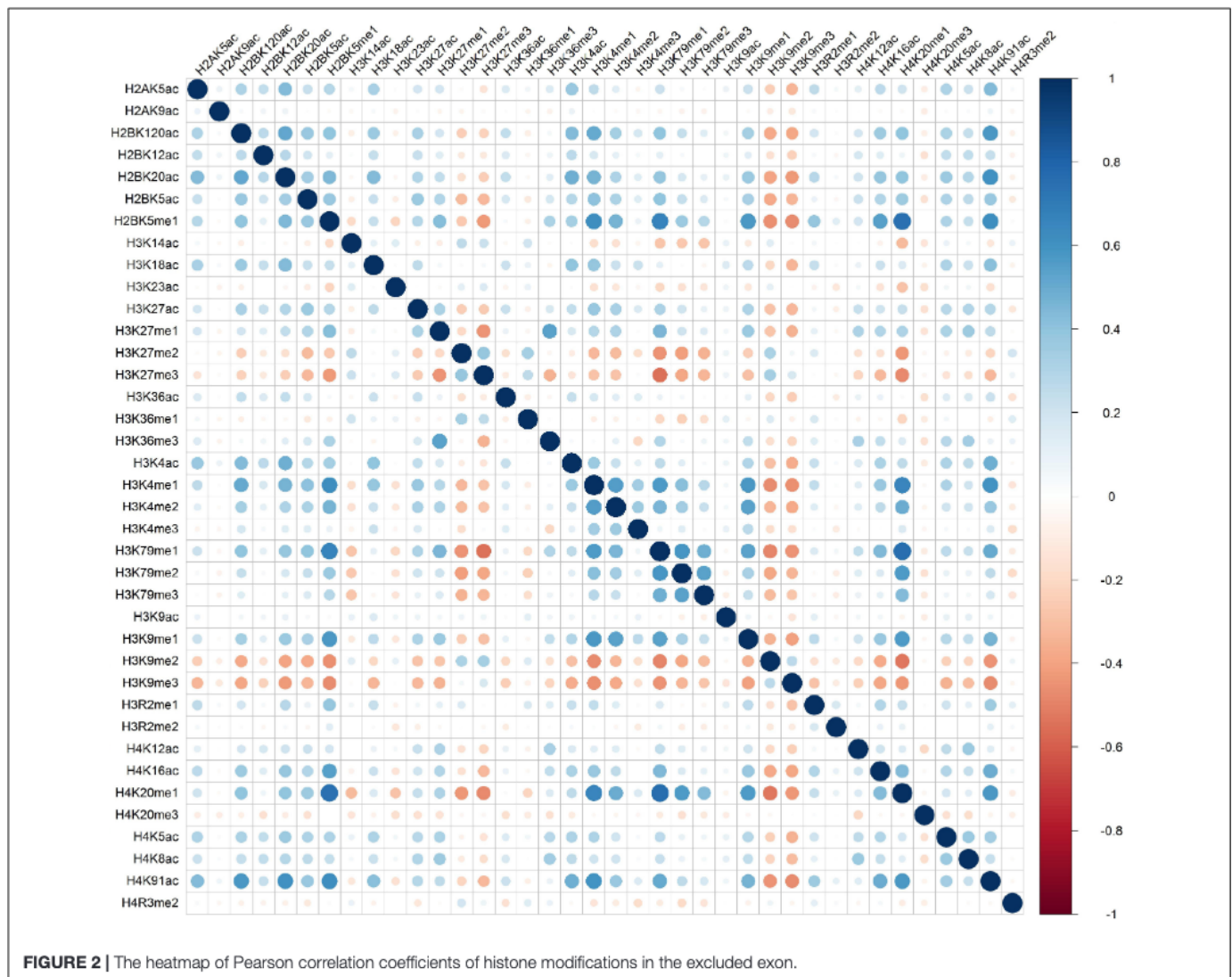


FIGURE 2 | The heatmap of Pearson correlation coefficients of histone modifications in the excluded exon.

and H3K79me3), between modifications on different amino acids (e.g., H2BK5me1 and H3K9me1), and between different kinds of modifications (e.g., H2BK5me1 and H4K16ac), **Figure 3A**. It can also be observed that the 10 histone modifications that have direct correlations with RNA splicing in excluded exon are H3K79me3, H3K79me2, H3K4me2, H4K16ac, H3K4me1, H3R2me1, H4K5ac, H2BK120ac, H3K18ac, and H3K4ac.

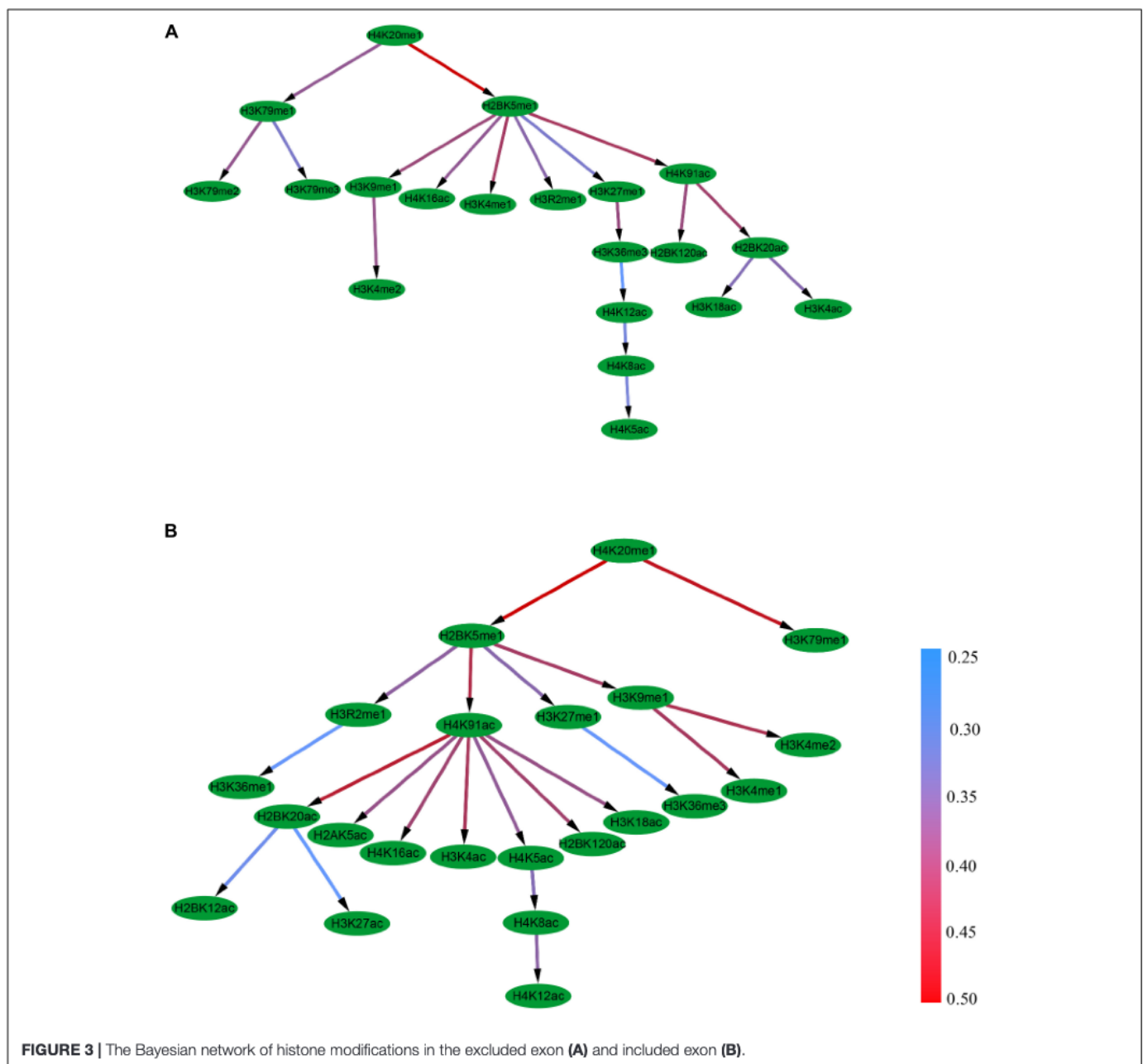
Distinct from the excluded exon, the Bayesian network for the included exon event contains 21 edges and 13 combinational patterns of histone modifications (**Figure 3B**). The interactions between modifications on different amino acids (e.g., H4K20me1 and H3K79me1), and between different kinds of modifications (e.g., H2BK5me1 and H4K91ac) were observed in this case. There are 13 histone modifications (H3K79me1, H3K36me3, H3K36me1, H3K4me1, H3K4me2, H2BK12ac, H3K27ac, H2AK5ac, H4K16ac, H3K4ac, H4K12ac, H2BK120ac, and H3K18ac) that have direct correlations with RNA splicing.

The above results demonstrate that the topologies of the Bayesian networks of histone modifications for the included

and excluded exon in the skipping event are different. Moreover, the differences also exist in the proceeding and succeeding intronic regions of the included and excluded exons (**Supplementary Figures S5–S6**). Therefore, it can be concluded that the casual relationship of histone modifications were obviously different between included and excluded exons.

CONCLUSION

Based on the Pearson correlation coefficients, the casual relationships of histone modifications in the process of RNA splicing were deduced by constructing their Bayesian networks. The results indicate that the inclusion or exclusion of exons is influenced by combinational patterns of histone modifications (**Figure 3** and **Supplementary Figures S5–S6**). Some of the histone modifications contribute directly to RNA splicing (e.g., H3K36me3 and H3K79me1), while other histone modifications indirectly contribute to the RNA splicing.



The result that H3K36me3 and H3K79me1 can affect RNA splicing is consistent with previous studies which have demonstrated that H3K36me3 and H3K79me1 are enriched in included exons (Shindo et al., 2013). The H3K36me3 can regulate alternative splicing by interacting with polypyrimidine tract-binding protein (PTB) (Luco et al., 2010). By interacting with the Tudor domain of TP53BP1, the H3K79me1 was also reported to interact with that interacts with snRNP (Huyen et al., 2004; Shindo et al., 2013).

By relaxing the chromatin structure, the H3 and H4 acetylation were also reported to regulating inclusion or exclusion of the skipping exon (Zhou et al., 2011). Besides the histone modifications located in exon regions, histone modifications located in intragenic regions can also influence RNA splicing by

regulating RNAPII elongation rates, or by directly binding to splicing factors and hence mediating their binding to pre-mRNA (Gomez Acuna et al., 2013).

Since there is no evidence for some of the histone modifications how they regulate RNA splicing, further experiments are needed in order to illustrate their roles in RNA splicing regulation. Taken together, we hope that this work could provide novel insights into the research on RNA splicing. Besides histone modifications, the method proposed here could also be used to analyze the relationship between RNA splicing with other modifications, such as m6A (Chen et al., 2015; Chen et al., 2018a; Wei et al., 2019), m4C (Chen et al., 2017; He et al., 2018), phosphorylation (Wei et al., 2017), GlcNAcylation (Jia et al., 2018), etc.

AUTHOR CONTRIBUTIONS

WC and HL conceived and designed the experiments. HL, XS, and WC wrote the manuscript. All authors performed the experiments, read, and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (31771471 and 61772119), Natural Science

Foundation for Distinguished Young Scholar of Hebei Province (No. C2017209244), and the Program for the Top Young Innovative Talents of Higher Learning Institutions of Hebei Province (No. BJ2014028).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00122/full#supplementary-material>

REFERENCES

- Badr, E., ElHefnawi, M., and Heath, L. S. (2016). Computational identification of tissue-specific splicing regulatory elements in human genes from RNA-Seq data. *PLoS One* 11:e0166978. doi: 10.1371/journal.pone.0166978
- Badr, E., and Heath, L. S. (2015). CoSREM: a graph mining algorithm for the discovery of combinatorial splicing regulatory elements. *BMC Bioinformatics* 16:285. doi: 10.1186/s12859-015-0698-6
- Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., et al. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837. doi: 10.1016/j.cell.2007.05.009
- Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72, 291–336. doi: 10.1146/annurev.biochem.72.121801.161720
- Chen, W., Ding, H., Zhou, X., Lin, H., and Chou, K. C. (2018a). iRNA(m6A)-PseDNC: identifying N(6)-methyladenosine sites using pseudo dinucleotide composition. *Anal. Biochem.* 56, 59–65. doi: 10.1016/j.ab.2018.09.002
- Chen, W., Feng, P., Ding, H., and Lin, H. (2018b). Classifying included and excluded exons in exon skipping event using histone modifications. *Front. Genet.* 9:433. doi: 10.3389/fgene.2018.00433
- Chen, W., Feng, P., Ding, H., Lin, H., and Chou, K.-C. (2015). iRNA-Methyl: identifying N-6-methyladenosine sites using pseudo nucleotide composition. *Anal. Biochem.* 490, 26–33. doi: 10.1016/j.ab.2015.08.021
- Chen, W., Lin, H., Feng, P., and Wang, J. (2014). Exon skipping event prediction based on histone modifications. *Interdiscip. Sci.* 6, 241–249. doi: 10.1007/s12539-013-0195-4
- Chen, W., Yang, H., Feng, P., Ding, H., and Lin, H. (2017). iDNA4mC: identifying DNA N⁴-methylcytosine sites based on nucleotide chemical properties. *Bioinformatics* 33, 3518–3523. doi: 10.1093/bioinformatics/btx479
- Cui, X. J., Li, H., and Liu, G. Q. (2011). Combinatorial patterns of histone modifications in *Saccharomyces cerevisiae*. *Yeast* 28, 683–691. doi: 10.1002/yea.1896
- David, C. J., Chen, M., Assanah, M., Canoll, P., and Manley, J. L. (2010). HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* 463, 364–368. doi: 10.1038/nature08697
- Enroth, S., Andersson, R., Wadelius, C., and Komorowski, J. (2010). SICTIN: rapid footprinting of massively parallel sequencing data. *BioData Min.* 3:4. doi: 10.1186/1756-0381-3-4
- Enroth, S., Bornelov, S., Wadelius, C., and Komorowski, J. (2012). Combinations of histone modifications mark exon inclusion levels. *PLoS One* 7:e29911. doi: 10.1371/journal.pone.0029911
- Garcia Blanco, M. A., Baraniak, A. P., and Lasda, E. L. (2004). Alternative splicing in disease and therapy. *Nat. Biotechnol.* 22, 535–546. doi: 10.1038/nbt964
- Gomez Acuna, L. I., Fiszbein, A., Allo, M., Schor, I. E., and Kornblihtt, A. R. (2013). Connections between chromatin signatures and splicing. *Wiley Interdiscip. Rev. RNA* 4, 77–91. doi: 10.1002/wrna.1142
- He, W., Jia, C., and Zou, Q. (2018). 4mCPred: machine learning methods for DNA N⁴-methylcytosine sites prediction. *Bioinformatics* [Epub ahead of print] doi: 10.1093/bioinformatics/bty1668
- Huyen, Y., Zgheib, O., Ditullio, R. A. Jr., Gorgoulis, V. G., Zacharatos, P., Petty, T. J., et al. (2004). Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* 432, 406–411. doi: 10.1038/nature03114
- Jia, C., Zuo, Y., and Zou, Q. (2018). O-GlcNAcPred-II: an integrated classification algorithm for identifying O-GlcNAcylation sites based on fuzzy undersampling and a K-means PCA oversampling technique. *Bioinformatics* 34, 2029–2036. doi: 10.1093/bioinformatics/bty039
- Kornblihtt, A. R., Schor, I. E., Allo, M., Dujardin, G., Petrillo, E., and Munoz, M. J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* 14, 153–165. doi: 10.1038/nrm3525
- Lai, T. S., and Greenberg, C. S. (2013). TGM2 and implications for human disease: role of alternative splicing. *Front. Biosci.* 18:504–519.
- Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science* 327, 996–1000. doi: 10.1126/science.1184208
- Mele, M., Mattioli, K., Mallard, W., Shechner, D. M., Gerhardinger, C., and Rinn, J. L. (2017). Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res.* 27, 27–37. doi: 10.1101/gr.214205.116
- Nellore, A., Jaffe, A. E., Fortin, J. P., Alquicira-Hernandez, J., Collado-Torres, L., Wang, S., et al. (2016). Human splicing diversity and the extent of unannotated splice junctions across human RNA-seq samples on the sequence read archive. *Genome Biol.* 17:266. doi: 10.1186/s13059-016-1118-6
- Oberdoerffer, S., Moita, L. F., Neems, D., Freitas, R. P., Hacohen, N., and Rao, A. (2008). Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPL. *Science* 321, 686–691. doi: 10.1126/science.1157610
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415. doi: 10.1038/ng.259
- Saint-Andre, V., Batsche, E., Rachez, C., and Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1gamma favor inclusion of alternative exons. *Nat. Struct. Mol. Biol.* 18, 337–344. doi: 10.1038/nsmb.1995
- Scotti, M. M., and Swanson, M. S. (2016). RNA mis-splicing in disease. *Nat. Rev. Genet.* 17, 19–32. doi: 10.1038/nrg.2015.3
- Shindo, Y., Nozaki, T., Saito, R., and Tomita, M. (2013). Computational analysis of associations between alternative splicing and histone modifications. *FEBS Lett.* 587, 516–521. doi: 10.1016/j.febslet.2013.01.032
- Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476. doi: 10.1038/nature07509
- Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddapah, S., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* 40, 897–903. doi: 10.1038/ng.154
- Wei, L., Su, R., Wang, B., Li, X., and Zou, Q. (2019). Integration of deep feature representations and handcrafted features to improve the prediction of N⁶-methyladenosine sites. *Neurocomputing* 324, 3–9. doi: 10.1016/j.neucom.2018.04.082
- Wei, L., Xing, P., Tang, J., and Zou, Q. (2017). PhosPred-RF: a novel sequence-based predictor for phosphorylation sites using sequential information only. *IEEE Trans. Nanobioscience* 16, 240–247. doi: 10.1109/tnb.2017.2661756

- Yu, H., Zhu, S., Zhou, B., Xue, H., and Han, J. D. (2008). Inferring causal relationships among different histone modifications and gene expression. *Genome Res.* 18, 1314–1324. doi: 10.1101/gr.073080.107
- Zhou, H. L., Hinman, M. N., Barron, V. A., Geng, C., Zhou, G., Luo, G., et al. (2011). Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner. *Proc. Natl. Acad. Sci. U.S.A.* 108, E627–E635. doi: 10.1073/pnas.1103344108
- Zhu, S., Wang, G., Liu, B., and Wang, Y. (2013). Modeling exon expression using histone modifications. *PLoS One* 8:e67448. doi: 10.1371/journal.pone.0067448

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Chen, Song and Lin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.