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.

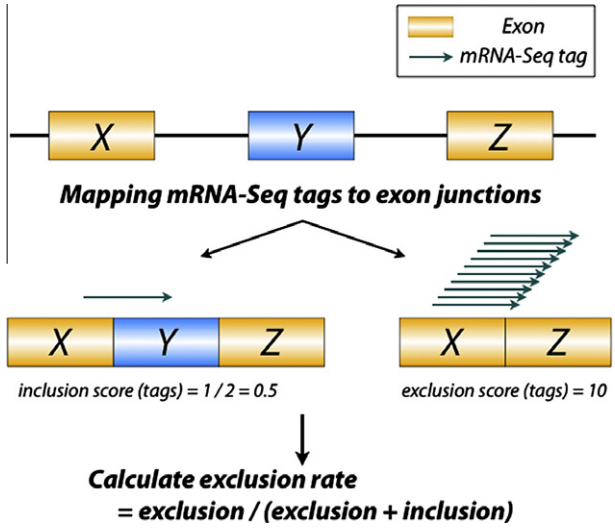

Fig. 1. Characterization of the exon skipping events by using mRNA-Seq. The inclusion score and exclusion score were defined based on the profile of mapped reads onto the exon–exon junction. The exclusion rate indicates how frequently exon skipping events occur and was defined as exclusion score/(exclusion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