FEBS Letters 587 (2013) 516–521

FEBS Letters

journal homepage: www.FEBSLetters.org

Computational analysis of associations between alternative splicing and histone modifications

Yuki Shindo [a,b], Tadasu Nozaki [a,b], Rintaro Saito [a,*,1], Masaru Tomita [a,b]

[a] Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0017, Japan
[b] Systems Biology Program, Graduate School of Media and Governance, Keio University, Fujisawa 252–8520, Japan

ARTICLE INFO

Article history:
Received 3 November 2012
Revised 13 January 2013
Accepted 17 January 2013
Available online 24 January 2013

Edited by Takashi Gojobori

Keywords:
Splicing
Histone modification
Computational analysis

ABSTRACT

Pre-mRNA splicing is a complex process involving combinatorial effects of *cis*- and *trans*-elements. Here, we focused on histone modifications as typical *trans*-regulatory elements and performed systematic analyses of associations between splicing patterns and histone modifications by using publicly available ChIP-Seq, mRNA-Seq, and exon-array data obtained in two human cell lines. We found that several types of histone modifications including H3K36me3 were associated with the inclusion or exclusion of alternative exons. Furthermore, we observed that the levels of H3K36me3 and H3K79me1 in the cell lines were well correlated with the differences in alternative splicing patterns between the cell lines.

1. Introduction

Alternative splicing is often finely regulated in a specific manner according to cell type, developmental stage or both, and dysfunction in alternative splicing is associated with several diseases [1]. NGS analyses of the human whole transcriptome have revealed that more than 90% of human genes undergo alternative splicing [2]. However, the mechanisms by which the correct exons are selected in cell type- or stage-specific manners remain unclear.

Traditionally, the regulation of alternative splicing has been thought to have been achieved by splicing enhancers and silencers, which are short RNA sequences located either in exons or introns. In fact, the sum of all splicing-related features in pre-mRNA sequences (i.e., splicing code) explains most of the differences in alternative splicing among several tissues [3]. Conversely, recent studies have reported that post-translational modification of the histone protein can regulate alternative splicing patterns in humans [4]. Histone is the core protein of the nucleosome, which is the basic unit of chromatin structure, and variations in the type of histone modifications and the genomic position of histones influences genomic functions [5]. Furthermore, genome-wide mapping of nucleosomes and modified histones has revealed their non-random distribution around exons, with exonic regions having high levels of nucleosome occupancy and modified histones compared with intronic regions [6–8]. For example, the level of H3K36me3 correlates with that of exon expression, and the expression level of each exon (either high or low) can be classified according to the patterns of combinatorial histone modifications with a maximum accuracy of ~78% [9,10]. These facts strongly suggest a role for histone modifications in the regulation of pre-mRNA splicing, although the extent of the association remains unclear. For instance, the following questions remain unanswered: Which types of histone modifications are enriched in which alternative splicing patterns? How often do changes in histone modifications within a specific gene change its splicing pattern among multiple cell types?

Here, we systematically analyzed the relationship between alternative splicing and histone modifications. We demonstrated not only global associations between exon inclusion/exclusion patterns and histone modifications but also global changes of histone modification profiles corresponding to cell-specific exon usages.

2. Materials and methods

2.1. Annotation datasets

We downloaded the Illumina iGenomes package (GRCh37), which is a collection of sequence and annotation files, from

Abbreviations: H2BK12ac, acetylated histone H2B at Lys12; H3K36me3, trimethylated histone H3 at Lys36; H3K4me1, mono-methylated histone H3 at Lys4; H3K4me2, di-methylated histone H3 at Lys4; H3K4me3, tri-methylated histone H3 at Lys4; H3K79me1, mono-methylated histone H3 at Lys79; H3K9ac, acetylated histone H3 at Lys9; H4K5ac, acetylated histone H4 at Lys5; NI, normalized intensity; NGS, next generation sequencing; NPS, nucleosome positioning from sequencing; Pol II, RNA polymerase II; SI, splicing index

* Corresponding author. Fax: +81 235 29 0525.
E-mail address: rsaito@sfc.keio.ac.jp (R. Saito).
[1] Current address: Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA.

0014-5793/$36.00
http://dx.doi.org/10.1016/j.febslet.2013.01.032

http://tophat.cbcb.umd.edu/igenomes.html. ChIP-Seq data for histone modifications in H1 IMR90, fetal brain cells, and fetal lung cells were obtained from [11]. The histone ChIP-Seq data were processed by using NPS [12] and positions emerging as significant peaks were determined ($P < 1e\text{-}5$; Table S1). We regarded the center of the peak of ChIP-Seq data as the position of the corresponding histone. Note that we excluded X and Y chromosome datasets from the analyses because the gender origins of the cells were different.

2.2. Characterization of transcripts by using mRNA-Seq

We downloaded the strand-specific shotgun sequencing reads for mRNAs in H1 and IMR90 cells from [11]. Reads were aligned to the human genome (GRCh37) and known splice junctions by using TopHat [13]. To characterize exon-skipping events, we considered sets of three consecutive exons (exon trios) and possible exon junctions generated by splicing of these exons, and focused on whether the second exon in each set was excluded or not. To avoid capturing chromatin features of the adjacent exonic regions, only exon trios with intron lengths of $\geqslant$500 bp were selected [14]. We also discarded exons whose length were <50 bp because of difficulties in aligning reads to short exons. For each exon trio, we counted the number of reads that aligned to the junction of the first and third exons, which implies that the second exon was spliced out (mature transcript X–Z, Fig. 1) and the number of aligned reads to the junctions of first and second exons, or second and third exons, which implies that the second exon was included in the mature transcript (mature transcript X–Y–Z, Fig. 1). We defined these counts as the "exclusion score" and the "inclusion score", respectively; inclusion scores were divided by two because the RNA-Seq can be mapped to two exon–intron junctions for the mature transcript X–Y–Z. In the case where the sum of two scores was $\leqslant$10, the corresponding exon trio was discarded from the analysis. We then calculated the exclusion rate as the ratio of exclusion score/(exclusion score + inclusion score) and assumed that an exon-skipping event had occurred only for cases where the exclusion rate was $\geqslant$0.95. With this criterion, we extracted 716 and 554 final sets of exons in H1 and IMR90 cells, respectively. For the following analysis, we labeled the second exons within exon trios as "excluded" and the third exons as "included". We did not use the first exons because of the possibility that there were either annotated or unannotated promoters near the first exons, and histone modification marks might reflect features associated with transcription initiation rather than splicing.

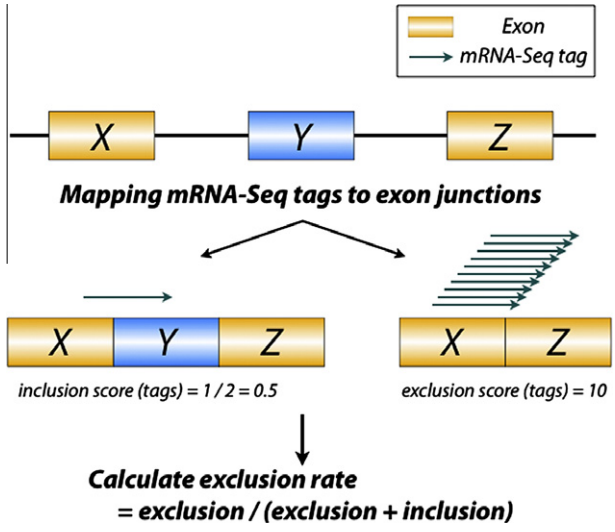

Fig. 1. Characterization of the exon skipping events by using mRNA-Seq. The inclusion score and exclusion score were defined based on the profile of mapped reads onto the exon–exon junction. The exclusion rate indicates how frequently exon skipping events occur and was defined as exclusion score/(exclusion score + inclusion score).

2.3. Comparison of the histone modification profiles between included exons and excluded exons within cell types

For each cell type and each type of histone modification, we calculated the fraction of exons that had at least one peak of modified histones within ±200 bp from their intron–exon boundaries. We performed the chi-squared test for independence followed by the Bonferroni correction to evaluate the significance of differences in this fraction between excluded and included exons ($P < 0.01$). For comparison, we randomly extracted 10000 sites (400 bp length) from the gene bodies and intergenic regions respectively, and calculated the fractions in the same fashion. We defined the relative frequency as (values of the fraction in the vicinity of intron–exon boundaries)/(values of the fraction in randomly selected regions from the gene body).

2.4. Calculation of SI values

We downloaded the exon array data pertaining to H1, IMR90, fetal brain and fetal lung cells from NCBI GEO (accessions GSE14863 and GSE18927). To assess the cell-specific alternative splicing, we adopted SI, which is a simple but efficient method for detecting alternative splicing events between samples [15]. SI is defined as the log fold change between two normalized intensities (NIs), where NI is defined as the ratio of exon expression/gene expression. There is a caveat that an SI value around 0 may imply either that the exon is retained in both cell lines or is spliced out in both cell lines. Because of this, we only included data with an absolute value of SI of >1 (i.e., at least two-fold change) in further analyses.

3. Results

3.1. Characterization of exon-skipping events and histone positioning

To investigate the relationship between splicing and histone modification, we first needed to characterize splicing events and the distribution of various modified histones in a genome-wide analysis. For splicing analysis, we focused on mRNA-Seq reads that were mapped to exon–exon junctions because they strongly reflect the structure of the mature transcript (Fig. 1). We assessed how often each exon in a set of three consecutive internal exons (exon trio) was alternatively spliced out (exclusion rate). On the basis of our definition of the exclusion rate, a value close to 0 indicates that the corresponding exon is included in the mature transcript in the cell, whereas a value close to 1 indicates that the exon is spliced out.

About 90% of exons showed a value close to 0, and most of the rest (~6%) showed a value close to 1 (Fig. S1). We regarded each exon with an exclusion rate of $\geqslant$0.95 as an "exon-skipping event" and defined the second exon of the trio in the event (i.e., exon Y in Fig. 1) as the "excluded exon" and the third exon (i.e., exon Z in Fig. 1) as the "included exon" (Section 2).

3.2. Associations between histone modifications and splicing patterns

Previous studies reported that several types of modifications (e.g., H3K36me3 and H3K79me1) were considerably enriched in exons compared with introns. These observations imply that a relationship exists between histone modifications and splicing patterns, although the extent of the association is still unclear.

We focused on the vicinity of intron–exon boundaries and calculated the fraction of exons harboring each particular modified histone.

We found that H3K36me3 was significantly enriched around intron–exon boundaries of included exons compared with excluded exons ($P < 0.01$; chi-squared test; Fig. 2). The observed trend was similar in H1 and IMR90 cells. This finding is consistent with a previous report showing that the change of H3K36me3 level around exons affects the pattern of alternative splicing, although only a few genes were examined in the previous study [4]. We also found that several types of histone modifications (e.g., H3K4me3, H2BK12ac, H4K5ac) around the intron–exon boundary were significantly enriched in excluded exons compared with included exons (Fig. 2). These results imply that histone modifications are associated not only with exon inclusion but also exon exclusion. Interestingly, there were several types of histone modifications where clear differences in frequencies were observed only in one of the two cell lines (H1 or IMR90), which might suggest a cell-specific function for these histone modifications (Fig. S2).

3.3. Relationship between histone modifications and alternative splicing in different cell lines

The above analyses explored exon inclusion within each of the two cell types. We next investigated the relationship between histone modifications and alternative splicing patterns by comparing different exon usage between the two cell types. We used publicly available exon-array data instead of mRNA-Seq data because very few exons that are alternatively spliced in H1 and IMR90 cells could be identified by using mRNA-Seq data. We assessed whether an exon that is included in one cell line is excluded in the other cell line by using the SI, which was defined as $SI = \log_2 (NI_{H1}/NI_{IMR90})$, where NI denotes normalized intensity of the exon in each cell line. In this study, a small value of SI indicates that the exon is specifically included in the IMR90 cell line and excluded in H1 cell line, whereas a large value indicates the opposite. We binned the sets of exons based on their SI value, with a bin interval of 0.25, and calculated the frequency of the exons harboring modified histones for each bin.

We found that the frequency of histone modifications around exons and the SI values of the exons were strongly associated for two types of histone modifications, H3K36me3 and H3K79me1 (Fig. 3). For example, exons having low SI values (i.e., those presumed to be included in IMR90 cells but excluded in H1 cells) harbored H3K36me3 more frequently in IMR90 than in H1. Conversely, exons having high SI values (i.e., those presumed to be included in H1 cells but excluded in IMR90 cells), harbored more H3K36me3 in H1 cells than in IMR90 cells ($r = 0.73$, -0.76 in H1 and IMR90, respectively). A similar association was observed for H3K79me1 (Fig. 3B, $r = 0.59$, -0.70 in H1 and IMR90, respectively) although significant differences were not observed between completely included or excluded exons (Fig. S3). Then, we investigated whether these trends could be observed under physiological conditions, for example, in primary cells. Although the publicly available datasets were limited and we used only H3K36me3 ChIP-Seq data, we successfully confirmed that H3K36me3 levels around exons were correlated with alternative splicing patterns in fetal brain and lung cells (Fig. S4, $r = 0.79$, -0.52 in fetal brain and fetal lung, respectively). These results further support the hypothesis that genome-wide differences in histone modifications are tightly associated with cell-specific alternative patterns.

Apart from H3K36me3 and H3K79me1, we did not find any clear correlations between frequency of modified histones and differences in alternative splicing patterns between the two cell types.

3.4. An example showing combinatorial histone modifications around spliced exons

Many splicing events would not be determined solely by a single type of histone modification [16]. There may be multiple histones having different modifications around an exon, and in such a case, the effect of one type of modification may be suppressed by other types of modification [9]. Fig. 4 shows histone modifications and splicing patterns in IMR90 cells in the genomic region of the *BIN1* gene as an example. The second exon from the left side was skipped, and H3K36me3 was specifically depleted around this

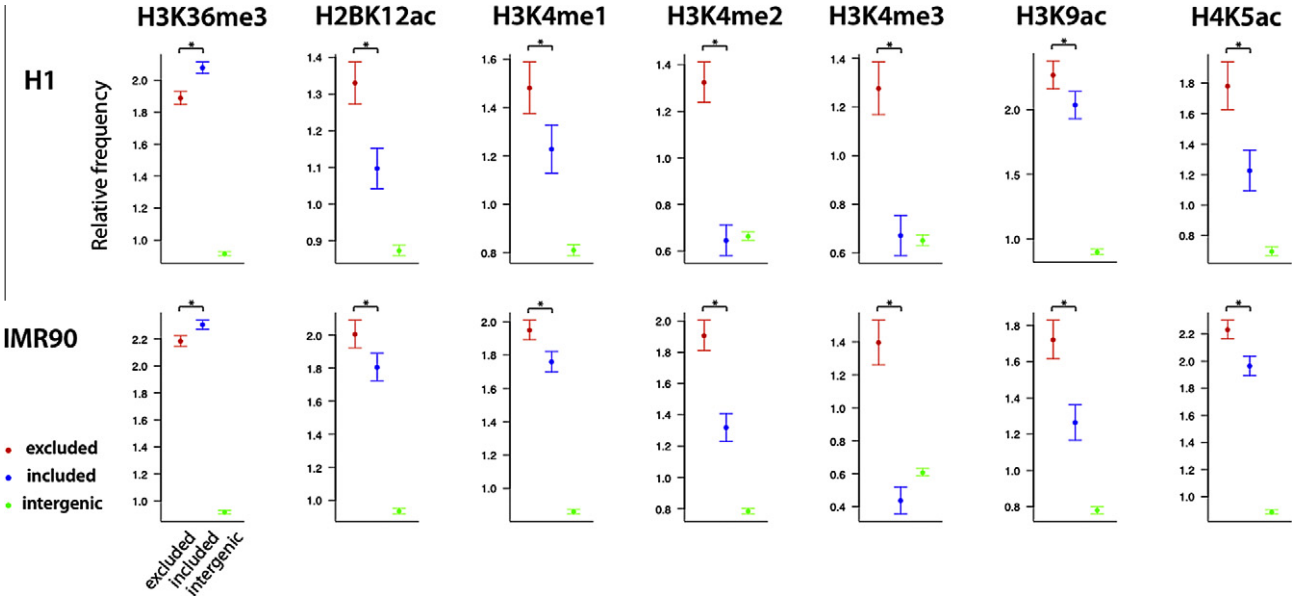

Fig. 2. Associations between histone modifications and exon inclusion patterns. Exons were separated into "excluded" (red) and "included" (blue) groups, and the frequencies of exons harboring modified histones within ±200 bp from the intron–exon boundary were calculated. As a control, we randomly selected 10 000 regions from the gene body and intergenic regions (green), respectively. The frequencies for each group were normalized by the values for gene body regions in the same cell type to give relative frequencies. Values represent means ± SD. *$P < 0.01$.

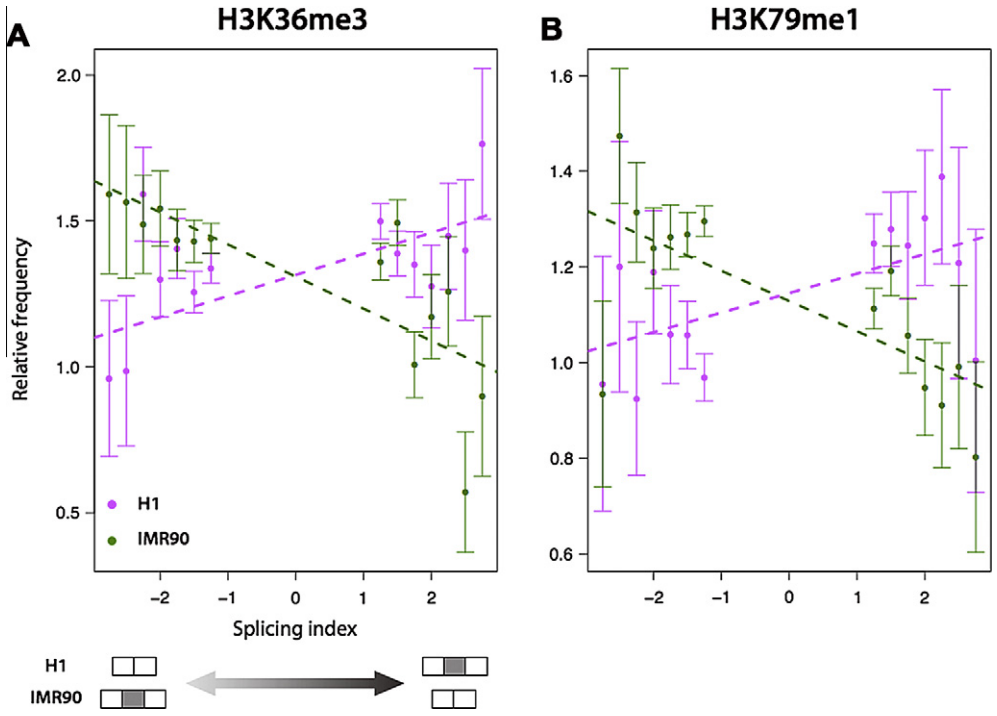

Fig. 3. Differences in the frequencies of modified histones around alternatively spliced exons between cell types. The *x*-axis indicates the SI and the *y*-axis represents the relative frequency of exons having H3K36me3 (A) or H3K79me1 (B) around their intron–exon boundaries. Exons were binned based on the values of SI, with a bin interval of 0.25, and relative frequencies were calculated for each binned group. Relative frequencies observed for H1 and IMR90 cells are shown in magenta and green, respectively. Values represent means ± SD.

exon but enriched around the included exons. However, there were two excluded exons (sixth and seventh exons) around which a number of peaks corresponding to H3K36me3 were observed. Interestingly, in this region, H2BK12ac, H4K5ac, and H3K4me3 were also enriched, which might suggest competition between modifications that enhance splicing and those that silence splicing. Similar examples observed in H1 are shown in Fig. S5.

4. Discussion

In this study, we performed a comprehensive analysis of the relationships between splicing and histone modifications. First, we used mRNA-Seq data to estimate which exons were excluded from mature mRNA. On the basis of this analysis, most of the exons were predicted to be either completely included or completely excluded, which may seem counter-intuitive; exon inclusion rates presumably span over a continuous range. This might be because currently available NGS technology is not efficient enough to quantitatively determine the precise amount of isoforms in the cell and thus the calculated balance of inclusion/exclusion for each exon based on NGS data may be biased. In fact, NGS data-based estimation of differential exon usage among samples remains a challenging problem [17,18]. Therefore, we avoided using mRNA-Seq data for our alternative splicing analysis. Nonetheless, mRNA-Seq is a helpful method for identifying the exon structure of mature transcripts on the basis of the junction reads [2]. Thus, although the sensitivity of our method to detect exon structure, which is dependent on exon junction reads, seemed to be limited, causing the number of extracted exon structures to be relatively low, we believe that at least the exon-skipping events extracted by our criterion were reliable (for example, exons shown in Fig. 4).

H3K36me3 and H3K4me3 have been reported to interact with the splicing machinery via adaptor protein MRG15 [4]. In addition, this study reported that H3K36me3 and H3K4me3 had opposite effects on exon inclusion/exclusion, which is consistent with our results (Fig. 2). In addition to these two types of histone modification, we also found other types of modified histones such as H2BK12ac and H4K5ac although there is no evidence of how these histone modifications are associated with splicing. There are several possible scenarios, including interaction with unidentified splicing-associated proteins or association with Pol II [19–21], which require further experimental investigation.

We also found that exons that were alternatively spliced out in a cell-specific manner tended to have less of the H3K36me3 and H3K79me1 modified histones (Fig. 3 and S4). This is an exceptional result in that it shows the genome-wide relationship between the histone modification landscape and alternative splicing patterns among cell types. Similar results were also observed in two types of primary cells. As mentioned previously, H3K36me3 is known to interact with splicing machinery [4]. The other modification type, H3K79me1, was reported to interact with the Tudor domain of TP53BP1 that interacts with snRNP [22]. These facts support the hypothesis that both cell-specific gene expression and alternative splicing patterns are globally regulated by epigenetics.

Several histone modifications showed significant differences in Fig. 2 but not in Fig. 3 and vice versa. In Fig. 2, we selected mRNA-Seq data with relatively strict criteria to accurately extract transcripts with skipped exons. As a result, the number of extracted exons was low and we frequently found identical splicing patterns between H1 and IMR90 (Fig. S6). Conversely, exons analyzed in Fig. 3 were determined based on the criterion of whether the exon was differentially spliced out between two cell types and we used exon array data instead of mRNA-Seq data due to the current limitation of quantifying exon expression levels by mRNA-Seq, as stated earlier. Therefore, the discrepancy between Figs. 2 and 3 could have occurred because the characteristics of the two sets of exons extracted based on these two distinct criteria were very different even if the exons harbored the same types of histone modification

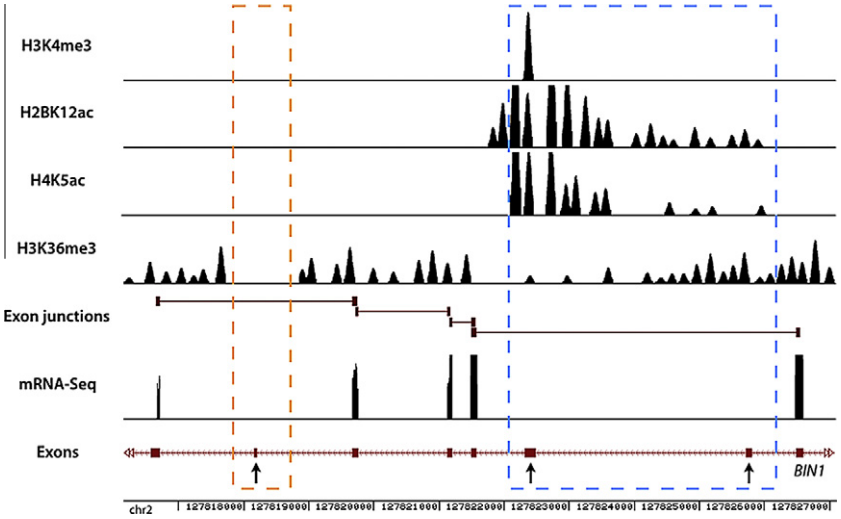

Fig. 4. An example of a genomic region showing associations between histone modifications and alternative splicing events. Exons marked with the black arrow in the dotted boxes were alternatively excluded from the mature transcript in IMR90 cells.

in both cell types. Another possible explanation is that the results for H3K79me1 could be interpreted as H3K79me1 not acting as an accurate "switch" within a particular cell type, and somehow changing the probability of exons being included among different cell types. In the future, when NGS technology provides sufficiently accurate quantitative exon expression levels, these analyses should be repeated with the same datasets and using the same criteria.

Finally, it is worth noting that recent studies showed that combinations of *cis*- and *trans*-elements are extremely important for the regulation of splicing [16]. In other words, histone modifications cannot explain the complete splicing patterns; in fact, a number of included exons showed low levels of H3K36me3 [6,9]. Moreover, bidirectional communication between the histone modification and splicing machinery has been reported recently. For example, splicing can enhance the recruitment of H3K36 methyltransferase HYPB/Setd2, resulting in significant differences in the enrichment of H3K36me3 between intron-containing and intronless genes [23]. Thus, it is important to pay close attention to the cause and effect. In addition, there is the situation where the chromatin state has little influence on the splicing event because the regulation of splicing occurs both co- and post-transcriptionally; the selection of the exon must be determined by a mixture of *cis*- and *trans*-factors, although the whole picture has not yet been unveiled. Nevertheless, our findings of non-random distribution of modified histones across exons and global differences in histone modification profiles between cells corresponding to differences in alternative splicing patterns (Fig. 3) suggest that at least some part of the specific alternative splicing patterns are regulated in an epigenetic manner, similar to the way that the cell type- and stage-specific chromatin state around transcription start sites drives the activation or repression of specific gene expression. Taken together, it may be possible to estimate alternative splicing patterns more precisely by combining splicing motifs (i.e., genome-level information) and histone modification profiles (i.e., epigenome-level information), as the "extended splicing code".

Acknowledgment

This research was supported by research funds from the Yamagata Prefectural Government and Tsuruoka City, Japan.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.febslet.2013.01.032.

References

[1] Licatalosi, D.D. and Darnell, R.B. (2010) RNA processing and its regulation: global insights into biological networks. Nat. Rev. Genet. 11, 75–87.
[2] Wang, E.T. et al. (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470–476.
[3] Barash, Y., Calarco, J.A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B.J. and Frey, B.J. (2010) Deciphering the splicing code. Nature 465, 53–59.
[4] 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.
[5] Kouzarides, T. (2007) Chromatin modifications and their function. Cell 128, 693–705.
[6] Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C. and Komorowski, J. (2009) Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Res. 19, 1732–1741.
[7] Dhami, P. et al. (2010) Complex exon–intron marking by histone modifications is not determined solely by nucleosome distribution. PLoS ONE 5, e12339.
[8] Schwartz, S., Meshorer, E. and Ast, G. (2009) Chromatin organization marks exon–intron structure. Nat. Struct. Mol. Biol. 16, 990–995.
[9] Enroth, S., Bornelov, S., Wadelius, C. and Komorowski, J. (2012) Combinations of histone modifications mark exon inclusion levels. PLoS ONE 7, e29911.
[10] Hon, G., Wang, W. and Ren, B. (2009) Discovery and annotation of functional chromatin signatures in the human genome. PLoS Comput. Biol. 5, e1000566.
[11] Bernstein, B.E. et al. (2010) The NIH roadmap epigenomics mapping consortium. Nat. Biotechnol. 28, 1045–1048.
[12] Zhang, Y., Shin, H., Song, J.S., Lei, Y. and Liu, X.S. (2008) Identifying positioned nucleosomes with epigenetic marks in human from ChIP-Seq. BMC Genomics 9, 537.
[13] Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111.
[14] Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S. and Ahringer, J. (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. Nat. Genet. 41, 376–381.
[15] Clark, T.A., Schweitzer, A.C., Chen, T.X., Staples, M.K., Lu, G., Wang, H., Williams, A. and Blume, J.E. (2007) Discovery of tissue-specific exons using comprehensive human exon microarrays. Genome Biol. 8, R64.
[16] Hertel, K.J. (2008) Combinatorial control of exon recognition. J. Biol. Chem. 283, 1211–1215.
[17] Anders, S., Reyes, A. and Huber, W. (2012) Detecting differential usage of exons from RNA-seq data. Genome Res. 22, 2008–2017.
[18] Hu, Y. et al. (2013) DiffSplice: the genome-wide detection of differential splicing events with RNA-seq. Nucleic Acids Res. 41, e39.
[19] Gunderson, F.Q. and Johnson, T.L. (2009) Acetylation by the transcriptional coactivator Gcn5 plays a novel role in co-transcriptional spliceosome assembly. PLoS Genet. 5, e1000682.

[20] Schor, I.E., Rascovan, N., Pelisch, F., Allo, M. and Kornblihtt, A.R. (2009) Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. Proc. Natl. Acad. Sci. USA 106, 4325–4330.
[21] Zhou, H.L., Hinman, M.N., Barron, V.A., Geng, C., Zhou, G., Luo, G., Siegel, R.E. and Lou, H. (2011) Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner. Proc. Natl. Acad. Sci. USA 108, E627–E635.
[22] Huyen, Y. et al. (2004) Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. Nature 432, 406–411.
[23] de Almeida, S.F. et al. (2011) Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. Nat. Struct. Mol. Biol. 18, 977–983.