

DNA hypermethylation of alternatively spliced and repeat sequences in humans

Andigoni Malousi · Sofia Kouidou

Received: 16 February 2012 / Accepted: 11 June 2012 / Published online: 28 June 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract DNA methylation is presently accepted as a tentative regulatory parameter in splicing. Recently, we reported significant methylation differences among various exonic splicing-enhancing elements and alternative splicing events, based on CpG methylation data from the Human Epigenome Project for chromosomes 6, 20 and 22. Presently, using a different computational approach and the same database, we report: (a) significant increase of hypermethylation in intronic and exonic sequences close to acceptor sites, relative to overall introns and exons, respectively (1,973 CpGs examined); (b) frequent CpGs, mostly hypomethylated, in donors and infrequent CpGs mostly hypermethylated, in acceptors; and (c) hypermethylation in cassette exons which are occasionally spliced and have weaker average splicing potential, relative to constitutive exons ($p < 0.0001$). CpGs are hypomethylated in non-coding exons (only 16 % hypermethylation). Single-exon genes, similarly to first exons, frequently contain hypomethylated CpGs, while in internal and last exons CpGs are more frequently hypermethylated. Methylation is also more frequent in strange introns and splice sites processed by the minor spliceosome, e.g., ATAC, ($p < 0.0001$ in all cases), but not in sites of incomplete processing, e.g., retained introns or bleeding exons, ($p = 0.706$ and $p = 0.313$, respectively). Most Alus, which are known to contribute to transcript presentation, are heavily

methyated, in contrast with other Alus, e.g., AluJo and mammalian interspersed repetitive elements which have been previously associated with alternative expression. These results elucidate the role of intragenic methylation in association with alternative splicing and facilitate the evaluation of genomic variations/polymorphisms and the development of tools for the prediction of alternative splicing events.

Keywords Methylation · Splice sites · Alternative splicing · SINEs · Alus · MIRs

Introduction

Although epigenetic mechanisms are actively investigated at present, our understanding of the impact of non-promoter DNA methylation is still very limited. Recently, it has become increasingly evident that DNA methylation might be a key parameter in the process of splicing (Luco et al. 2011). This novel role for DNA methylation, which is based on the finding that splicing is a co-transcriptional process (Pandya-Jones and Black 2009; Tilgner and Guigo 2010), is expected to provide a better understanding of the splicing mechanisms and differentiation of gene expression (Pandya-Jones and Black 2009).

In fact, the possible involvement of DNA methylation in the regulation of the splicing process has been recently addressed. Specifically, the studies of Chodavarapu et al. on *Arabidopsis thaliana* reveal that DNA methylation is highly enriched in exons and nucleosome positioning might play a significant role in DNA methylation (Chodavarapu et al. 2010), which, in turn, is associated with the selection of intron–exon boundaries (Schwartz and Ast 2010; Tilgner and Guigo 2010; Choi 2010). Transitions in

Communicated by S. Hohmann.

A. Malousi · S. Kouidou (✉)
Department of Biological Chemistry, School of Medicine,
Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
e-mail: kouidou@auth.gr

A. Malousi
e-mail: andigoni@med.auth.gr

the methylation state correlating with genomic landmarks such as intron–exon junctions were reported by Hodges et al. (2009) based on single molecule human DNA studies. Finally, the role of DNA methylation in alternative splicing is supported by the recent work of Shukla et al. (2011) in CD45 cells. In this work, the authors showed that the activity of the CCCTC-binding factor (CTCF), which can promote inclusion of weak upstream exons by mediating local RNA polymerase II pausing, can be reciprocated by DNA methylation in a specific exon.

Correlation of the CpG density/methylation state with expression in alternatively spliced sequences is therefore critical for elucidating the epigenetic regulatory mechanisms in splicing. CpG-dense motifs (Malousi et al. 2008) and specific Alu repeats, e.g., AluJb, AluSx and AluJo (Sorek et al. 2002, 2004) have been strongly associated with alternative splicing. In addition, the latter have been proposed to cause frame-shifting or premature termination codons, reduced expression frequencies (Lin et al. 2008) in different pathological conditions, such as X-linked dilated cardiomyopathy (Rimessi et al. 2010), ATM (Pastor et al. 2009) and other human diseases (Vorechovsky 2010). Similarly, mammalian interspersed repeats (MIRs) have been associated with frequently diverse, translational regulation (Huh et al. 2008; Lin et al. 2008).

These observations are in support of a new model for the regulation of splicing, in which DNA methylation plays an important role in constitutive and alternative splicing. More specifically, chromatin remodeling via chromatin-binding proteins such as CTCF, or other protein factors appears to be also related to polymerase stalling, which in turn, is necessary for transcriptional elongation and splicing (Tao et al. 2010; Shukla et al. 2011; Alexander et al. 2010; Oberdoerffer 2012). In addition, nucleosomal positioning, which has been associated with polymerase activity regulation, is reported to be preferentially associated with methylated sequences (Chodavarapu et al. 2010; Shukla et al. 2011). This data, which combines nucleosomal positioning and polymerase activity with splicing efficiency and DNA methylation reveals an immersing role for DNA methylation, related to splice site selection, and possibly alternative splicing. Such a model would in turn provide a better interpretation of disease-related mutations and polymorphism selection (Karambataki et al. 2010). However, to date, there is little direct information from large-scale data on the association of methylation with specific alternative splicing types, splice sites and short interspersed nuclear elements (SINEs), as well as the distribution of methylation in splice sites.

A recent analysis (Anastasiadou et al. 2011) based on the methylation data obtained for 33,352 CpG sequences from different human tissues reported by the Human Epigenome Project (HEP), showed that DNA methylation is

significantly associated with alternative splicing events and varies among different categories of exonic motifs (exonic splicing enhancers), which play a regulatory role in splicing. Using a different computational procedure and the same raw data from a variety of tissues, we presently investigated CpG methylation in introns, splice sites, and different SINEs. Moreover, we examined its presence with respect to different forms of alternative splicing events. Our present results provide considerable insight on the association of methylation with splicing selection and promote our understanding of alternative splicing.

Materials and methods

The present study is based on computational analyses of DNA methylation data from HEP (Rakyan et al. 2004; Eckhardt et al. 2006). HEP is the first systematic and large-scale study of DNA methylation on the human genome that aims to map DNA methylation sites through the 30,000 gene loci. Specifically, we used the 26th release of the HEP pilot data (June 2006) containing the methylation status of 1.88 million CpG sites that were obtained from the analysis of 2,524 amplicons of human chromosomes 6, 20 and 22. In order to examine potential co-localization of the analyzed CpG sites and regulatory elements affecting gene expression, we mapped genomic regions of interest (hg16 genome assembly, July 2003), e.g., transcribed sites, repeats and alternative splicing events to the available methylation sites together with the assigned average methylation scores from 12 tissues ranging from 0 to 100 % for each record. CpG sites with known data that was not further analyzed by the HEP Project and data with no available methylation score was ignored.

We classified the resulted 33,352 CpGs with recorded methylation score, including zero methylation, in three categories depending on their average methylation values. CpG sites with methylation score greater or equal to 80 % (9,012 sites) belong to the dataset of the hypermethylated cytosines (HmC), while CpGs with methylation scores ranging from zero up to 20 % (15,296 sites) are characterized as hypomethylated or low-methylated (LmC). The remaining 9,044 methylation data with methylation levels between 20 and 80 % correspond to a third category containing data with intermediate methylation scores (ImC). This classification has been also applied in other studies and can safely distinguish sequences containing HmC and LmC (Rakyan et al. 2004; Eckhardt et al. 2006).

In the analysis of the 60-nt sequences centered at the splice junctions that contain HmC, ImC or LmC we extracted all hg16 introns, rather than exons, and retrieved their flanking sequences in order to avoid the selection of splice sites preceding and following first and last exons,

respectively. Introns and exons of less than 30 nt were excluded in this study in order to avoid overlaps between consecutive splice sequences. Similarly, in the analysis of the 130 nt intronic sequences flanking donor and acceptor sites, overlapping sequences were excluded. The UCSC Table Browser (Karolchik et al. 2004) through the Galaxy platform (Goecks et al. 2010) was used to extract the coordinates and the sequences of the examined regions and the Galaxy's lift-over tool was used to map genomic coordinates between different assemblies, where needed.

The set of 27,866 constitutive exons is a compilation of all hg16 reference exons subtracted by the 8,735 alternative splicing events that were extracted from the UCSC altEvents track of the same human genome assembly. In addition, the Repeatmasker track of the UCSC Table Browser (Smit AFA 1996–2007) was used to screen annotated SINEs in chromosomes 6, 20 and 22 (hg16 human genome assembly). Finally, for the combined analysis of the methylation and genomic data we developed Perl scripts and for the statistical evaluation of the identified associations we employed Chi-square and *t* tests.

Results

Methylation in transcribed sequences of chromosomes 6, 20 and 22

The proposed computational approach was first applied on the adjusted HEP data in order to measure the distribution

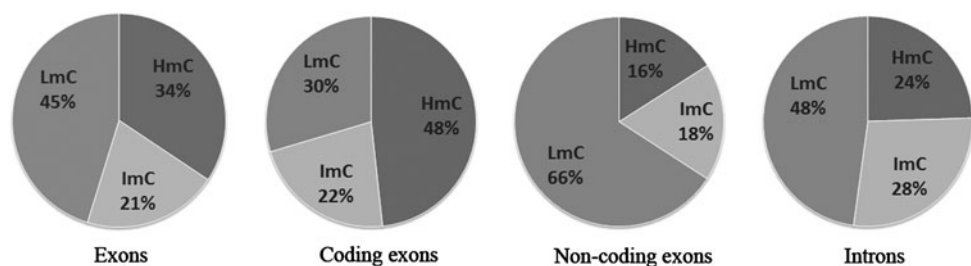
Table 1 Number of methylated CpGs in selected sequences depending on their coding potential

	Total CpGs	CpGs in exons	CpGs in coding exons	CpGs in non-coding exons ^a	CpGs in introns
HmC	9,012	3,644	2,915	729	2,844
ImC	9,044	2,167	1,336	831	3,219
LmC	15,296	4,795	1,788	3,007	5,553
Total	33,352	10,606	6,039	4,567	11,616

Multiple CpGs may be located in a single sequence

^a The number of non-coding exons resulted by subtracting the coding exons from exons

Fig. 1 Methylation frequencies (obtained from data in Table 1) in exons, coding exons, non-coding exons, and introns

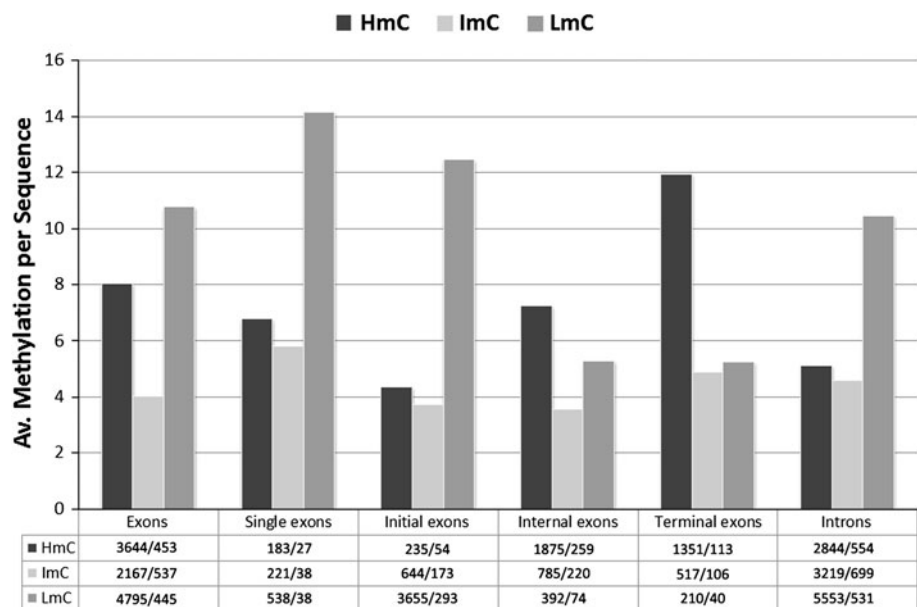


of methylation in exons compared to introns (Table 1). Exonic CpGs were divided in two categories: coding and non-coding. The latter are mostly positioned at both ends of the transcribed sequences and play a regulatory role in expression (Lilischkis et al. 2001; Medvedeva et al. 2010). This analysis reveals, as expected (Hodges et al. 2009; Chodavarapu et al. 2010), significant hypermethylation (HmC, 48 %) among coding exons, while CpG-containing sequences in introns are less frequently hypermethylated (HmC, 24 %; Fig. 1). It should be noted, however, that since all presently examined intronic and exonic sequences are CpG-containing, while introns also contain extended CpG-free sequences, the actual intronic methylation frequency is probably lower than that presently shown in Table 1. The results in Table 1 also show that a substantial number of the examined methylated CpGs are included in non-coding exons, and the size of this sample is suitable for valid statistical comparisons. In this case, CpGs in non-coding exons are scarcely hypermethylated (HmC, 16 %; Fig. 1).

We then analyzed the frequencies of hypermethylated, occasionally methylated and hypomethylated CpGs (HmC, ImC and LmC, respectively), in different exonic and intronic sequences. For this purpose, the CpGs of each of the above categories were divided by the corresponding number of non-duplicated sequences containing at least one CpG in the corresponding category (Fig. 2). In addition, exonic sequences were classified as initial, terminal, single and internal exons. Initial and terminal exons (mostly non-coding) are more CpG-dense compared to internal (coding) exons (total number of CpGs per number of sequences: 20.55 for initial, 22.08 for terminal exons, 16.10 for internal, and 26.75 for single exons), while initial exons are also more hypomethylated compared to intronic sequences.

The hypermethylation frequency in initial exons is comparable to that in introns (Fig. 2). Single exons have common epigenetic characteristics both with initial and partially with internal exons. Specifically, the average number of hypermethylated CpGs per sequence is similar in single and internal exons, but single, similarly to initial exons, also have higher LmC frequencies and are more CpG-dense than other exons. In addition, they share similar average numbers of CpG sites per sequence of variable methylation with terminal exons.

Fig. 2 Average methylation per sequence, i.e., the number of identified HmC, ImC and LmC versus the number of analyzed sequences (no. of CpGs/no. of seqs), in total, single, initial, internal and terminal exons and in introns, containing at least one recorded CpG. The number of CpGs per the number of sequences is shown in the embedded table



Overall, the data in Table 1 and Fig. 2 show that, in addition to dense exonic versus poor intronic methylation, limited methylation close to the transcription start site is probably required to establish the transcription process, while high CpG density is necessary to block transcription initiation or to terminate it (in single exons and in terminal exons, respectively), when required. Furthermore, this data might also indicate that high, possibly specific methylation, is required to regulate transcription of single, initial, terminal exons. In internal exons, where the average hypermethylation per sequence is high, though considerably lower than in terminal exons the CpG density is lower possibly introducing a transient halt during transcription. In this case the presence of methylation might be also associated with the spliceosome formation (Oberdoerffer 2012; Anastasiadou et al. 2011).

Distribution of methylation in donors and acceptors

The distribution of methylation and the CpG frequencies in splice sites were evaluated with respect to the position of each analyzed CpG relative to the intron/exon boundaries. These sites have been previously shown to be associated with distinct methylation transitions in single cell experiments (Hodges et al. 2009). Figure 3a illustrates the distribution for each methylation level in donors and acceptors in 20 nt sliding windows. In total, 1,973 CpGs (763 HmC, 491 ImC and 719 LmC) were identified in 60 nt regions centered at splice junctions. At the splice donor sequence corresponding to the U1snRNP binding site (Roca et al. 2005) the LmC frequency exhibits a local minimum compared to the HmC ($p = 0$). In addition, the frequency of methylated cytosines (HmC and ImC) in the intronic region

close to acceptors is also very low compared to the intronic sites close to donors, indicating that, CpGs in donors are frequently hypomethylated contrary to those in the corresponding acceptor region ($p = 0$).

The hypermethylation frequency, i.e., the number of HmC divided by the total CpGs, either HmC, ImC or LmC identified in each region is shown in Fig. 3b. Close to the splice junction, the hypermethylation frequency is significantly higher at the acceptor site (both intronic and exonic) compared to the average frequency in introns or exons (Fig. 1). In donors, the hypermethylation frequency is comparable to the average frequency in introns and exons. This data reveal that, although CpGs might be frequent in splice donors, only a small fraction is hypermethylated, while the limited number of CpGs present in acceptors is mostly hypermethylated. Thus, in 5' intron ends there is probably a strict selection of CpGs undergoing methylation, while hypermethylation is very frequent in 3' intron ends compared to all other sites examined.

We then analyzed the hypermethylation frequencies among CpGs in the intronic sequences, flanking splice junctions using 20-nt sliding windows (Fig. 4). Up to 90 nt from the acceptor junction, the hypermethylation frequency decreases but remains significantly higher compared to the average intronic, or even to the corresponding donor sites. Approximately 100 nt from the acceptor site the intronic hypermethylation frequency decreases. Extending by 30 nt this region it is evident that the HmC frequency becomes less variable and it remains relatively low. At the donor site the hypermethylation frequency extended in the 130-nt intronic region is less variable than the corresponding frequency at the acceptor site and also continuously lower than the average frequency of introns.

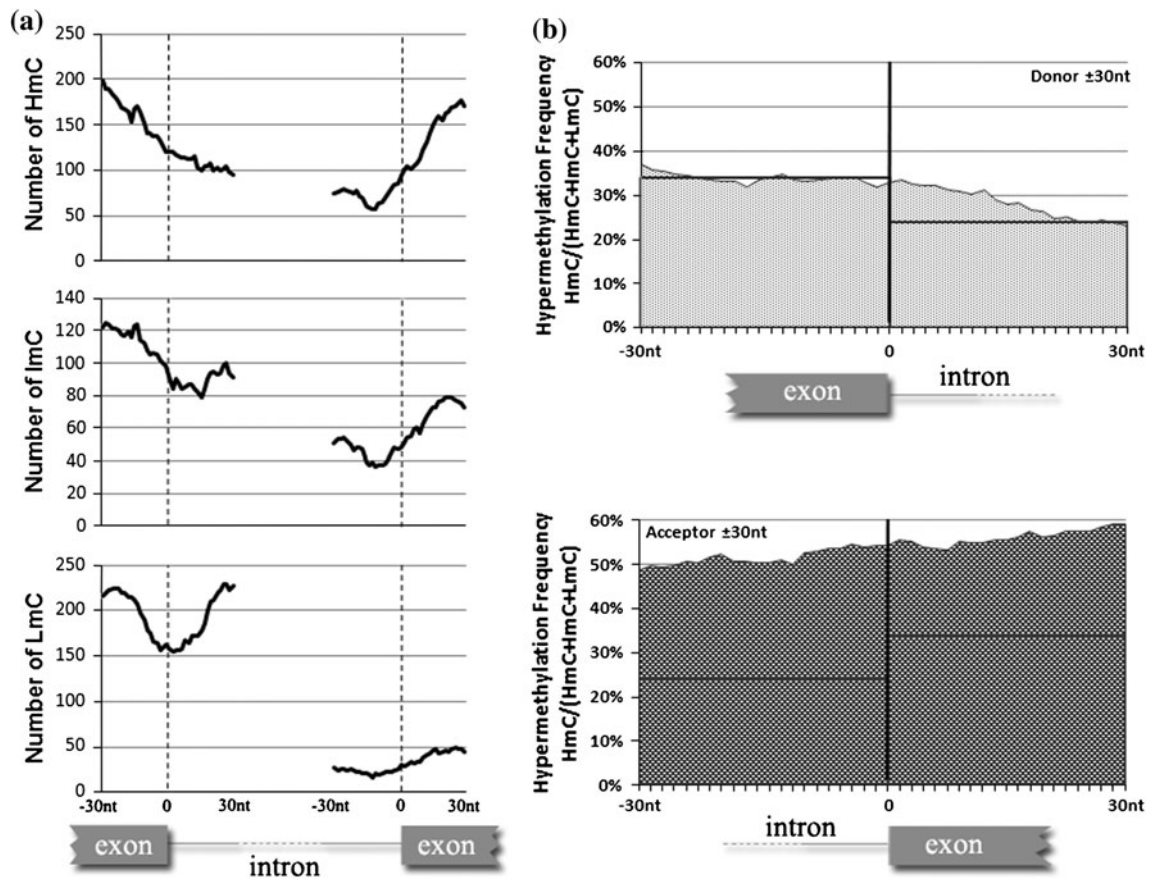


Fig. 3 a Distribution of methylation at both sides of donors and acceptors in 20-nt sliding windows of 60 nt total length centered at the splice junctions. The exon–intron boundaries are shown in vertical

dashed lines. b Hypermethylation frequency corresponding to the same region shown in a. The average HmC frequency in whole exons and introns is shown in *horizontal lines* (obtained from Fig. 1)

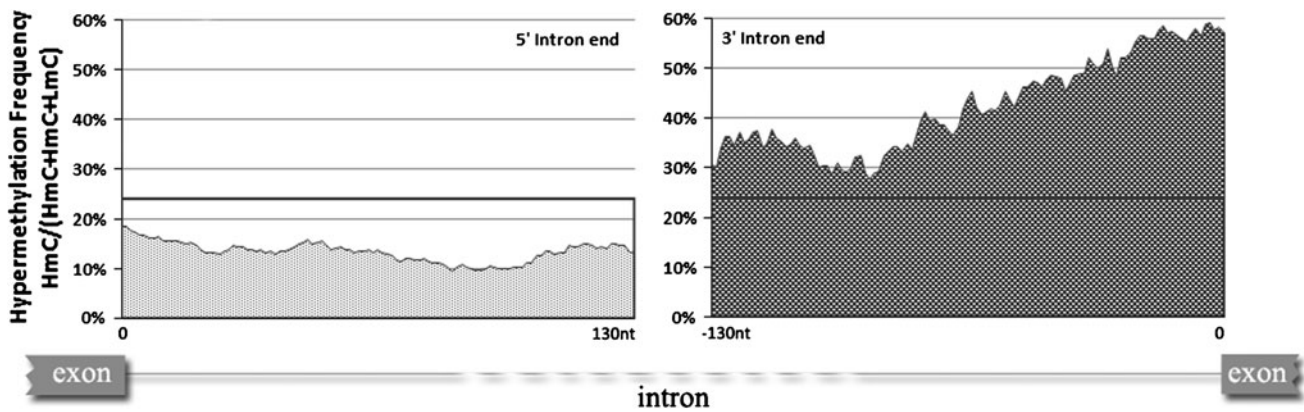


Fig. 4 Hypermethylation frequency in the 130-nt intronic sequences flanking donor (*left*) and acceptor (*right*) sites, using 20-nt sliding windows. The average HmC frequency in introns, obtained from Fig. 1, is shown in *horizontal lines*

In addition, the methylation density in the 100-nt intronic region close to splice junctions was found to be comparable to that in exons and approximately 100-fold higher than the corresponding density in total intronic sequences shown in Table 2. The density of the LmC and

partly methylated CpGs (ImC) close to the splice junctions is very high compared to introns. On the contrary, the hypermethylated CpG density is very low in donors (Table 2). Thus, in 5' intron ends close to splice junctions a strict selection of CpGs undergoing methylation, extends

Table 2 Methylation density in different sequences

	5' intron ends (100 nt)	3' intron ends (100 nt)	Introns	Cassette exons	Constitutive exons
HmC	0.024	0.022	0.0003	0.0259	0.0095
ImC	0.021	0.018	0.0002	0.0134	0.0054
LmC	0.059	0.045	0.0005	0.0153	0.0167

Total CpGs/(number of non-duplicated sequences containing at least one CpG \times average length of the sequences)

for at least 100 nt within introns, while the probability of hypermethylation within acceptor CpGs is higher than in any other site presented in Table 1.

Provided that previous studies based on a single cell analysis (Hodges et al. 2009) have shown sharp methylation changes in intron/exon boundaries, it is conceivable that the presently shown high methylation scores and slow demethylation transitions in intron/exon boundaries, are related to the fact that this analysis is based on average methylation frequencies for all tissues and cells. Together, these results show that methylation in the 100-nt edges of intronic sequences is probably very significant and selective for determining the “epitype” (Hodges et al. 2009), i.e., epigenetic regulation and splice site selection in individual tissues.

Methylation and alternative splicing

Methylation frequencies in cassette and constitutive exons

We then tested whether the methylation frequency might be altered in association with alternatively spliced sequences such as cassette exons, exons with weak, non-canonical splice sites, or in splice sites processed by the minor spliceosome. For this purpose, sequences containing CpGs with recorded methylation data were further analyzed for their presence in alternatively spliced forms. The most frequently observed alternative splicing events involve cassette exons which exhibit well-defined intron-exon boundaries and are characterized by lower splicing potential (scores) compared to constitutive exons (Clark and Thanaraj 2002; Garg and Green 2007).

As shown in Table 2, the density of hypomethylated CpGs is higher in the 990 analyzed constitutive exons (1.67 % LmC vs. 0.95 and 0.54 % for HmC and ImC, respectively), while the hypermethylation density is higher in cassette (64 in total) than constitutive exons (2.59 vs. 0.95 %, respectively). Thus, cassette exons are characterized by a considerable increase of the methylation density, while their total CpG density does not significantly vary from that of constitutive exons. Although the sample of cassette LmC is limited, the methylation differences of

Table 3 Statistical analysis of hypermethylation differences (p values) between total exons, introns relative to alternative splicing events

	Total exons		Total introns	
Cassette exons	$p < 0.0001$	Retained introns	$p = 0.706$	
Bleeding exons	$p = 0.313$	ATAC introns	$p < 0.0001$	
		Strange introns	$p < 0.0001$	

HmC and LmC between total exons and cassette sequences are very significant ($p < 0.0001$; Table 3).

The above findings (summarized in Fig. 5) reveal that significant methylation differences are observed between alternative and constitutive exons depending on the type of alternative splicing process. Particularly prominent is also the “infiltration of methylation” in intronic sequences neighboring splice junctions and the methylation differences between donors and acceptors. Information regarding alternative splicing in association with the distribution of methylation in exonic splicing enhancers (Anastasiadou et al. 2011) will facilitate the design of improved computational tools for alternative splice site prediction (Richard et al. 2010; Wang and Marin 2006) and evaluation of disease-related mutations and polymorphisms (Kouidou et al. 2009).

Splicing potential in cassette and constitutive exons

The flanking 5' and 3' intron ends of the analyzed cassette and constitutive exons were then evaluated for their compliance with the canonical forms, i.e., GT dinucleotides for the following donors and AG for the preceding acceptors as regards to their methylation and their splicing strength. The flanking intron ends were identified in cassette and constitutive exons using UCSC altEvents track in our sample and the splicing potential of canonical splice sites was calculated using the identification method proposed by Shapiro and Senapathy (1987).

Using this approach we observed that 63 out of 64 cassette exons containing at least one HmC, ImC or LmC were canonical (overlaps included), showing that the presently examined cassette exons comply with the canonical splicing forms ($p < 0.0001$). The only cassette exon with non-canonical splice sites flanks a GC donor and AG acceptor. Of the 990 identified constitutive exons that contain at least one HmC, ImC or LmC (Table 4), a substantial number corresponds to first and last exons. Almost all internal constitutive exons exhibit canonical AG/GT intron ends (99.34 %). As expected from the results shown in Fig. 2, we also observed that hypermethylation is frequent in terminal and less prominent in initial exons, while the opposite is found for rarely methylated CpGs (LmC). However, intermediate methylation was equally frequent in

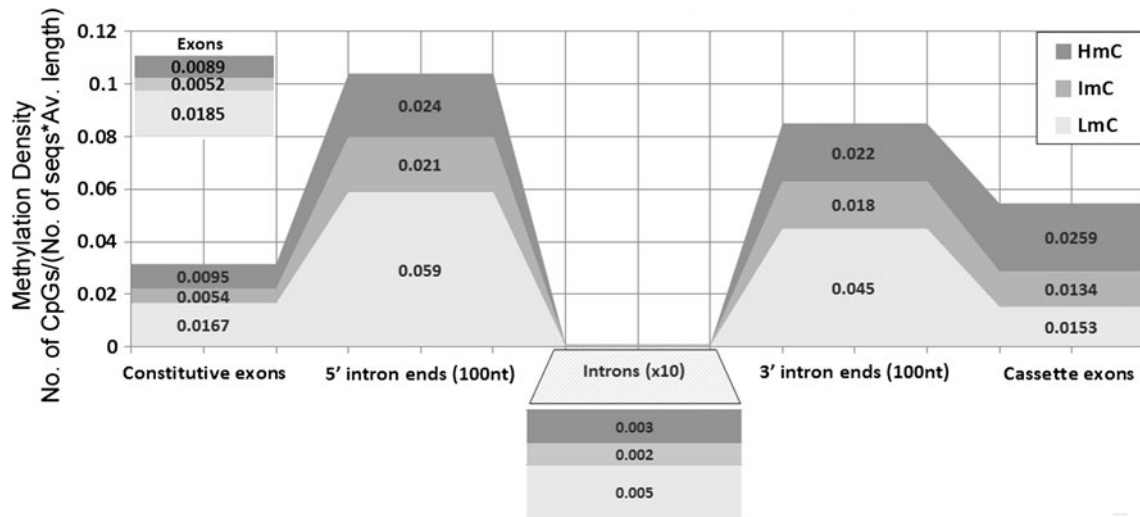


Fig. 5 Graphical representation of methylation densities in exons (cassette, constitutive and total), introns and 100 nt 5', 3' intronic sites. The relative widths of the *different grey tones* correspond to methylation frequencies. The shown frequency scale is enlarged 10× for introns

Table 4 Analysis of splice sites flanking constitutive exons (including initial and terminal exons) which contain at least one HmC, ImC or LmC

	CpG-containing exons	Acceptor		Donor		% Paired canonical (AG/GT)
HmC	348	AG	295	GT	232	98.6 % (205/208)
		(Initial exons)	51	(Terminal exons)	113	
		AT	1	AT	1	
		AC	1	GC	2	
ImC	382	AG	256 (<i>p</i> = 0)*	GT	264 (<i>p</i> = 0.48)*	100 % (176/176)
		(Initial exons)	126	(Terminal exons)	115	
		Other	0	AT	1	
				GC	2	
LmC	260	AG	95 (<i>p</i> = 0)*	GT	201 (<i>p</i> = 0.004)*	100 % (69/69)
		(Initial exons)	165	(Terminal exons)	57	
		Other	0	GC	2	

“Other” refers to non-canonical donors and acceptors of internal exons unless otherwise specified. Paired canonical: Exons flanked by canonical splice sites (both acceptor and donor)

* Relative to HmC

initial and terminal exons. Among constitutive acceptors, methylation differences (HmC/ImC and HmC/LmC) are statistically significant, and the same is observed in some cases in donors (HmC/LmC). However, in donors we also observed that there is no significant difference between the frequency of hypermethylated CpGs (HmC) and ImCs, i.e., CpG cytosines which are either methylated only in specific tissues, or at specific developmental stages. This finding probably indicates that in donors and acceptors the methylation process is regulated by different parameters. All

other differences are statistically significant. As expected from the results shown in Fig. 2, in constitutive terminal exons hypermethylation is more frequent relative to initial exons while the opposite is found for rarely methylated CpGs (LmC). Intermediate methylation was equally frequent in initial and terminal exons.

The splicing scores of the examined donors and acceptors also differ between cassette and constitutive exons. Acceptors with canonical form that precede cassette exons exhibit lower splicing scores compared to constitutive

Table 5 Splice site scores and average scores of the three categories for cassette and constitutive exons examined (all canonical, containing CpGs)

	Cassette	Constitutive
Donors		
HmC	78.53 (33)	82.71 (232)
ImC	80.04 (26)	83.44 (264)
LmC	77.84 (4)	83.86 (201)
Av. score	78.80	83.34
Acceptors		
HmC	48.95 (34)	54.03 (295)
ImC	49.99 (26)	54.32 (256)
LmC	46.16 (4)	54.31 (95)
Av. score	48.37	54.22

The corresponding number of exons is shown in parenthesis

exons (Table 5). The data on cassette exons are in agreement with previous observations by Choi (2010), which reveal that methylation is enriched in exons with weak splice sites and is probably associated with spliceosome positioning and with previous findings for cassette exons (Garg and Green 2007; Clark and Thanaraj 2002). These data are in support of a splicing-enhancing role for exons with weak splice sites.

Methylation in other alternatively expressed loci

In order to further verify the association of methylation with various, less frequent types of alternatively spliced sequences (other than cassette exons), we identified the alternative splicing events among the analyzed CpG-containing sequences studied by HEP (hg16, July 2003) and correlated them with the methylation status in these sequences. Specifically, each methylation locus was searched through the list of annotated alternative splicing events. Figure 6 shows the matched events per methylation category.

The results in Fig. 6 reveal that most significant differences in methylation are observed in specific types of alternative splicing events, e.g., in alternative promoters, in alternative 3' ends, and in strange introns and bleeding exons. Specifically, Fig. 6 shows that hypermethylation is more frequent among alternatively spliced sequences in which the intron-exon boundaries are well defined, relative to other splicing events defined as “bleeding” in which the intron-exon boundaries are not well defined. Thus, bleeding exons are less hypermethylated compared to cassette exons (32.04 % HmC, vs. 60.52 % in cassette exons) or compared to the average coding exons (48 %, Table 1). In addition, other types of sequences resulting from splicing events which are considered products of partial spliceosomal processing, e.g., retained introns, are characterized by low methylation levels (HmC, 29.03 %; LmC, 48.39 %).

In both bleeding exons and retained introns the statistical difference with total exons and introns is not significant ($p = 0.313$ and $p = 0.706$, respectively; Table 3). In addition, strange introns are significantly more hypermethylated than total introns (Table 1; Fig. 6) and the same is observed with ATAC introns, a specific intron category with non-canonical intron ends, which are processed by the minor spliceosome (HmC, 57.14 %, similar to the frequency of cassette exons), indicating that there might be a significant correlation of methylation frequencies and selection of different intron ends.

Once more, this type of analysis is in agreement with the results shown in Table 1 and Fig. 2 and reinforces the finding that methylation is significantly correlated with the selection of expressed sequences and acts, in a sequence-specific, orchestrated manner, depending on the splice site acceptor and donor sequences (Tables 4, 5; Figs. 5, 6). In contrast, “deviations” occurring by incomplete processing of one exon boundary, i.e., bleeding exons, are associated with lower methylation frequencies and other selection processes. All previously described differences are very significant (p values are shown in Table 3).

Methylation in SINEs

The frequency of methylation in SINEs, which are also known regulatory parameters for transcript inclusion, was then evaluated using RepeatMasker (Smit AFA 1996–2007) in order to identify methylation differences which might discriminate the various SINEs relative to their epigenetic modification state. These differences were then compared to the previously described impact on transcript inclusion for the different SINEs. We identified 308 SINEs containing HmC, ImC or LmC among the 167,304 detected SINEs in chromosomes 6, 20 and 22 (hg16, July 2003). The identified SINEs correspond to 0.184 % of the total SINEs in these chromosomes, regardless of their CpG content (Table 6). Analysis of the different SINE categories (Fig. 7) reveals a variety of different Alus present in this sample. Among these, AluSx, MIR and MIRb are most frequent.

Most of these SINEs are hypermethylated (HmC, 54.87 %; Table 6), but strong differences of the methylation frequencies among Alus and particularly MIRs (Fig. 8) are also observed. While most Alus, including AluSx, which is the most frequent Alu present in this sample and in chromosomes 6, 20 and 22 (Fig. 7), are rarely hypomethylated, MIRs (MIR, MIR3 and MIRb), as well as AluJo are characterized by lower hypermethylation frequencies (Fig. 8). The difference of hypermethylation between the above repeat sequences, i.e., AluJo and MIRs, relative to all the remaining SINEs is statistically significant ($p < 0.0001$). Methylation among MIRs also varies. In view of these

Fig. 6 Distribution of the studied **a** HmC, **b** ImC and **c** LmC with respect to their co-occurrence with alternative splicing events. The frequency of all reported alternatively spliced introns and exons in chromosomes 6, 20 and 22 is shown in **d**

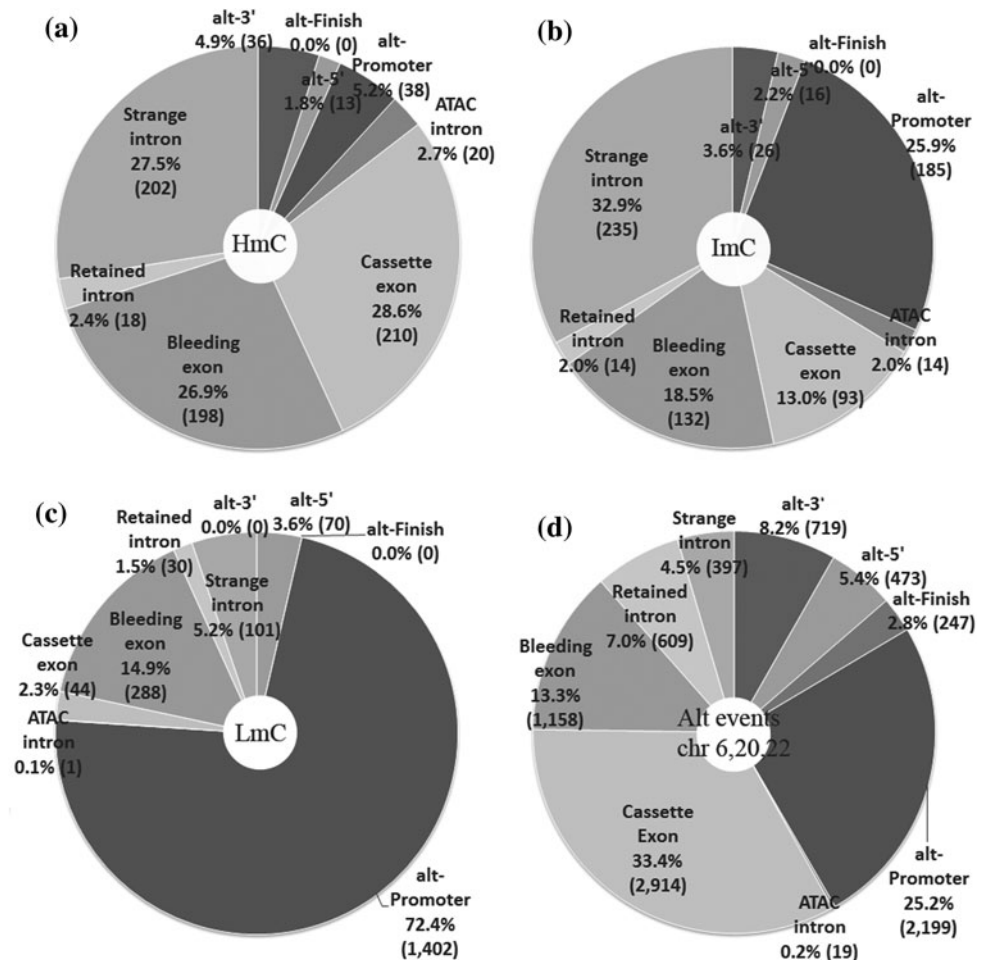


Table 6 Total CpGs observed in SINEs and the number of SINEs containing at least one HmC, ImC and LmC

	Total CpGs in SINEs	SINEs containing at least one CpG (CpGs per SINE)
HmC	169	55 (3.07)
ImC	89	51 (1.74)
LmC	50	23 (2.17)
Total	308	129 (2.38)

The average number of CpGs per SINE sequence is shown in parenthesis

significant variations of the methylation frequencies in AluJo and MIRs which have been previously associated with alternative splicing (Lin et al. 2008; Shen et al. 2011) and transcript inclusion in coding sequences, it is probable that methylation might also contribute to the regulation of “exonization” in SINE-containing methylated sequences.

Discussion

Contrary to promoter methylation, the role of exonic and intronic methylation is still ill-defined. Recent evidence

points out the association of promoter methylation with chromatin modification, transcriptional elongation and RNA polymerase stalling (Chen et al. 2011). This data also indicates a similar role for intragenic methylation in splicing and its regulation. Our data provides a detailed description of the distribution of methylation and of epigenetically unmodified CpGs in different human sequences with respect to splicing, “weak” and “strong”, alternative and constitutive splice sites, and “exonization” (single, initial and terminal exons, as well as Alus which are known to contribute to “exonization”). In addition, we provide a detailed account of sites, which are partially or rarely epigenetically modified. Based on our results, it becomes evident that the cytosine methylation density is modified according to the transcriptional and splicing potential of a sequence, and could thus act as a supplementary parameter for regulating the splicing strength of splice sites (in defining “spliceable” sequences). Moreover, it becomes evident that different “rules” regulate the methylation frequency in donors and acceptors, as well as at the 5', relative to the 3' ends of genes. More specifically, our results reveal that intragenic methylation is strongly associated with transcribed regions in which specific transcript

Fig. 7 Relative frequencies of different types of SINEs in chromosomes 6, 20 and 22. Some Alus (AluSq/x, AluSg1, Alu, FAM, AluYd2, AluYc2, AluJ/FLAM, and AluYd3a1) generally denoted as ‘other’ have frequency greater than 0.3 % (not shown)

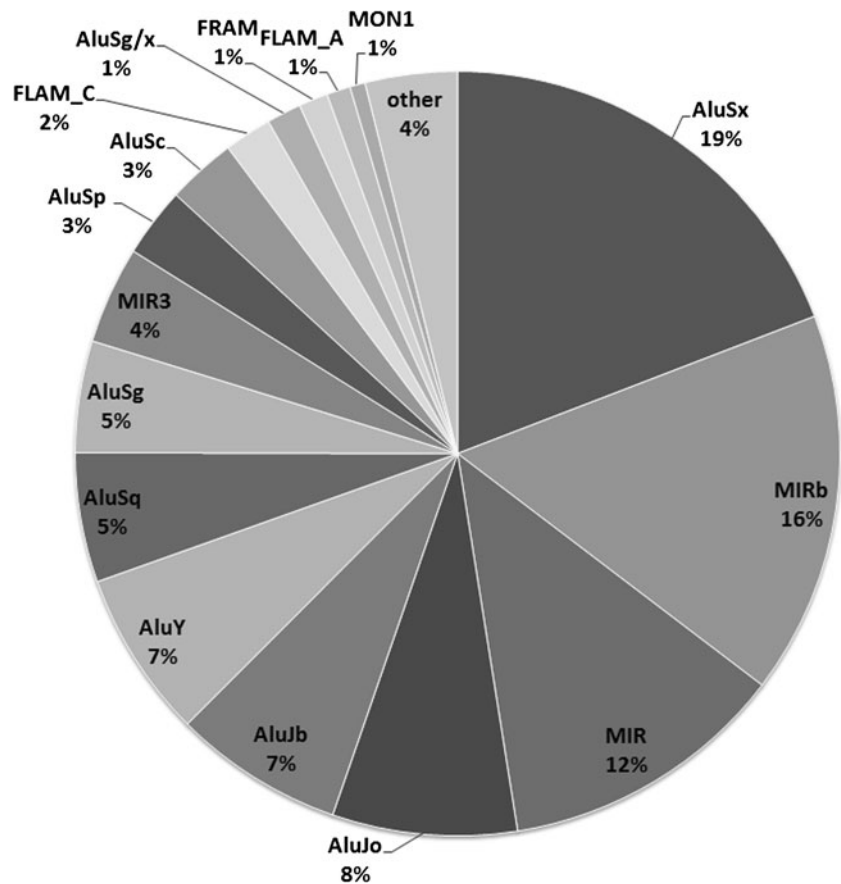
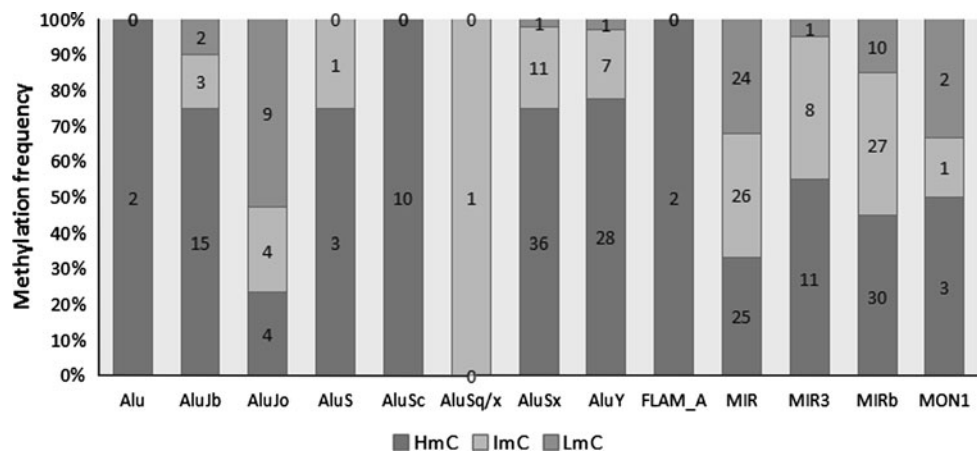


Fig. 8 Distribution of methylated CpGs among different types of SINEs. The total number of SINEs in chromosomes 6, 20 and 22 is 167,304



processing is required, such as alternative exons, weak acceptors, as well as rare ATAC introns. Epigenetic marks, such a reduction of methylation appear to define exon/intron boundaries. It should be noted that in a previous publication we showed an increase of methylation in nucleosome-binding consensus sequences, which has also recently been shown to regulate the splicing efficiency (Anastasiadou et al. 2011).

During the preparation of this manuscript, two studies on genome-wide methylation in rat (Sati et al. 2012), as

well as in specific cell types have been reported (Zhou et al. 2012). Although different analytical procedures have been applied in these studies and the results are not directly comparable, they both emphasize the role of DNA methylation in the alternative transcript inclusion. Our study presents a detailed account of the average methylation in more than one human tissue as well as the average degree of methylation (hyper-, inter-, hypomethylation) for each CpG site in different developmental stages. In conclusion, the above data are critical for understanding the role of the

vast majority of DNA methylation, which is focused in intragenic regions. It is, however, conceivable that methylation involving, but yet undefined regulatory regions related to miRNA coding and other regulatory sites of the genome is expected to play significant role in expression and differentiation.

Moreover, we analyzed data with respect to the CpG frequency, i.e., the number of methylation targets for each sequence. Previous studies (Davey and Allan 2003; Dhasarathy and Wade 2008) have shown that the presence of CpG sequences alters the DNA conformation and nucleosome binding independently of the methylation density. Our detailed analysis indicates that the CpG and methylation densities might act as independent parameters defining the selection and efficiency of splice sites and promoters, and the transcript inclusion of exons or intronic sequences. In UTRs and splice donors, where transcriptional stalling is required, the appropriate protein assembly is probably dependent on the very frequent presence of methylation and the recognition by different methyl-binding proteins (Dhasarathy and Wade 2008). Similarly, methylation might facilitate splicing of weak splice sites and alternatively spliced exons by assisting spliceosome recruitment. In promoters and first exons, high CpG density probably denotes the presence of a tentative regulatory target, while their accessibility is regulated by methylation. In introns where splice sites and intronic acceptors which are CpG-poor, the chromatin architecture, looping and absence of nucleosomes, probably facilitates spliceosome binding and splicing. Finally, the expression of intronic CpG-dense sites, such as Alus, might be dependent on their methylation density. Thus, analysis of the CpG density relative to the methylation distribution presented in this study contributes to a better understanding of a new role of DNA methylation in transcription and introduces new constituents to the process of alternative splicing.

Comprehensive analysis of all available data and large-scale studies on epigenetic modifications, are now required to promote our understanding of the epigenetic regulation of splicing and transcript inclusion. However, since DNA imprinting includes several forms of epigenetic modification, such as non-CpG methylation, also frequent in splice sites (Shen et al. 2011), and other forms of cytosine modification, e.g., hydroxymethylcytosine, also presently considered critical for DNA expression (Pastor et al. 2011), it is evident that the present data provides only “quick glances” at the highly complex epigenetic code and its impact on expression.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Alexander RD, Innocente SA, Barrass JD, Beggs JD (2010) Splicing-dependent RNA polymerase pausing in yeast. *Mol Cell* 40(4):582–593. doi:10.1016/j.molcel.2010.11.005
- Anastasiadou C, Malousi A, Maglaveras N, Kouidou S (2011) Human epigenome data reveal increased CpG methylation in alternatively spliced sites and putative exonic splicing enhancers. *DNA Cell Biol* 30(5):267–275. doi:10.1089/dna.2010.1094
- Chen Y, Jorgensen M, Kolde R, Zhao X, Parker B, Valen E, Wen J, Sandelin A (2011) Prediction of RNA polymerase II recruitment, elongation and stalling from histone modification data. *BMC Genomics* 12:544. doi:10.1186/1471-2164-12-544
- Chodavarapu RK, Feng S, Bernatavichute YV, Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, Casero D, Bernal M, Huijser P, Clark AT, Kramer U, Merchant SS, Zhang X, Jacobsen SE, Pellegrini M (2010) Relationship between nucleosome positioning and DNA methylation. *Nature* 466(7304):388–392. doi:10.1038/nature09147
- Choi JK (2010) Contrasting chromatin organization of CpG islands and exons in the human genome. *Genome Biol* 11(7):R70. doi:10.1186/gb-2010-11-7-r70
- Clark F, Thanaraj TA (2002) Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. *Hum Mol Genet* 11(4):451–464. doi:10.1093/hmg/11.4.451
- Davey C, Allan J (2003) Nucleosome positioning signals and potential H-DNA within the DNA sequence of the imprinting control region of the mouse *Igf2r* gene. *Biochim Biophys Acta* 1630(2–3):103–116. doi:10.1016/j.bbexp.2003.09.009
- Dhasarathy A, Wade PA (2008) The MBD protein family—reading an epigenetic mark? *Mutat Res* 647(1–2):39–43. doi:10.1016/j.mrfmmm.2008.07.007
- Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, Haefliger C, Horton R, Howe K, Jackson DK, Kunde J, Koenig C, Liddle J, Niblett D, Otto T, Pettett R, Seemann S, Thompson C, West T, Rogers J, Olek A, Berlin K, Beck S (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 38(12):1378–1385. doi:10.1038/ng1909
- Garg K, Green P (2007) Differing patterns of selection in alternative and constitutive splice sites. *Genome Res* 17(7):1015–1022. doi:10.1101/gr.6347907
- Goecks J, Nekrutenko A, Taylor J (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11(8):R86. doi:10.1186/gb-2010-11-8-r86
- Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M, Zhang MQ, Ye K, Bhattacharjee A, Brizuela L, McCombie WR, Wigler M, Hannon GJ, Hicks JB (2009) High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Res* 19(9):1593–1605. doi:10.1101/gr.095190.109
- Huh JW, Kim DS, Ha HS, Lee JR, Kim YJ, Ahn K, Lee SR, Chang KT, Kim HS (2008) Cooperative exonization of MaLR and AluJo elements contributed an alternative promoter and novel splice variants of RNF19. *Gene* 424(1–2):63–70. doi:10.1016/j.gene.2008.07.030
- Karambataki M, Malousi A, Maglaveras N, Kouidou S (2010) Synonymous polymorphisms at splicing regulatory sites are associated with CpGs in neurodegenerative disease-related genes. *Neuromolecular Med* 12(3):260–269. doi:10.1007/s12017-009-8111-0
- Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ (2004) The UCSC Table Browser data

- retrieval tool. *Nucleic Acids Res* 32 (Database issue):D493–D496. doi:[10.1093/nar/gkh103](https://doi.org/10.1093/nar/gkh103)
- Kouidou S, Malousi A, Maglaveras N (2009) Li-Fraumeni and Li-Fraumeni-like syndrome mutations in p53 are associated with exonic methylation and splicing regulatory elements. *Mol Carcinog* 48(10):895–902. doi:[10.1002/mc.20537](https://doi.org/10.1002/mc.20537)
- Lilischkis R, Kneitz H, Kreipe H (2001) Methylation analysis of CpG islands. *Methods Mol Med* 57:271–283. doi:[10.1385/1-59259-136-1:271](https://doi.org/10.1385/1-59259-136-1:271)
- Lin L, Shen S, Tye A, Cai JJ, Jiang P, Davidson BL, Xing Y (2008) Diverse splicing patterns of exonized Alu elements in human tissues. *PLoS Genet* 4(10):e1000225. doi:[10.1371/journal.pgen.1000225](https://doi.org/10.1371/journal.pgen.1000225)
- Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T (2011) Epigenetics in alternative pre-mRNA splicing. *Cell* 144(1):16–26. doi:[10.1016/j.cell.2010.11.056](https://doi.org/10.1016/j.cell.2010.11.056)
- Malousi A, Maglaveras N, Kouidou S (2008) Intronic CpG content and alternative splicing in human genes containing a single cassette exon. *Epigenetics* 3(2):69–73. doi:[10.4161/epi.3.2.6066](https://doi.org/10.4161/epi.3.2.6066)
- Medvedeva YA, Fridman MV, Oparina NJ, Malko DB, Ermakova EO, Kulakovskiy IV, Heinzl A, Makeev VJ (2010) Intergenic, gene terminal, and intragenic CpG islands in the human genome. *BMC Genomics* 11:48. doi:[10.1186/1471-2164-11-48](https://doi.org/10.1186/1471-2164-11-48)
- Oberdoerffer S (2012) A conserved role for intragenic DNA methylation in alternative pre-mRNA splicing. *Transcription* 3(3)
- Pandya-Jones A, Black DL (2009) Co-transcriptional splicing of constitutive and alternative exons. *RNA* 15(10):1896–1908. doi:[10.1261/rna.1714509](https://doi.org/10.1261/rna.1714509)
- Pastor T, Talotti G, Lewandowska MA, Pagani F (2009) An Alu-derived intronic splicing enhancer facilitates intronic processing and modulates aberrant splicing in ATM. *Nucleic Acids Res* 37(21):7258–7267. doi:[10.1093/nar/gkp778](https://doi.org/10.1093/nar/gkp778)
- Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, Tahiliani M, Daley GQ, Liu XS, Ecker JR, Milos PM, Agarwal S, Rao A (2011) Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 473(7347):394–397. doi:[10.1038/nature10102](https://doi.org/10.1038/nature10102)
- Rakyan VK, Hildmann T, Novik KL, Lewin J, Tost J, Cox AV, Andrews TD, Howe KL, Otto T, Olek A, Fischer J, Gut IG, Berlin K, Beck S (2004) DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol* 2(12):e405. doi:[10.1371/journal.pbio.0020405](https://doi.org/10.1371/journal.pbio.0020405)
- Richard H, Schulz MH, Sultan M, Nurnberger A, Schrinner S, Balzereit D, Dagand E, Rasche A, Lehrach H, Vingron M, Haas SA, Yaspo ML (2010) Prediction of alternative isoforms from exon expression levels in RNA-Seq experiments. *Nucleic Acids Res* 38(10):e112. doi:[10.1093/nar/gkq041](https://doi.org/10.1093/nar/gkq041)
- Rimessi P, Fabris M, Bovolenta M, Bassi E, Falzarano S, Gualandi F, Rapezzi C, Coccolo F, Perrone D, Medici A, Ferlini A (2010) Antisense modulation of both exonic and intronic splicing motifs induces skipping of a DMD pseudo-exon responsible for x-linked dilated cardiomyopathy. *Hum Gene Ther* 21(9):1137–1146. doi:[10.1089/hum.2010.010](https://doi.org/10.1089/hum.2010.010)
- Roca X, Sachidanandam R, Krainer AR (2005) Determinants of the inherent strength of human 5' splice sites. *RNA* 11(5):683–698. doi:[10.1261/rna.2040605](https://doi.org/10.1261/rna.2040605)
- Sati S, Tanwar VS, Kumar KA, Patowary A, Jain V, Ghosh S, Ahmad S, Singh M, Reddy SU, Chandak GR, Raghunath M, Sivasubbu S, Chakraborty K, Scaria V, Sengupta S (2012) High resolution methylome map of rat indicates role of intragenic DNA methylation in identification of coding region. *PLoS One* 7(2):e31621. doi:[10.1371/journal.pone.0031621](https://doi.org/10.1371/journal.pone.0031621)
- Schwartz S, Ast G (2010) Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. *EMBO J* 29(10):1629–1636. doi:[10.1038/emboj.2010.71](https://doi.org/10.1038/emboj.2010.71)
- Shapiro MB, Senapathy P (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 15(17):7155–7174. doi:[10.1093/nar/15.17.7155](https://doi.org/10.1093/nar/15.17.7155)
- Shen S, Lin L, Cai JJ, Jiang P, Kenkel EJ, Stroik MR, Sato S, Davidson BL, Xing Y (2011) Widespread establishment and regulatory impact of Alu exons in human genes. *Proc Natl Acad Sci USA* 108(7):2837–2842. doi:[10.1073/pnas.1012834108](https://doi.org/10.1073/pnas.1012834108)
- Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479(7371):74–79. doi:[10.1038/nature10442](https://doi.org/10.1038/nature10442)
- Smit AFA, Green P (1996–2007) RepeatMasker Open-3.0
- Sorek R, Ast G, Graur D (2002) Alu-containing exons are alternatively spliced. *Genome Res* 12(7):1060–1067. doi:[10.1101/gr.229302](https://doi.org/10.1101/gr.229302)
- Sorek R, Lev-Maor G, Reznik M, Dagan T, Belinky F, Graur D, Ast G (2004) Minimal conditions for exonization of intronic sequences: 5' splice site formation in Alu exons. *Mol Cell* 14(2):221–231. doi:[10.1016/S1097-2765\(04\)00181-9](https://doi.org/10.1016/S1097-2765(04)00181-9)
- Tao Y, Xi S, Briones V, Muegge K (2010) Lsh mediated RNA polymerase II stalling at HoxC6 and HoxC8 involves DNA methylation. *PLoS One* 5(2):e9163. doi:[10.1371/journal.pone.0009163](https://doi.org/10.1371/journal.pone.0009163)
- Tilgner H, Guigo R (2010) From chromatin to splicing: RNA-processing as a total artwork. *Epigenetics* 5(3). doi:[10.4161/epi.5.3.11319](https://doi.org/10.4161/epi.5.3.11319)
- Vorechovsky I (2010) Transposable elements in disease-associated cryptic exons. *Hum Genet* 127(2):135–154. doi:[10.1007/s00439-009-0752-4](https://doi.org/10.1007/s00439-009-0752-4)
- Wang M, Marin A (2006) Characterization and prediction of alternative splice sites. *Gene* 366(2):219–227. doi:[10.1016/j.gene.2005.07.015](https://doi.org/10.1016/j.gene.2005.07.015)
- Zhou Y, Lu Y, Tian W (2012) Epigenetic features are significantly associated with alternative splicing. *BMC Genomics* 13(1):123. doi:[10.1186/1471-2164-13-123](https://doi.org/10.1186/1471-2164-13-123)