Links between chromatin structure and regulation of alternative pre-mRNA splicing in mammalian cells

Camilla Iannone

DOCTORAL THESIS UPF/ 2014

THESIS DIRECTOR

Dr Juan Valcárcel

Department of Gene Regulation, Stem Cells and Cancer, CRG

«"O frati," dissi, "che per cento milia
perigli siete giunti a l'occidente,
a questa tanto picciola vigilia

d'i nostri sensi ch'è del rimanente
non vogliate negar l'esperïenza,
di retro al sol, del mondo sanza gente.

Considerate la vostra semenza:
fatti non foste a viver come bruti,
ma per seguir virtute e canoscenza".»
Dante, L'Inferno
Canto XXVI, vv. 112-120

Acknowledgements

I need to thank many people that have been with me, encouraged me, helped during this incredible journey that has been my PhD. And without whom I wouldn't have made it here, writing these words in this exciting, a little hectic moment of my journey.

First of all I need to give a big thank you to my thesis supervisor Juan, for giving me the opportunity to work on this exciting, risky project (which at times drove me nuts). I am grateful for all the time we discussed science in the past five years, and thank you for teaching me so much not only about science, but also about being a scientist.

I want to thank also Miguel and Roderic, who made a pair of enthusiastic co-supervisors, who have always been interested about my projects and pushed and challenged me to do better.

This thesis has been a love and hate relationship with bioinformatics and bioinformaticians. I need to thank all of you, Hagen, Andy, João and Pan, for your enthusiasm, for being part of these projects, teaching me a little bit of R here and there and being patient when I wasn't getting a plot, and for making me ralize what are the right (p) Values in life.

I am also thankful to all the members of Juan and Miguel's labs for sharing protocols, data, and for countless discussions.

My lab mates made going to work a pleasant experience. Thank you for all the time we spent together, from scientific discussions to beach volley session, from the nights out at the bar to weekends out of town, from conferences to music sessions in the lab.
I was lucky to have MP sitting in the lab next to me for the first years, who has been a good friend, a great scientific inspiration, and great fun.
Thanks to Elias. Or better, thank you Emily. Well, you tried to sabotage my experiments, but you know, I know how to be emotional indifferent and blablabla…
Thanks Anna, Elena, Jordi, Joao, Villy, for all the fun, and for not just being great colleagues but also good friends. Thanks Luisa, for being a great adventure companion, and for always sharing your food stocks!

And thank you, Raimon. TIO! YA ESTA'! te lo puedes creer? Yo no! my birthday buddy, and thesis buddy, and great science talker. Gracias por todo este tiempo que hemos pasados juntos aqui'.

As someone said *"I am gonna try with a little help from my friends",* and I need to thank to all of you, that near or far away have been supportive.
Fra and Cri, I don't even know where to start, but you have been just essential over here. Thank you. Same for Anna, that in one way or the other managed to be always around.
Thanks for all the betty fords' nights, Kiana, great drinking buddy, feminist comrade. Thanks to Anne, because "if it's crowded all the better, because we knew we were gonna be up late".
Soggetts. Grazie, for not trying to understand what the hell was I doing, for invading my life here multiple times and in multiple ways, and for being supportive, and friends, no matter what. You are my favorite dancing queens!
Alejandro, thanks for teaching me from day 1 the importance of stupidez en la ciencia.
Thank you to Bruno, and Zu. And to Laure, Marija, Klaus, all the crowd from Ramin's lab, for making me feeling welcome from the very beginning.

I need to thank everyone that, back when I decided to come here, supported me and pushed me to follow my way.
Thanks to my family, my grandparents, my aunt. For always being supportive. To my small-big Bro, to my Mum and Dad. Grazie. Mi sento persino scema a scrivere. You have been so unconditionally supportive and I felt you were close to me all this time. Thank you.

And thank you Peter. For many things, all represented in the coffee you bring me every morning.

Final thanks go to, in alphabetical order:
abba,asturies 66, el barrio de gracia, Belgian chocolate and beers, the beatles, the beer sessions, carders 14, collserola,, incuqina, kibuka, Murakami, la pasta madre, paris and the "final rush", all the people of the GRSC program, the phd student community, primavera sound, sant joan, shot_me, sonar, spotify, thesaurus, vampire weekend, wes anderson, zuluna y el yoga.

Grazie.

Thesis abstract

Intron removal is a necessary step for expression of most genes in higher eukaryotes, and alternative splice selection is a highly regulated mechanism that endows a single gene with the possibility to codify for multiple transcripts. Pre-mRNA splicing occurs largely co-transcriptionally, and its outcome is influenced by transcription elongation and chromatin structure. In this thesis we have used two different approaches to study novel links between chromatin structure and alternative splicing regulation.

In the first approach, we identified genome-wide progesterone-regulated cassette exons and compared them with nucleosome density profiles, with the aim of finding correlations between changes in nucleosome positioning and changes in alternative splicing. We find that, even if all exons harbor a well-positioned exonic nucleosome, four different classes of nucleosome density profiles can be identified around alternative exons, which strongly correlate with the DNA sequence GC content. Transitions between these profiles occur upon hormone stimulation and can be correlated with alternative splicing changes, although changes in nucleosome profiles are also observed in non-regulated exons. In particular hormone-induced exon inclusion is more frequently linked to changes in nucleosome density than hormone-induced skipped exons, which tend to have low nucleosome density profiles even before hormone treatment. Peaks of nucleosome density before alternative exons tend to correlate with exon inclusion.

In the second approach, we took advantage of ENCODE data of chromatin epigenetic signatures and RNA-Seq in multiple cell lines

to evaluate functional enrichment of histone modifications over alternative exons. We find that three histone modifications (H3K4me3, H3K27ac and H3K9ac) co-occur in a subset of exons when they are highly included. These features are sufficient to predict differential inclusion levels in other cell lines. Moreover, they are enriched in exons characterized by the presence of DNase hypersensitive sites, promoter signatures and RNA Pol II accumulation. These observations suggest a functional role for 3-dimensional genome structure in the regulation of alternative splicing.

Resumen de tesis

La eliminación de intrones es un paso necesario para expresar la mayoría de los genes en eucariotas superiores. La selección alternativa del corte y empalme (pre-mRNA splicing) es un mecanismo altamente regulado que dota a un solo gen con la posibilidad de codificar para múltiples transcritos. El splicing del pre-mRNA se produce en gran parte de manera cotranscripcional y por esto resultado está influenciado por la elongación de la transcripción y la estructura de la cromatina.
En esta tesis se han utilizado dos enfoques diferentes para estudiar nuevos vínculos entre la estructura de la cromatina y la regulación del splicing alternativo.

En el primer enfoque hemos identificado, a nivel de todo el genoma, exons internos regulados por progesterona y los hemos comparados con los perfiles de densidad de nucleosomas, con el objetivo de encontrar correlaciones entre los cambios en el

posicionamiento de nucleosomas y en el splicing alternativo. Hemos encontrado que, aunque todos los exones albergan un nucleosoma bien posicionado, se pueden identificar cuatro clases diferentes de perfiles de densidad de nucleosomas alrededor de exones alternativos, que se correlacionan fuertemente con el contenido G+C en la secuencia del ADN. Las transiciones entre estos perfiles se producen tras la estimulación con la hormona y se pueden correlacionar con los cambios en splicing alternativo, aunque se observan también cambios en los perfiles de nucleosomas en exones no regulados. La inclusión de exones inducida por la hormona está relacionada más a menudo con cambios en la densidad de nucleosomas que la exclusión. Estos exones excluidos tienden a tener perfiles de baja densidad nucleosomal incluso antes del tratamiento hormonal. Los picos de densidad de nucleosomas antes de exones alternativos tienden a correlacionarse con la exclusión de exones.

En el segundo enfoque nos aprovechamos de los datos de ENCODE de marcas epigenéticas de la cromatina y de RNA-Seq en múltiples líneas celulares, para evaluar el enriquecimiento funcional de las modificaciones de histonas en exones alternativos. Encontramos que tres modificaciones de las histonas (H3K4me3, H3K27ac y H3K9ac) co-ocurren en un subconjunto de exones mas incluidos. Estas características son suficientes para predecir los niveles de inclusión diferenciales de estos exones entre líneas celulares. Además, estos exones se caracterizan también por la presencia de sitios hipersensibles a la DNasa, de marcas de promotores y la acumulación de ARN Pol II . Estas observaciones sugieren un papel funcional para la estructura en 3 dimensiones del genoma en la regulación del splicing alternativo.

Preface

The emerging concept that chromatin dynamics and epigenetic modifications can influence and forecast alternative splicing decisions, shaping exon-intron architectures during evolution, is a powerful complement to predictive models based on regulatory sequence motifs identifiable in pre-mRNAs (Barash et al., 2010). Evidence, however, remains largely correlative in nature and rigorous mechanistic proof that recruitment of splicing factors through chromatin modifications can be rate-limiting for splicing decisions is not yet generally available.

In this thesis work, we aimed to better understand connections and interplay between chromatin structure and alternative splicing regulation using two distinct approaches. One of the approaches investigated links between progesterone-induced regulation of splicing and nucleosome occupancy. The second approach took advantage of recent massive datasets comparing epigenetic marks and RNA profiling in a variety of cell lines.

Table of Contents

Acknowledgements ..... **v**

Thesis abstract ..... **ix**

Resumen de tesis ..... **x**

Preface ..... **xiii**

Introduction **17**
1. Alternative pre-mRNA splicing 17
1.1 Co-trascriptional splicing 18
2. Exonic sequences in nucleosomes 20
2.1 Nucleosomes as "speed bumps" 22
2.2 Histone marks on exonic regions 24
2.2.1 Adaptors 26
2.2.2 Modifications affecting RNA pol II elongation 29
3. DNA methylation in exonic regions 31
4. Splicing reaches back 33
5. Steroid hormones 36
5.1 Progesterone Receptor 37
5.2 Progesterone and breast cancer 39

Objectives: **41**

Results-Part I **43**

Results-Part II **83**

Discussion **121**
Nucleosome occupancy and splicing regulation 122
Co-occurrence of promoter-like histone modifications and splicing regulation 125

Conclusions **127**
Part I 127
Part II 128

Bibliography **129**

Introduction

1. Alternative pre-mRNA splicing

Pre-mRNA splicing is the process by which intronic sequences are removed from primary transcripts and exonic sequences are spliced together to generate functional mRNAs. Alternative splicing is a highly regulated mechanism that endows a single gene with the capacity to generate multiple mRNAs, which can be subject to distinct regulation (e.g. by microRNAs) and/or encode proteins with different functions (Kelemen et al., 2013) The process makes a major contribution to proteome diversity in metazoans, with more than 90% of protein-coding genes undergoing alternative pre-mRNA processing in mammals (Pan et al., 2008; Wang et al., 2008).

Sequence elements at or near intron/exon boundaries are recognized by components of the spliceosome, the complex and dynamic cellular machinery in charge of intron removal, composed of five small nuclear ribonucleoprotein particles (snRNPs) and more than 150 proteins (Wahl et al., 2009). The binding of additional trans-acting factors (such SR- and hnRNP-proteins) to cis-regulatory sequence elements (splicing silencers and enhancers) located in intronic or exonic regions of the pre-mRNA facilitate or prevent spliceosome assembly on regulated splice sites (Chen et al., 2009). In this classical view, the interplay between positive and negative sequence elements and their cognate factors acting on a transcribed pre-mRNA determines the alternative splicing outcome (Barash et al., 2010).

1.1 Co-trascriptional splicing

Another dimension of alternative splicing regulation was revealed by the realization that pre-mRNA splicing and transcription are integrated processes. Electron microscopy images of nascent transcripts from Drosophila embryos showed for the first time that splicing can occur co-transcriptionally (Beyer and Osheim, 1988). Multiple studies have provided consistent evidence that introns can be removed while the nascent transcript is still tethered to the DNA through the RNA polymerase II complex (Ameur et al., 2011; Dye et al., 2006; Khodor et al., 2012; Khodor et al., 2011; Listerman et al., 2006; Pandya-Jones and Black, 2009; Tilgner et al., 2012; Vargas et al., 2011).

The fraction of introns that are spliced co-transcriptionally can vary between species, at least to some extent reflecting differences in intron-exon length architecture (Khodor et al., 2012). Nevertheless, even for yeast introns, which are typically short and located near the 5' end of the transcript, RNA polymerase pausing within terminal exons enforces co-transcriptional splicing of the majority of intron-containing genes (Carrillo Oesterreich et al., 2010).

Three aspects of co-transcriptional RNA processing can influence alternative splicing regulation. First, the timing of appearance of alternative splice sites during transcription influences their kinetic competition, and consequently differences in RNA polymerase II elongation rates can modulate alternative splice site choice (Kornblihtt, 2007) (Figure 1).

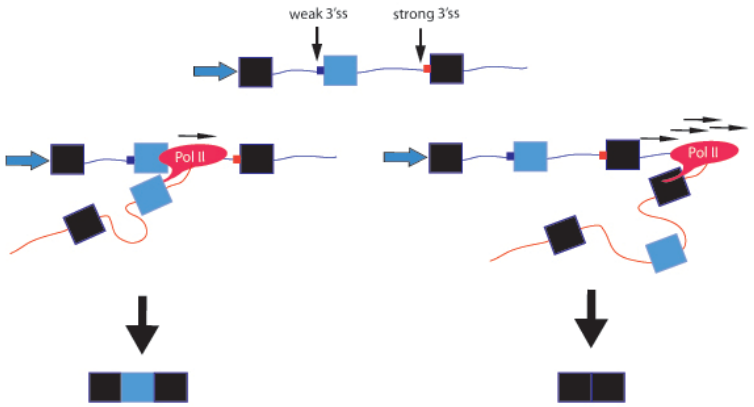


Fig.1 RNA polymerase II elongation rates can influence alternative splice site choices. In the model, an alternatively spliced internal exon flanked by weak splice sites is more included when the elongation rate of RNA polymerase II is low (left), while it is skipped when elongation is faster (right). Slower elongation rates provide more time for excision of the intron preceding the alternative exon before the alternative splice sites enter in competition. Faster elongation limits kinetic competition and favors alternative splicing decisions based on the relative strength of the alternative splice sites.

Second, components of the transcription and splicing complexes can mutually influence each other, modulating splice site recognition. In particular, the Carboxy-Terminal Domain (CTD) of RNA polymerase II interacts with a variety of splicing factors and therefore can couple transcription with RNA processing (McCracken et al., 1997). Third, given the essential role of chromatin in transcriptional control, chromatin structure and dynamics have been proposed to influence the coupling between transcription and splicing and consequently alternative splicing regulation (Kornblihtt, 2007; Luco et al., 2011). Two (not mutually exclusive) models have been proposed to functionally link

chromatin structure and alternative splicing regulation. One argues for direct recruitment of splicing factors on the nascent pre-mRNA, facilitated by interactions with histones and their modifications (Luco et al., 2011). Another model proposes that modulation of chromatin compaction (caused by changes in nucleosome positioning and/or by histone modifications) indirectly affects alternative splicing through changes in RNA polymerase II elongation rate (Kornblihtt et al., 2013).

2. Exonic sequences in nucleosomes

Nucleosomes are the structural units of chromatin: a single unit is formed by an octamer of histones around which DNA fragments of approximately 147 base pairs are wrapped. Interestingly, in contrast with the much longer and variable intron sizes, exon length seems to be under strong evolutionary pressure in vertebrates, remaining within a constant range of about 140 base pairs (Wang and Burge, 2008). Beckmann and Trifonov were the first to put forward the concept that chromatin structure may reflect the underlying intron-exon structure of the gene (Beckmann and Trifonov, 1991). They observed a 205-bp periodicity in between consecutive 3' and 5' splice sites and speculated that this could reflect a “nucleosomal repeat” that could possibly protect splice sites from mutational hazards.

Micrococcal nuclease digestion of chromatin followed by high-throughput DNA sequencing revealed that, indeed, regions protected from nuclease digestion are enriched in exons over introns, suggesting that exons represent regions of more stable nucleosome occupancy (positioning), a phenomenon that appears

to be evolutionarily conserved (Hodges et al., 2009; Nahkuri et al., 2009; Schwartz et al., 2009; Spies et al., 2009; Tilgner et al., 2009) (Figure 2). While some technical caveats remain (e.g. GC content is significantly higher in exons than in introns (Schwartz et al., 2009), which can influence patterns of Micrococcal nuclease digestion), several correlative features argue in favor of a functional role for nucleosomes in splice site recognition and exon definition. First, isolated exons flanked by long introns display higher nucleosome positioning than more clustered exons separated by small introns (Spies et al., 2009). Second, nucleosome occupancy levels increase with exon length, shifting towards the center of the exon (Tilgner et al., 2009). Third, while the pattern of nucleosome occupancy on Transcription Start Sites (TSS) and Transcription Termination Sites (TTS) is tightly connected with gene expression, nucleosome positioning at internal exons seems to be independent of transcription levels (Tilgner et al., 2009) and be higher in included than in excluded exons (Wang and Burge, 2008). Fourth, nucleosome occupancy is inversely proportional to the strength of the splice sites, suggesting that well-positioned nucleosomes can be particularly effective to assist in the recognition of weak splicing signals (Schwartz et al., 2009; Tilgner et al., 2009). Fifth, pseudoexons (intronic sequences with well-defined splice site sequences that are not normally included in mature transcripts) show a profile appearing as the mirror image of the occupancy observed in true internal exons, with pseudoexons depleted and the flanking sequences enriched in nucleosomes (Tilgner et al., 2009). Indeed, it has been proposed that nucleosome profiling can predict the levels of exon inclusion of pseudoexons (Tilgner et al., 2009). Collectively, these observations also suggest that, in addition to the evolutionary pressure to maintain coding capacity

and regulatory splicing motifs, enforcing nucleosome occupancy can be an important constraint for the evolution of exonic sequences (Schwartz et al, 2009; Wang and Burge, 2008).
Additional data suggest that the splicing process can, in turn, influence chromatin structure (see also the section “Splicing reaches back”). For example, the pyridimine-rich sequences preceding 3’ splice sites not only serve for the binding of essential splicing factors in the pre-mRNA, but may also act as “exclusion zones” for nucleosomes and thus influence nucleosome positioning in exons. Furthermore, recent results indicate that reinforcing 5’ splice site recognition has consequences on nucleosome occupancy (Keren-Shaul et al, 2013), again arguing for mutual influences between chromatin and splicing.

2.1 Nucleosomes as "speed bumps"

How can nucleosome pos itioning influence splice site recognition? One simple hypothesis is that nucleosomes represent physical barriers for the progression of RNA polymerase complexes and therefore that regions of high nucleosome density significantly reduce transcription elongation rates. Indeed, recent genome-wide studies of nascent transcripts document RNA polymerase II pausing at nucleosomes (Churchman and Weissman, 2011; Kwak et al., 2013) and at the 3' end of introns (Kwak et al., 2013), consistent with higher nucleosome density in exons.
Slowing down RNA polymerase can provide more time for the deposition on nascent pre-mRNAs of splicing factors associated with RNA polymerase complexes and therefore improve splice site

recognition, which can be particularly relevant for intrinsically weak splice sites (Muñoz et al., 2010). A second consequence of slowing down transcription elongation is the delay in the synthesis of 3' sequences corresponding to competing splice sites, thus influencing the timing at which alternative sites enter in kinetic competition (Figure 1) (de la Mata et al., 2010).

Consistent with these concepts, elegant work using an RNA polymerase II mutant that displays slower elongation rates demonstrated the influence of transcription elongation on alternative splicing decisions in vitro and in vivo (de la Mata et al., 2003) (Figure 1). Also consistent with the interplay between transcription elongation and splice site choice was the observation that the ratio between isoforms of an alternatively spliced region was different depending on whether transcription was activated by factors that promote RNA polymerase recruitment or by factors that promote elongation (Nogués et al., 2002). Even in *Saccharomyces cerevisiae*, an organism whose pre-mRNAs rarely contain more than one intron, skipping of an internal exon is enhanced by overexpression of an elongation factor and is reduced upon genetic or pharmacological inhibition of RNA polymerase II elongation (Howe et al, 2003). Still another example was provided by the observation that while both the Ewing sarcoma protein (EWS) and a fusion between EWS and the transcription factor FLI1 (characteristic of Ewing sarcomas) promote cyclin D1 gene expression, EWS-FLI1 decreases RNA pol II elongation rate and favors accumulation of the more oncogenic D1b isoform (Sánchez et al., 2008).

Two examples further illustrate the validity of these models. First, a study reported that DBIRD, a two-subunit complex that interacts with both RNA polymerase II and with nascent ribonucleoprotein

complexes (mRNPs), promotes RNA pol II elongation through A-T rich regions and facilitates skipping of alternative exons located within these regions (Close et al., 2012). The second example involves Brahma (Brm), a protein component of the SWI/SNF chromatin remodeling complex, which can influence nucleosome occupancy. Brm associates with various components and regulators of the spliceosome, slows down RNA polymerase II and favors inclusion of alternative exons (Batsche et al., 2006).

2.2 Histone marks on exonic regions

Core histones are globular proteins except for their long, unstructured N-terminal tails that can undergo a variety of post-translational modifications in multiple residues. Such modifications influence higher-order chromatin structure and the recruitment of non-histone proteins to chromatin (Kouzarides, 2007), consequently influencing transcription (Li et al., 2007).

Genome-wide analysis revealed preferential accumulation of particular histone modifications in exonic regions of genomic sequences (Andersson et al., 2009; Kolasinska-Zwierz et al., 2009; Schwartz et al., 2009; Spies et al., 2009). For example, trimethylation of lysine 36 in histone H3 (H3K36me3) was found to be enriched in exons compared to introns (initially in *C. elegans*, then confirmed in human and mouse), correlating with transcription levels. H3K36me3 levels increased towards the 3' end of genes and alternative exons displayed lower levels of H3K36me3 compared to constitutive exons (Andersson et al., 2009; Dhami et al., 2010; Kolasinska-Zwierz et al., 2009). H3K79me1 displays a similar profile, albeit less pronounced. Internal exons are also

enriched in H3K27me2, while H3K4me3, H3K27me1 and H3K36me1 signals are enriched only in internal exons flanked by long introns (Spies et al., 2009) (Figure 2). Specific marks for intronic regions, such as monoubiquitination of H2b, have been also found (Dhami et al., 2010; Shieh et al., 2011).

Apparent enrichment of histone modifications can be the consequence of the higher density of nucleosomes in exonic regions discussed in the previous paragraph. After normalization for nucleosome density, H3K36me3, H3K4me3, and H3K27me2 maintain their enrichment profile, while other modifications like H3K9me3 are comparatively depleted (Dhami et al., 2010; Spies et al., 2009) (Figure2).

The pattern of histone modifications profoundly influences chromatin accessibility and the recruitment of chromatin modifiers and transcription factors (Kouzarides, 2007). The density of epigenetic marks in exons could also influence splice site recognition and selection, and examples of this are emerging.

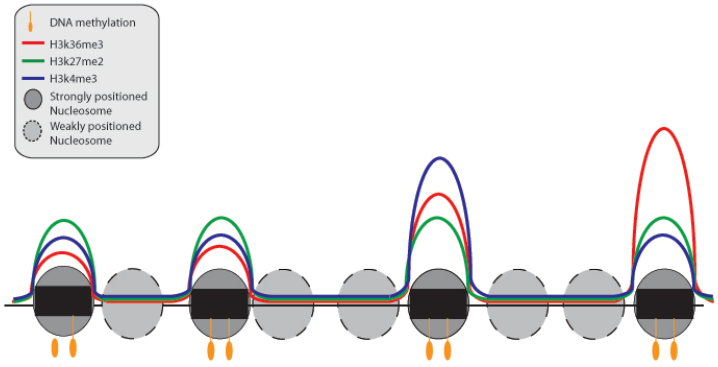


Figure 2. Distribution of chromatin features along the exon / intron architecture. The relative nucleosome density as well as the levels of epigenetic marks, including DNA methylation and histone modifications,

are represented for a model genomic region containing four internal exons. Enrichment in the histone modifications represented in the figure persists after normalization by nucleosome density. Notice that some marks (e.g. H3K36me3) increase in internal exons from the 5' to the 3' end of the gene, while others (e.g. H3K27me2) are more uniformly distributed. Other marks (e.g. H3K4me3) are enriched in particular exons (e.g. those flanked by long introns).

2.2.1 Adaptors

A common theme in transcription regulation is the role of adaptor molecules that on one hand recognize specific epigenetic marks and on the other interact with factors and enzymes involved in chromatin modification. An adaptor molecule that could play a similar role in splicing regulation is the protein MRG15, which recognizes H3K36me3 and interacts with the splicing regulator Polypyrimidine Tract Binding protein (PTB) (Luco et al., 2010) Receptor 2 (FGFR2) exon IIIb, which is included in epithelial cells, while the alternative exon IIIc is used in mesenchymal cells. Interestingly, these two cell types display differences in histone modifications in the alternatively spliced regions, with higher levels of H3K36me3 and H3K4me1 correlating with repression of exon IIIb / inclusion of exon IIIc, while the inverse correlation is observed for H3K27me3 and H3K4me3. Alteration of the levels of H3K36me3 or H3K4me3 (e.g. by depletion of enzymes involved in introducing these marks) result in altered splicing patterns of PTB-regulated exons. These results are consistent with a model in which the presence of histone tail modifications like H3K36me3 help to recruit, via adaptor molecules like MRG15, factors like PTB that regulate alternative splicing (Luco et al., 2010) (Figure 3).

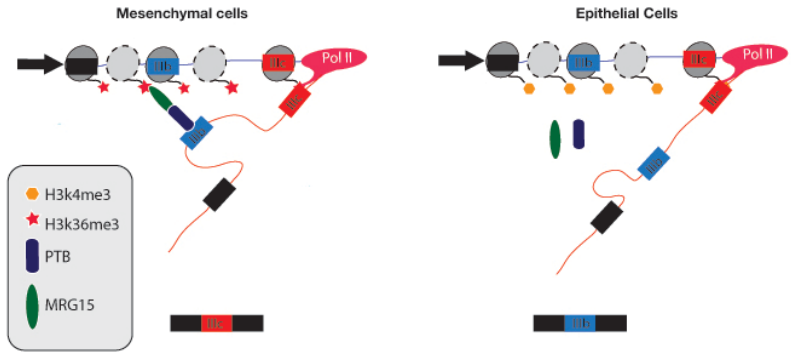


Figure 3. An example of adaptor models to explain functional connections between epigenetic marks and alternative splicing. The presence or absence in different cell types of two epigenetic marks in a genomic region corresponding to an alternatively spliced exon influences regulation by the splicing factor PTB. The adaptor protein MRG15 recognizes H3K36me3 and also interacts with PTB, thus helping to recruit this regulatory factor and promote skipping of the alternative exon.

Importantly, genome-wide analyses uncovered an inverse correlation between alternatively spliced exons affected by MRG15 and the strength of PTB binding sites in the target pre-mRNA (Luco et al., 2010), suggesting that the contribution of epigenetic marks to regulation is particularly important when intrinsic binding of the regulator to the target RNA is not strong enough to mediate its direct recruitment. It remains to be established whether epigenetic marks also contribute to the activation of the alternative exon IIIc.

H3K36me3 can also be recognized by Psip1/p52, which could serve as an adaptor for the recruitment of the serine/arginine-rich (SR) protein SRSF1. Indeed, knock down of Psip1/p52 affects particular alternative splicing events and reduces the association of SRSF1 with H3K36me3 (Pradeepa et al., 2012). SRSF1 and

SRSF3 associate with chromatin in interphase, while during mitotis their interaction is reduced by H3S10 phosphorylation (Loomis et al, 2009). Intriguingly, in mitotis SRSF1 was found associated with HP1, a characteristic heterochromatin component (Loomis et al, 2009). It was observed that depletion of the HP1γ (CBX3) isoform is associated with accumulation of unspliced nascent transcripts and deficient recruitment of splicing factors (Smallwood, 2012), Taken together, these data suggest that HP1γ could serve as an adaptor between heterochromatin marks (e.g. H3K9m39) and splicing factors, although effects for HP1 proteins on transcription elongation rates have also been reported (see below).

Another example of an adaptor molecule is CHD1, a chromodomain protein that recognizes H3K4me3 and interacts with SF3a, a subcomplex of U2 snRNP spliceosomal particles involved in early 3' splice site recognition. Decreased levels of H3K4me3 or CHD1 reduce the association of U2 snRNP with the pre-mRNA in vitro and in vivo, revealing a mechanism by which epigenetic marks can increase splicing efficiency (Sims et al., 2007).

Adaptor-based mechanisms have been also proposed to explain connections between histone acetylation and changes in alternative splicing (Luco et al., 2011; Shukla and Oberdoerffer, 2012). Studies in yeast revealed genetic interactions between the histone acetyltransferase GCN5 and protein components of U2 snRNP (Gunderson and Johnson, 2009) as well as links between spliceosome assembly and histone acetylation dynamics (Gunderson and Johnson, 2009; Gunderson et al., 2011). Treatment of HeLa cells with sodium butyrate, an inhibitor of Histone Deacetylases (HDAC), induced alternative splicing changes in 700 genes, which in the case of Fibronectin exon EDB,

skipping could be correlated with increased histone H4 acetylation and decreased recruitment of the activating factor SRSF5 (Hnilicova et al., 2011). In other genes, however, the inhibitor caused exon inclusion, suggesting a variety of mechanisms affected by HDAC inhibition (Hnilicova et al., 2011).

2.2.2 Modifications affecting RNA pol II elongation

In addition to adaptor-based mechanisms, several epigenetic modifications have been proposed to influence alternative splicing by modulating RNA polymerase elongation rates, along the lines discussed above for nucleosome positioning.

a) Histone acetylation. A clear example of the impact of histone acetylation on alternative splicing regulation through changes in Pol II elongation has been identified in the Neural Cell Adhesion Molecule (NCAM) exon 18 upon neuron membrane depolarization. Depolarization induces H3K9 acetylation, which is accompanied by more open chromatin, enhanced RNA polymerase II elongation and exon skipping, consistent with the kinetic model (Schor et al., 2009).

b) H3K9me3.
Several studies highlight the potential of H3K9 tri-methylation, which is typically associated with transcriptional repression and heterochromatin formation, on splicing regulation. One example is provided by PMA-mediated enhanced inclusion of exons in the variable region of the CD44 gene, which correlates with increased H3K9me3 and with binding of Heterochromatin (chromodomain)

Protein 1 gamma (HP1γ) in the region. HP1γ stabilizes the association of nascent transcripts with chromatin and reduces RNA polymerase II elongation rates, without affecting the overall expression of the gene (Saint-Andre et al., 2011). As mentioned above, however, HP1γ has also been implicated in the recruitment of splicing factors (Smallwood et al, 2012).

Other studies implicated the Argonaute proteins AGO1 and AGO2 in this mechanism, facilitating the recruitment of splicing factors and reducing RNA polymerase II elongation (Ameyar-Zazoua et al., 2012). Argonaute proteins mediate RNA silencing in the cytoplasm and also transcriptional silencing in the nucleus. For these activities, Argonautes associate with small double stranded RNAs that serve to direct silencing complexes to the target RNAs or DNAs (Filipowicz et al., 2005). Antisense transcription could be a source of double stranded RNA that could mediate the regulatory effects of Argonaute proteins in alternative splicing (Ameyar-Zazoua et al., 2012). Indeed, the activity of RNAse III Dicer, which is involved in the generation of Argonaute-associated small RNAs from longer double stranded precursors, is also required for the effects of Argonaute proteins in alternative splicing (Ameyar-Zazoua et al., 2012).

Importantly, small double stranded RNAs (siRNAs) directed against an intron of the fibronectin gene 3' of a regulated exon were found to induce exon inclusion in an AGO1-dependent manner, involving also an HP1 protein (HP1α) and H3K9me3 (Allo et al., 2009). These effects were not observed when targeting intronic sequences 5' of the regulated exon, and could be reversed by factors favoring chromatin opening or transcription elongation, suggesting a common mechanism for regulation by Argonaute proteins (Allo et al., 2009). Recent results in Drosophila, however,

revealed that AGO2 regulates over 100 alternative splicing events and that these effects are independent of AGO2 catalytic activity and are possibly associated with direct RNA binding of the protein to G-rich sequence motifs, suggesting additional mechanisms for AGO proteins in alternative splicing regulation (Taliaferro et al., 2013)

3. DNA methylation in exonic regions

Genome-wide analyses in plants and human cells revealed an enrichment of DNA methylation (5-methyl cytosine, 5-mC) in nucleosome-associated DNA and, consequently, in exonic compared to intronic regions (Chodavarapu et al., 2010; Hodges et al., 2009). DNA methylation correlates with closed chromatin structures and, as a consequence, with reduced RNA polymerase II elongation (Lorincz et al., 2004), although intragenic methylation associated with increased transcription has also been reported (Rauch et al., 2009).

A detailed mechanism has been proposed for the DNA methylation-induced skipping of CD45 exon 5 during lymphocyte development (Shukla et al., 2011). The mechanism involves binding of the chromatin insulator protein CTCF, which recognizes specific genomic sequences within exon 5, slows down RNA polymerase II elongation and promotes exon 5 inclusion. DNA methylation prevents CTCF binding, allowing faster elongation of RNA polymerase II and therefore exon 5 skipping (Shukla et al., 2011) (Figure4). Genome-wide analyses of CTCF binding and splicing changes upon CTCF depletion suggest that the interplay between CTCF binding and DNA methylation at the 5' of exons can

regulate a substantial number of alternative splicing events (Shukla et al., 2011).

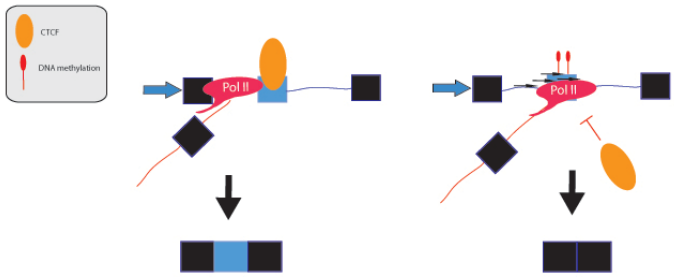


Figure 4. Model for alternative splicing regulation by DNA methylation and CTCF binding. Binding of the CTCF protein to genomic sequences reduces RNA polymerase II elongation, which favors exon inclusion (Figure 1). Methylation of CTCF binding sites prevents binding of the protein to the DNA and therefore faster elongation rates, which favor exon skipping.

Other observations support a role for DNA methylation in alternative splicing regulation. Analysis of methylation patterns in honeybees found that 5-mC is almost exclusively restricted to exons and absent from intronless histone-encoding genes. Remarkably, increased 5mc in workers compared to queens is associated with the production of a short variant of a transcript which is almost absent in queens (Lyko et al., 2010).

A recent analysis of different exon-intron architectures characterized by genomic regions of distinct GC content revealed a clear enrichment of DNA methylation in exons, particularly in genes with leveled GC content between exons and introns (Gelfman et al., 2013). Significantly, alternative exons were found to display

lower enrichment of 5-mc compared to constitutive ones (Gelfman et al., 2013), reinforcing the notion that DNA methylation can contribute to mark exon recognition.

Interestingly, a recent study found that while 5-mC is enriched at exon-intron boundaries in non-neural tissues, hydroxymethyl cytosine (5-hmC) is enriched at these boundaries in brain, with constitutive exons showing higher levels of 5-hmC compared to alternatively spliced exons in human frontal cortex (Khare et al., 2012). These observations raise the possibility that 5-mC and 5-hmC play distinct, perhaps tissue-specific roles in splice site recognition and regulation, although wider use of techniques able to distinguish between these modifications will be needed to evaluate this possibility.

4. Splicing reaches back

The mechanisms described so far postulate effects of nucleosome density or histone modifications on RNA polymerase elongation and/or recruitment of splicing factors. Recent evidence indicates that recruitment of the splicing machinery can in turn influence chromatin structure/dynamics and transcription and therefore modulate alternative splicing through these effects (de Almeida and Carmo-Fonseca, 2012)

a) Effects on chromatin.

Correlations between decreasing nucleosome occupancy and a switch in splicing from exon inclusion to skipping have been documented using a plasmid reporter (Keren-Shaul et al., 2013). Using this system, it was shown that strengthening the 5' splice site

associated with the alternative exon led to an increase in exon inclusion and, importantly, also to an increase in nucleosome density in the region, suggesting that splicing activity can influence chromatin organization (Keren-Shaul et al., 2013).

Similarly, levels of H3K36me3 can be affected by splicing activity, as illustrated by the effects of mutation of 3' splice sites in a β-globin reporter or by inhibition of splicing by treatment with the antitumor drug Spliceostatin A (Kim et al., 2011), which binds to and inhibits components of U2 snRNP (Bonnal et al., 2012). A shift in H3K36me3 profiles away from the 5' end and towards the 3' end of the gene is observed under these conditions, suggesting that splicing activity influences patterns of histone modification (Kim et al., 2011). Mechanistically, splicing can influence H3K36me3 levels by favoring the recruitment of the HYPB/Setd2 methyltransferase to the CTD of the RNA pol II (de Almeida et al., 2011).

Finally, another study found correlations between levels of active chromatin features, transcription and the length of the first exon (Bieberstein et al., 2012). This work supports a model in which recognition of the first splice site in a pre-mRNA contributes to the recruitment of activating H3K4me3 and H3K9ac chromatin marks at the promoter, accurate initiation of transcription and gene activity (Bieberstein et al., 2012).

b) Effects on transcription.

Spliceosome assembly was found to be coupled to RNA polymerase II dynamics at the 3' end of human genes, with splice site mutations leading to polymerase stalling and retention of unspliced RNAs at the site of transcription (Martins et al., 2011). This is similar to the splicing-dependent stalling of polymerase observed at the terminal exons of intron-containing genes in yeast

(Carrillo Oesterreich et al., 2010). These findings suggest that intron removal is required for transcription termination and transcript release, which could in turn provide a proofreading mechanism for splicing completion. Other precedents of function of splicing on transcription have been reported in budding yeast, where RNA polymerase II was found to accumulate transiently around the 3' end of introns depending on the presence of a functional 3' splice site, and upon appearance of spliced products the pausing is released (Alexander et al., 2010). Interestingly, genome-wide analysis of pausing sites in Drosophila also detects accumulation of RNA polymerase II near the 3' end of introns, (Kwak et al., 2013), which could conceivably be due to nucleosome positioning in exons and/or to a splicing proofreading mechanism like the one we just discussed for budding yeast.

Direct effects of splicing factors on transcription elongation have also been reported. These include core splicing components (U snRNPs) as well as splicing regulators (e.g. the SR protein SRSF2) that recruit elongation factors, particularly the P-TEFb complex, involved in CTD Ser 2 phosphorylation and promote efficient transcription elongation (Fong and Zhou, 2001; Lin et al., 2008; Zhong et al., 2009). Recent work unveils a detailed molecular mechanism for this activity of SRSF2, which targets the frequent phenomenon of RNA polymerase II pausing in the proximity of promoters (Ji et al, 2013). The results reveal that SRSF2 is part of an elongation inhibitory complex which includes the non-coding 7SK RNA and P-TEFb. Upon binding of SRSF2 to promoter-associated nascent RNA, SRSF2 and P-TEFb are released from the 7SK complex and this facilitates elongation of promoter-proximal paused RNA polymerase II. This mechanism is similar to transcription activation in HIV viruses mediated by HIV Tat/TAR

proteins, thus offering a common model for transcription activation by RNA binding proteins operating in viral and cellular genes (Ji et al, 2013). In vivo analysis argue, however, that transcription elongation kinetics proceed independently of splicing, at least in some genes (Brody et al., 2011).

Splicing regulatory Hu proteins influence transcription changes through an indirect mechanism that involves modulation of chromatin structure. Hu proteins recognize uridine-rich sequences and function in neuron-specific alternative splicing (Zhu et al., 2006). A recent report revealed that recruitment of Hu to pre-mRNA sequences near target alternative exons in the Nf1 and Fas genes results in local inhibition of HDAC2 and, consequently, local increase in histone acetylation leading to enhanced RNA polymerase transcription rates and exon skipping (Zhou et al., 2012).

5. Steroid hormones

Steroid hormones receptors (SRs) comprise a family of nuclear receptors that act as ligand-inducible transcription factors. Steroid hormones are small cholesterol-derived hydrophobic molecules that can traverse the cell membrane and activate their nuclear receptors. They include androgen, estrogen, progesterone, mineralocorticoid, and glucocorticoid receptor, and fulfill a variety of physiological function (Beato, 1996). They share similar modular structure, consisting of a N-terminal regulatory domain, a highly conserved DNA binding domain, a flexible hinge region and a moderately conserved C-terminal ligand-binding domain (Montano et al., 1995). All SRs serve mainly as transcription factors,

activating or repressing gene expression of their targets. Upon hormone stimulation they get activated in a concerted process that involves several steps: dissociation of chaperone complexes, receptor dimerization, translocation into the nucleus and binding to specific sequences on the DNA (hormone responsive elements, HREs) of their target genes (Beato, 2008). Although the majority of SRs are present in the nucleus or cytoplasm, a small fraction is constantly bound to the plasma membrane, where it activates a signaling cascade: this was initially known as the “non-genomic” action of hormones; however, recent studies revealed an intense cross talk between kinase cascades and the transcriptional functions of SRs (Vicent et al., 2006).

5.1 Progesterone Receptor

As for the other SRs, upon binding of the hormone to its C-terminal domain, progesterone receptor (PR) gets activated and exerts its effect by binding to HREs and modulating gene expression.
The DNA-bound PR recruits a plethora of co-regulators, including chromatin modifiers that remodel nucleosome organization and recruit further interactors that mediate gene expression changes. We can generally distinguish two classes of co-regulators: co-activators and co-repressors, which respectively help to increase or inhibit transcription of target genes. The chromatin remodeling complex SWI/SNF is recruited as a co-activator, as well as members of the steroid receptor co-activator family (SRC) (Beato and Klug, 2000). These, in turn, serve as adaptors that mediate the recruitment of histone acetyltransferases (HAT) (Torchia et al., 1997; Voegel et al., 1998). Conversely, the multiprotein complexes

NCoR and SMRT act as co-repressors, and exhibit deacetylase activity.

Hormone-mediated activation of the Mouse Mammary Tumor Virus (MMTV) has been extensively characterized. The provirus is integrated in the host DNA and responds rapidly to hormone stimulation by glucocorticoids or progestin. The hormone receptors bind to a cluster of HREs within the promoter and facilitate the interaction of other transcription factors whose have target sites are located between the HREs and the TATA box (reviewed in Vicent et al., 2010) SWI/SNF and SNF related complexes are recruited to the integrated MMTV promoter,, where upon PR binding selective displacement of histones H2A and H2B from the nucleosome positioned over one of the HRE occurs (Vicent et al., 2004; 2006)

How PR binds to its A targets is closely linked and influenced by nucleosome positioning and chromatin compaction. It is generally accepted that nuclear receptors bind preferentially to sequences located in nucleosome–depleted chromatin regions, marked by the so-called “pioneer factors” such as AP1 and FoxA1 (Hurtado et al., 2011; Carroll et al., 2005; John et al., 2011). A recent genome-wide analysis in breast cancer cells studied the relationship between transcriptional changes induced by hormones, PR binding and chromatin dynamics. The study revealed that PR binds preferentially to HREs that are enriched in nucleosomes prior to hormone induction, but also marked by DHS (DNAseI hypersensitive sites). Upon PR binding, nucleosomes get remodeled (but not evicted) and activation of gene expression begins (Ballare et al., 2013).

5.2 Progesterone and breast cancer

Ovarian steroid hormones 17β-oestradiol and progesterone are pivotal in the control of breast development and physiology ((Nandi, 1958)). ER and PR are expressed in 30-50% of luminal epithelial cells, which act as “sensors” for the hormone stimulation and send signals to the “responders” cells, which do not have receptors for these hormones (Clarke et al., 1997). Cell proliferation is induced in the ER+ cells but it is even more sustained, through paracrine signaling, in the neighboring ER-/PR- cells (Beleut et al., 2010). Both experimental and epidemiological studies revealed a strong link between these steroid hormones and mammary carcinogenesis. Sustained exposure to progesterone and activation of its downstream targets (cyclinD1, WNT4 and RANKL) has been proven to increase breast cancer risk. It is then essential to dissect the molecular mechanisms triggered by progesterone stimulation (reviewed in Brisken, 2013).

Objectives:

The observation that splicing can occur co-transcriptionally, and that chromatin dynamics and epigenetic modifications can influence and forecast alternative splicing decisions, complement the classic view of splicing regulation being limited to RNA *cis*-regulatory sequence motifs and their cognate binding factors. Evidence for the interplay between chromatin and splicing, however, remains correlative. In this thesis work we aim to:

1-Describe and understand the connections between progesterone-induced nucleosome density changes and alternative splicing regulation

2-Identify across multiple cell lines histone modifications enriched or depleted over alternatively spliced exons and discover whether they correlate with and influence exon inclusion levels

Results-Part I

Manuscript:

Relationship between nucleosome positioning and alternative splicing in progesterone-induced breast cancer cells

Camilla Iannone[1,2], Andrew Pohl[1,2], Panagiotis Papasaikas[1], Daniel Soronellas[1], Guillermo P. Vicent[1], Miguel Beato[1] and Juan Valcárcel[1]*

1 Centre for Genomic Regulation (CRG) and UPF, Dr. Aiguader, 88, 08003 Barcelona, Catalonia, Spain
2 Departament de Ciencies Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona.

* To whom correspondence should be addressed

Iannone C, Pohl A, Papasaikas P, Soronellas D, Vicent GP, Beato M et al. Relationship between nucleosome positioning and progesterone-induced alternative splicing in breast cancer cells. RNA. 2015 Mar; 21(3): 360-74. DOI: 10.1261/rna.048843.114

Summary

Splicing of mRNA precursors can occur co-transcriptionally and it has been proposed that chromatin structure influences splice site recognition and regulation. Here we have systematically explored potential links between nucleosome positioning and alternative splicing regulation upon progesterone stimulation of breast cancer cells. We confirm preferential nucleosome positioning in exons and report four distinct profiles of nucleosome density around alternatively spliced exons. RNA polymerase II accumulation closely follows nucleosome positioning. Hormone stimulation induces switches between profile classes, correlating with a subset of alternative splicing changes. In particular, peaks of nucleosome density in the intronic region preceding alternative exons correlate with exon inclusion. In contrast, exons skipped upon hormone stimulation display low nucleosome densities even before hormone treatment, suggesting that chromatin structure primes alternative splicing regulation. Collectively, our results suggest that a variety of chromatin architecture mechanisms influence alternative splicing decisions.

Introduction

Eukaryotic genomes are tightly packaged inside the nucleus in the form of chromatin nucleoprotein complexes that modulate the access of the DNA to replication, recombination and transcription factors and therefore play an important role in the regulation of gene expression. The primary unit of chromatin is the nucleosome, composed of an octamer of four different histone proteins around which ~147 bp of DNA are wrapped (Luger et al., 1997). Histones are small, globular proteins characterized by an unstructured tail protruding outside of the nucleosome core, which can be post-transcriptionally modified in a highly dynamic manner (Kouzarides, 2007). The state of these modifications is essential to regulate chromosome compaction, transcription and gene silencing.

The affinity of DNA for histone octamers is influenced by its primary sequence, which is therefore an important determinant of nucleosome positioning along the genome (Segal et al., 2006; Kaplan et al., 2010). Nevertheless, nucleosome positioning is highly dynamic and its regulation is crucial to control chromatin accessibility and recruitment of chromatin modifiers and transcription factors. Genome-wide studies correlated chromatin structure with the exon-intron organization of the genes and showed that nucleosomes are preferentially positioned in exons compared to introns in a variety of species from worms to humans (Nahkuri et al., 2008; Schwartz et al., 2009a; Spies et al., 2009; Tilgner et al., 2009). While the pattern of nucleosome occupancy at the transcription initiation and termination sites (TSS and TTS, respectively) is tightly connected with gene expression levels,

nucleosome positioning over internal exons seems independent of the transcription changes and levels of expression of the gene (Tilgner et al., 2009). It has also observed that exons with weak splice sites display stronger nucleosome densities, suggesting a possible role for this chromatin feature in exon recognition (Tilgner et al., 2009).

About 90% of human genes are alternatively spliced (Pan et al., 2008; Wang and Burge, 2008) both in a cell/tissue-specific manner and in response to stimuli. Defects in alternative splicing regulation are linked to different diseases (reviewed in Singh and Cooper, 2012)). Therefore intron removal needs to be both precise and plastic and thus is tightly controlled at multiple levels. In addition to the assembly of core components of the spliceosome to splice site signals, additional factors bind to intronic and exonic sequences of the pre-mRNA to regulate the inclusion of specific exons (Chen and Manley, 2009). The interplay between these sequence elements and their binding factors is the main determinant of alternative splicing outcome (Barash et al., 2010). Nonetheless, intron removal happens in large part co-transcriptionally (Khodor et al., 2012; Khodor et al., 2011; Tilgner et al., 2012), and thus chromatin structure and transcription dynamics can also be involved in the regulation of exon.

Polymerase II kinetics can be a determinant of the regulation of specific alternatively spliced exons. Slow elongation rates can favor inclusion of exons with weak splice sites by providing them with a kinetic advantage over competing downstream sites (de la Mata et al., 2003; de la Mata et al., 2010; de la Mata et al., 2003; de la Mata et al., 2010). Moreover, the Carboxy-Terminal-Domain (CTD)

of RNA pol II interacts with different splicing factors at transcription loci and can help to recruit them to nascent transcripts (Bird et al., 2004) . Furthermore, histone modifications can serve as a landing platform for “adaptor” proteins that recruit splicing factors (Luco et al., 2010; Saint-Andre et al., 2011), or indirectly impact splicing outcomes by modulating Pol II elongation rate through changes of level of compaction of chromatin (Schor et al., 2009).

It has been proposed that nucleosome positioning over alternative exons could regulate pol II processivity and consequently exon inclusion (Schwartz et al., 2009), but the general effect of nucleosome density on alternative splicing regulation has not been investigated systematically. We have explored this issue by analyzing in parallel changes in nucleosome positioning and alternative splicing induced by progesterone.

Progesterone is a steroid hormone that can traverse the cell membrane and activate its nuclear receptors (PR), which act as ligand-inducible transcription factor (Beato, 1996). A small fraction of PRs is also bound to the plasma membrane, where upon ligand binding activates signaling cascades. This was initially known as the “non-genomic” pathway; however recent studies revealed an intense cross talk between kinase cascades and the transcriptional functions of PRs (Vicent et al., 2006; Lange, 2004). Upon binding of the hormone to its C-terminal domain, progesterone receptor (PR) gets activated, binds to its DNA target sequences (HREs) and modulates gene expression by recruiting a plethora of co-regulators, among which chromatin modifiers that remodel nucleosome organization (Beato and Klug, 2000).

As for the other transcription factors, PR binding is strongly influenced by nucleosome density and chromatin compaction. PRs are thought to bind preferentially to HREs exposed by nucleosome–depleted chromatin regions, in which “pioneer factors” such as AP1 and FoxA1 are already bound (Hurtado et al., 2011; Carroll et al., 2005; John et al., 2011). A recent genome-wide analysis of global correlations between PR binding and chromatin dynamics induced by hormones in breast cancer cells (Ballare et al., 2013) revealed that PR binds preferentially to HREs that are at the same time enriched in nucleosomes prior to hormone induction and marked by DHS (DNAseI hypersensitive sites). Upon PR binding, nucleosomes get remodeled (but not evicted) and activation of gene expression begins (Ballare et al., 2013).

Our results confirm the higher density of nucleosomes associated with exons and its general decrease upon hormone stimulation. Clustering establishes four different classes of nucleosome density profiles that correlate with the distribution of GC content and that show a strong correlation with RNA polymerase II accumulation. Particular switches between nucleosome profile classes are correlated with hormone-induced alternative splicing. In particular high nucleosome density 5’ of alternative exons is associated with increased exon inclusion, while increased exon skipping is typically associated with low nucleosome densities even before hormone treatment. Our results suggest that changes in nucleosome positioning can influence alternative splicing decisions by a variety of mechanisms, not limited to influences of RNA pol II dynamics.

Results

Progesterone induces alternative splicing changes largely independent of gene expression changes

T47D-MMTVluc are breast cancer cells containing a single copy of the Mouse Mammary Tumor Virus promoter fused to a luciferase reporter, frequently used to study progesterone-induced transcriptional regulation (REFs). These cells were treated with 10nM of the synthetic progestin R5020. Total RNA was extracted after 6h of treatment (R6) or from untreated cells (T0) and high-throughput RNA-sequencing was carried out using Paired-end reads (see methods). To identify alternative exons regulated by R5020, we analyzed RNA-Seq data using the Mixture of Isoform Probabilistic Model (MISO) (Katz et al., 2010-See methods) focusing only on internal cassette exons. We identified 248 R5020-regulated alternative exon cassettes in 239 genes. RT-qPCR and semiquantitative analysis confirmed the alternative splicing changes in 5 out of 7 events analyzed. Progesterone treatment was associated with a similar number of higher inclusion (124 exons, hereto forth “R5020-included” exons) and higher skipping (123 exons, hereto forth “R5020-skipped” exons) events (Figure 1A).

It has been previously reported that R5020 treatment modulates transcription of 3835 genes (Ballare et al., 2013). However, the majority of R5020-regulated cassette exons occur in genes that are not transcriptionally regulated by the hormone, as determined by global microarray analysis and RNA-seq data (Ballare et al., 2013) (74% for genes with included exons, and 73,7% for genes with

skipped exons) (Figure 1B). Among the genes displaying changes in both gene expression and alternative splicing, a preference was observed for included exons to be more frequent in down-regulated genes and for skipped exons to be more frequent in upregulated genes, perhaps related to effects of changes in transcriptional elongation. Of possible relevance for the mechanisms behind the changes in alternative splicing, expression of a number of splicing regulatory factors was observed upon hormone induction (Table 1).

Gene Ontology analysis using Gorilla software suggested that genes displaying alternative splicing changes belong to functional categories distinct from those displaying hormone-induced transcriptional changes (Figure 1C). This is in line with previous observations in a variety of systems suggesting that gene expression and alternative splicing coordinate different layers of cellular function (Pan et al., 2004). Neither RNA processing nor RNA binding proteins were among the GO categories enriched in regulated genes or alternative splicing events.

Previous ChIP-Seq results identified over 25000 DNA progesterone receptor (PR) binding sites in the genome, occupied at different time points after R5020 treatment, Progesterone-modulated genes show an enrichment in PR binding sites compared to non-regulated genes, some of which nevertheless also show PR binding upon hormone stimulation. To understand whether PR binding could influence alternative splicing regulation, we looked for PR binding sites around R5020-regulated cassette exons. Less than 3% of the regulated exons had a ChIP-Seq signal in a window of 3000 base pairs flanking the alternative exon. Around 40% of genes containing regulated exons display PR binding sites anywhere in

the gene body or within 3000 base pairs upstream of the TSS, at any of the analyzed time points. This distribution is similar for regulated and not regulated genes (Ballare et al., 2013) and for genes with regulated or not regulated exons. We conclude that PR binding is unlikely to substantially contribute to the changes in alternative splicing observed,

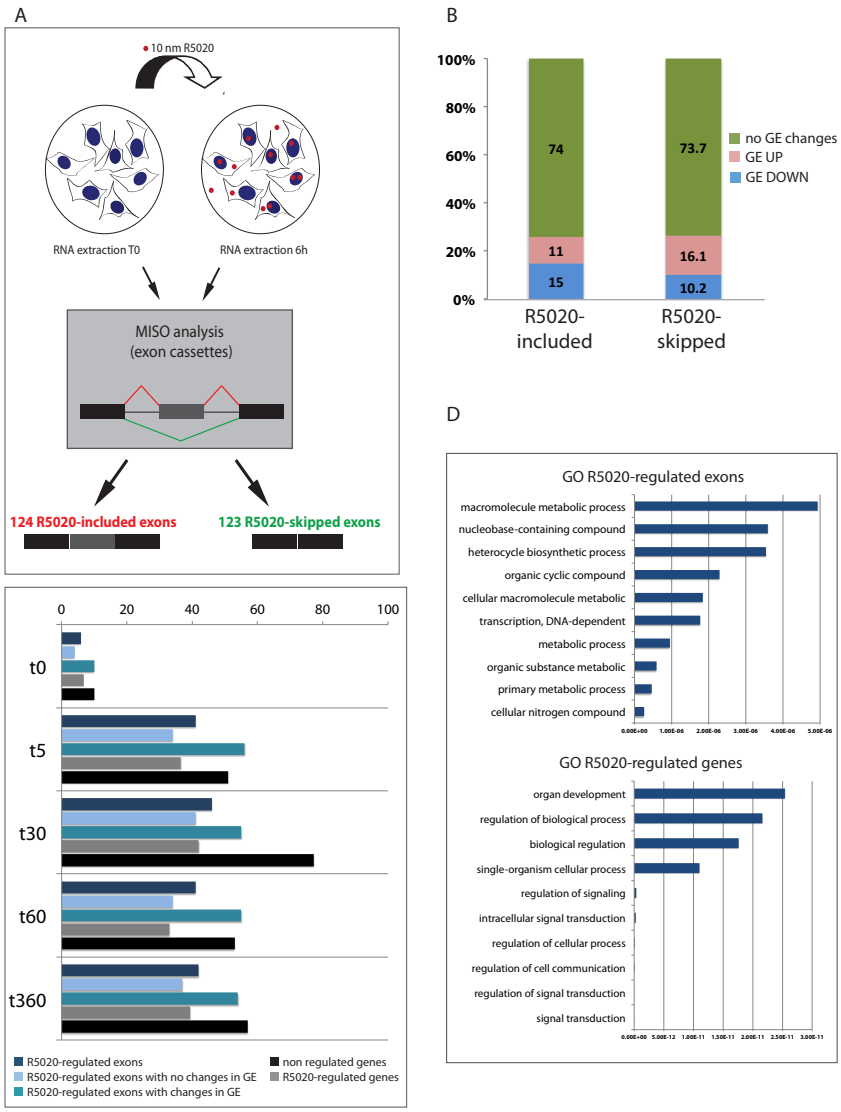


Figure 1. Progesterone-induced alternative splicing changes in T47D breast cancer cells. A. Scheme summarizing the experimental procedure and number of alternative splicing changes identified. B. Fraction of alternative splicing changes (increased inclusion or skipping) that occurs in genes without gene expression changes (no change) or in upregulated (up) or downregulated (down) genes. C. Fraction of genes displaying progesterone receptor (PR) binding sites at different times after hormone treatment. The different gene categories are represented with different colors, as indicated. D. Gene Ontology terms enrichment for progesterone-regulated genes and exons. The statistical significance of the GO terms enrichment are indicated.

Table 1

HnRNP/SR protein	
HNRNPA0	UP
HNRNPAB	UP
HNRNPH3	UP
HNRNPM	UP
SRSF10	DOWN
SRSF2	UP
SRSF3	UP
TRA2B	UP
SRRM2	UP
Core spliceosome	
CCAR1	DOWN
DHX16	UP
DHX8	UP
RBM5	DOWN
SNRPE	DOWN
U2AF1	UP
WBP4	DOWN
CRNKL1	DOWN
Varius splicing related	
DDX42	UP
EIF4A3	UP
FRA10AC1	DOWN
GEMIN4	UP
LSM5	DOWN
MBNL3	DOWN
PRPF38B	UP
RBM23	DOWN
RBMX	DOWN
STRBP	UP
TARDBP	UP
THOC5	UP
ZNF830	DOWN
GEMIN5	UP
PCBP4	UP
DDX26B	DOWN
Chromatin remodelling splicing related	
BRD4	UP
HDAC4	DOWN
MBD2	UP
TET1	DOWN
EIF2C2	UP
KAT2B	UP
PHC2	UP
RPS6KA5	UP
SUV39H1	UP
BMI1	DOWN

Nucleosome occupancy and RNA Polymerase II dynamics in alternatively spliced genomic regions

To correlate genome-wide, high resolution maps of nucleosome density with alternative splicing changes, we analyzed data produced by Ballaré et al. (2013) obtained by sequencing mononucleosomes generated by MNase digestion of chromatin in permeabilized T47D-MMTVluc cells untreated or treated with R5020 for 60 minutes.

We first analyzed the distribution of nucleosome density over all internal exons. In agreement with previous reports (Nahkuri et al., 2008; Schwartz et al., 2009a; Spies et al., 2009; Tilgner et al., 2009), a clear peak of nucleosome density centered on the exon and surrounded by regions of lower density in the flanking introns was observed (Figure 2A). This enrichment was detected both in control cells and in cells treated with R5020, but a clear decrease in the exonic nucleosome peak was observed upon hormone treatment, concomitant with a slight increase in nucleosome density in the flanking intronic regions (Figure 2A).

To evaluate whether nucleosome occupancy changes can be correlated with alternative splicing regulation, we investigated nucleosome occupancy of R5020-regulated exon cassettes and of their flanking exons in treated and untreated samples (Figures 2 B-C). Both R5020- included and R5020-skipped exon cassettes display higher nucleosome occupancy on exonic regions compared to flanking introns in the absence of hormone treatment, although the peak of skipped exons is distinctly lower than those of the flanking exons. This suggests that hormone-induced skipped exons have distinctive features (relatively lower nucleosome peaks compared to their flanking introns than the rest of the exons) even

before hormone treatment. Upon hormone treatment, nucleosome peaks are generally reduced, except the peak of the alternative skipped exons, which is not reduced further. These patterns were similar for exons displaying splicing changes in genes whose expression was not altered by hormone treatment (Figures 2D, E), and for regulated exons changing more than two fold in inclusion or skipping (33 skipped and 42 included exons). Of relevance, exons that display inclusion changes higher than two-fold have relatively low inclusion ratios in the absence of hormone (0.16 on average, increasing to an average of 0.4 upon hormone treatment). This suggests that these exons have a clear nucleosome peak even though they show low basal levels of exon inclusion. Conversely, strongly R5020-skipped exons, which have an average exon inclusion ratio of 0.40 (which is reduced to 0.15 upon hormone treatment), display low nucleosome occupancy in the absence or presence of hormone. This suggests that these exons have intrinsic features associated with low nucleosome density. Taken together, the results suggest that particular profiles of nucleosome occupancy around alternatively spliced exons can influence their regulation.

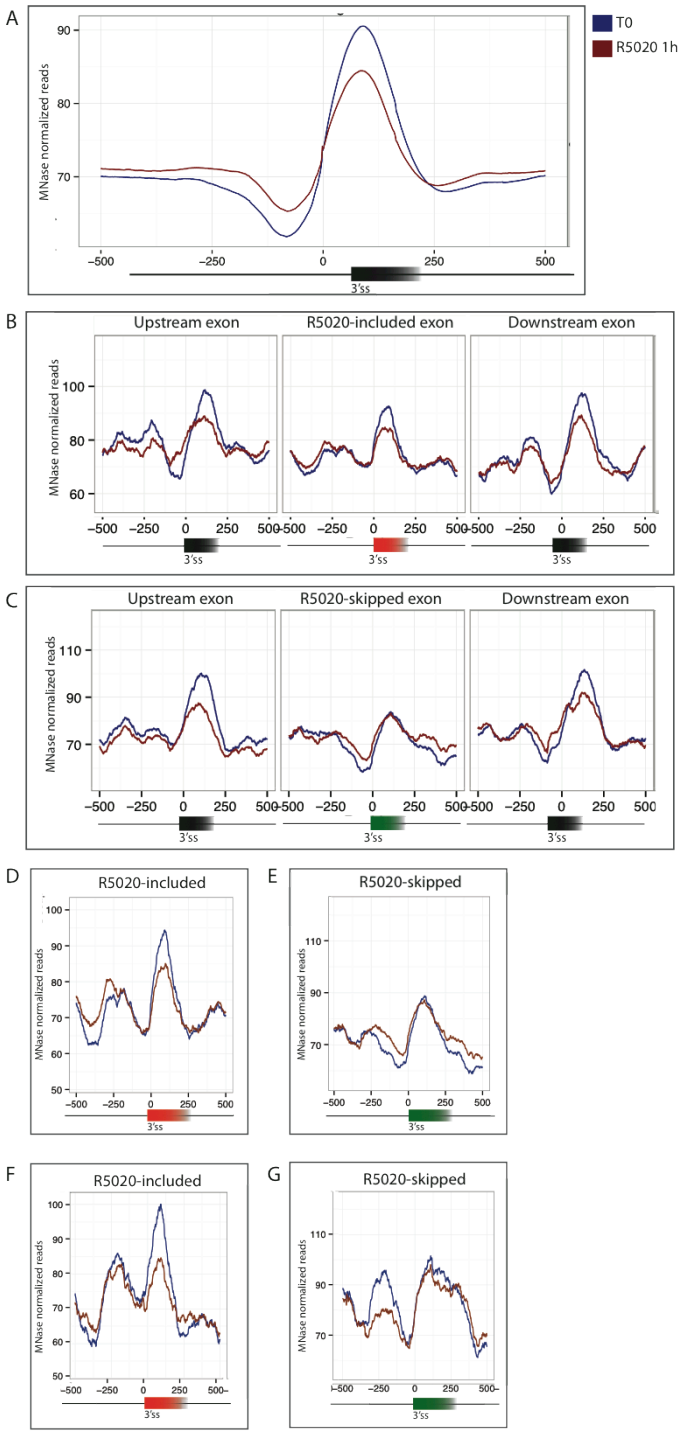

A
T0
R5020 1h
MNase normalized reads
3'ss
B
Upstream exon
R5020-included exon
Downstream exon
C
Upstream exon
R5020-skipped exon
Downstream exon
D
R5020-included
E
R5020-skipped
F
R5020-included
G
R5020-skipped

Figure 2. Progesterone-induced changes in nucleosome positioning in exonic regions. A. General profile of nucleosome density in exons and flanking sequences in control and hormone-stimulated T47D cells. Chromatin from permeabilized cells treated with MNase was isolated, mononucleosome-associated DNA was deep-sequenced and the number of reads covering each nucleotide are represented, after normalization by genomic DNA reads, relative to the 3' splice site of exons. B, C. Profiles of nucleosome densities in R5020-included (B) and R5020–skipped (C) exons and their flanking exons and introns in control and hormone-stimulated T47D cells. D, E. Profiles of nucleosome densities as in B-C, for alternative exons in genes that do not display changes in gene expression upon hormone stimulation. F,G. Profiles of nucleosome densities as in B-C for alternative exons that experience a more than 2-fold change in inclusion or skipping, as indicated.

Transcribing RNA polymerase II pauses at nucleosomes, (Churchman and Weissman, 2011; Kwak et al., 2013) and at the 3′ end of introns (Kwak et al., 2013) and its kinetics have been associated with alternative splicing regulation (de la Mata et al., 2003; de la Mata et al., 2010). Nucleosome positioning has been proposed to represent a barrier for RNA pol II progression, leading to reduced elongation rates that can provide more extensive opportunities to facilitate exon recognition by polymerase-associated splicing factors (Schwartz et al., 2009). In contrast, our results reveal decreased nucleosome density upon R5020-induced exon inclusion, and low densities with or without hormone in exons with induced skipping. In fact, skipped exons display a particularly pronounced peak of low nucleosome density in the 3' splice site region before the exon.

To investigate RNA Pol II dynamics in response to hormone induction over regulated exons we performed total Pol II ChIP-Seq in T47D-MMTVluc cells, either un-induced or treated for 60 minutes with 10 nM R5020. Meta-profile of the ChIP-Seq signal shows, as

expected, accumulation of RNA pol II on TSS and a very low signal in the gene body in both conditions (Figure S1). We investigated RNA Pol II accumulation over R5020-regulated exons and their upstream introns (bins of 150bp), and the same regions of the corresponding downstream constitutive exons as a control. RNA Pol II tends to accumulate in exonic regions compared to upstream intronic regions, consistent with nucleosomes acting as barriers for polymerase elongation. This is particularly clear in the constitutive exons and in R5020-included exons, with or without hormone treatment. In fact, hormone treatment accentuates this effect in included exons (P-value=0.0056). Consistent with the absence of a clear nucleosome peak associated with R5020-skipped exons, no accumulation of RNA pol II is observed in these regions, either in the presence or absence of the hormone (Figures 3A, B, D and E). These trends are less obvious in exons displaying more than two-fold changes in inclusion or skipping, perhaps because the number of exons involved in the analysis is not sufficient for statistical significance.

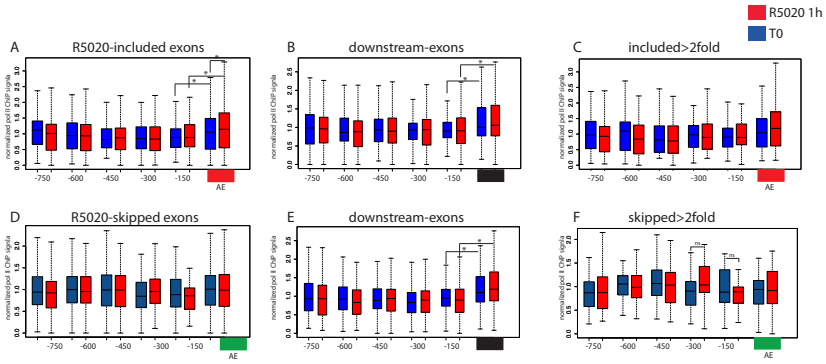
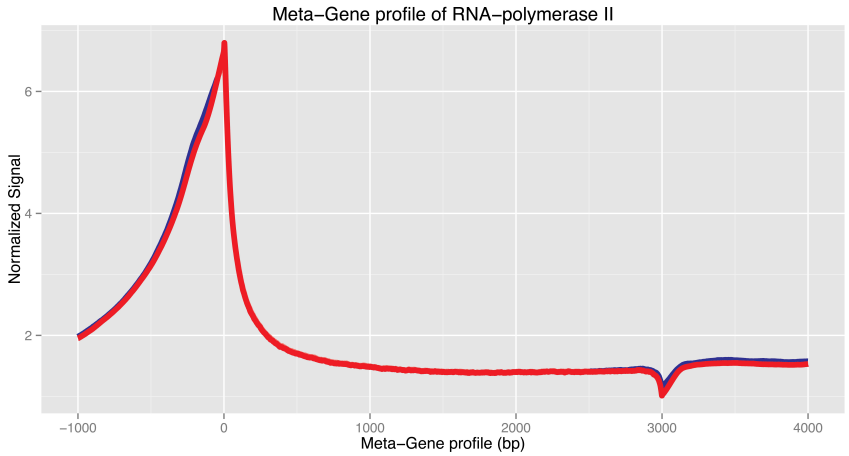


Figure 3. Distribution of RNA polymerase II in exons and upstream intronic regions. Boxplots represent the density of deep sequencing reads of DNA molecules associated with RNA polymerase II, corresponding to alternative exons and 750 nucleotides upstream, normalized to local average, in the absence (blue) or presence (red) of

R5020 for 60 minutes. Density maps correspond to R5020-included or –skipped exons (A and D, respectively), their constitutive downstream exon (B, E) and exons regulated by more than two-fold (C,F). */ns indicates p-value significative or not, calculated to Two-sample T-test.



Supplementary figure 1: Meta-profile of the ChIP-Seq signal of total RNA Polymerase II in untreated (blue line) and treated cells (red line) showsaccumulation of RNA pol II on TSS and reduced signal over the gene body. The two plots are globally overlapping.

Alternative exons can be classified according to their nucleosome profile

The pattern of nucleosome occupancy over internal exons has been so far described as a single sharp peak standing out on exons compared to the flanking introns (Nahkuri et al., 2009; Tilgner et al., 2009, Schwartz et al., 2009, Spies et al., 2009). Such patterns are generated by averaging the profile of nucleosomes across all the exons, possibly masking distinct profiles associated with particular subclasses of exons. We analyzed the MNase-Seq data of un-induced T47D-MMTVluc cells on all cassette exons whose expression is detectable in the MISO annotation, We

subsequently classified them in four different categories according to their profile of nucleosome occupancy using K-means clustering (see methods) (Figure 4a). While all the categories display a clear exonic nucleosome peak, distinct features were also revealed. Cluster A is the most populated class (45% of analyzed exons), and is characterized by a relatively low peak of nucleosome density on the alternative exon (Figure 4A). The remaining exons are similarly distributed among three other categories (B=18.6%, C=18.4% and D=17.9 %) (Figure 4C). Two clusters show higher nucleosome density either in the upstream intron (cluster B) or downstream intron (cluster C) compared to the cassette alternative exon (Figure 4A).. Cluster D exons are characterized by low intronic nucleosome density and a strongly positioned exonic nucleosome (Figure 4A).

To evaluate the possible impact of the nucleosome profiles characteristic of these categories on exon inclusion, we binned the alternative exons with the 5% lowest and highest inclusion level in untreated cells and analyzed their distribution in the four clusters. The distribution of exons in the different categories was essentially maintained, with cluster B (harboring a nucleosome peak before the exon) only slightly more frequent among less included exons, and more included exons displaying a slightly increased fraction in cluster A (Figures 4 D-F).

Analysis of splice site strengths showed that exons classified in cluster D (strongly positioned nucleosome in the exon) are usually associated with weaker 3' and 5' splice sites (Figure 4F), in agreement with previous reports that correlated higher nucleosome occupancy with weaker splice sites (Tilgner et al., 2009). We found that cluster B exons, characterized by high nucleosome occupancy

upstream of the exon, are associated with weaker 3' splice sites but not with weaker 5' splice sites, suggesting a role for nucleosome positioning upstream of the exon in 3' splice site recognition (data not shown). , The analogous correlation was not found for cluster C exons, harboring a nucleosome downstream of the exon, and weaker 5' splice sites (data not shown).
While there is no difference in average exon length, Cluster D exons (marked by strongly positioned nucleosomes) are associated with longer flanking introns. Clusters B and C are not enriched in exons with short introns, arguing against the possibility that upstream / downstream nucleosome accumulation was caused by peaks associated with other exons nearby (Figure 4H).

DNA sequence is a strong determinant of nucleosome positioning (Kaplan et al., 2010), and exons show a different GC composition compared to introns (Schwartz et al., 2009). We analyzed the average GC content distribution in the four different clusters and found a clear correspondence between peaks of high GC content and nucleosome occupancy. In fact, the distribution of GC content of each cluster mimics the corresponding nucleosome occupancy profiles, pointing to the DNA sequence as the main determinant for nucleosome distribution (Figure 4C).

Changes in nucleosome profiling associated with alternative splicing regulation

Next we analyzed whether hormone treatment induced changes in nucleosome profile that would represent switches between cluster classes. For this we compared the MNase-Seq data obtained from control cells and from R5020-induced T47D-MMTVluc cells and

clustered the alternatively spliced exons according to their nucleosome profiles. The same 4 main patterns of nucleosome occupancy were identified upon hormone treatment (Figure 4B), although the peaks were less sharp, as expected from to the less dense nucleosome profiles observed in all the exons (Figure 2A).

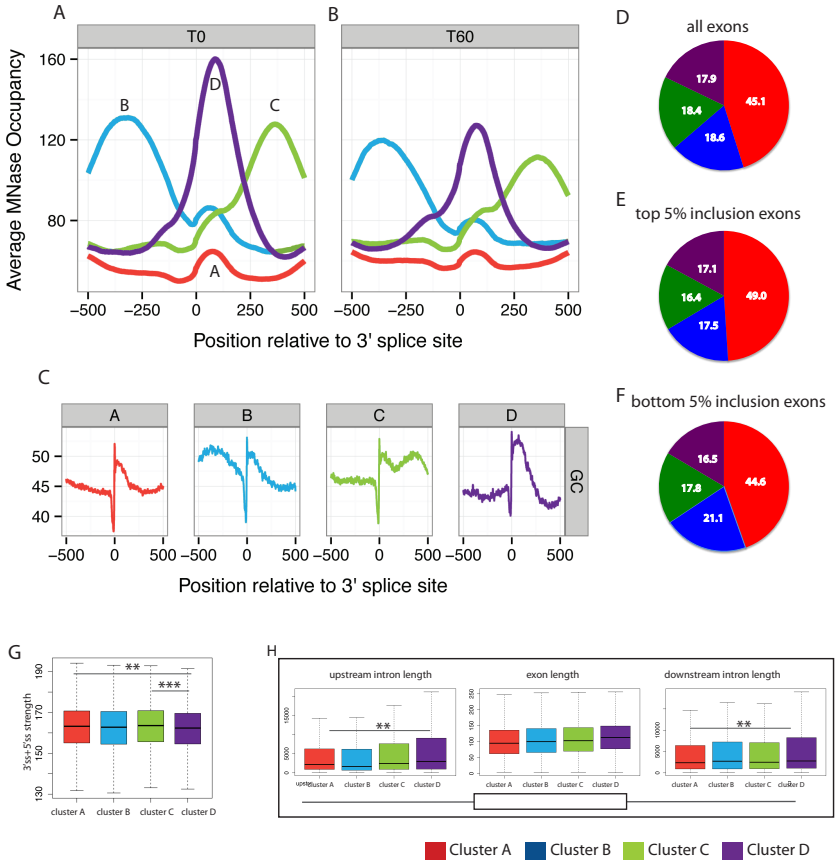


Figure 4. Clusters of nucleosome profiles around alternatively spliced exons. A, B. Four classes of nucleosome profiles identified in T47D cells before (A) and after (B) progesterone stimulation. Profiles of MNase reads relative to the position of 3' splice sites (0) are represented as in Figure 2 for four distinct clusters of exons before (T0) and 60 min after progesterone treatment (T60). C-E. Distribution of exons in the four nucleosome profile clusters for all exons (C), exons in the top 5% of exon

inclusion range (D) or exons in the bottom 5% of exon inclusion range (E). The colors correspond to each of the clusters defined in panels A-B. I. Distribution of splice site strengths associated with exons in the four classes of nucleosome profiles. Combined strength of 5' and 3' splice sites was measured by Analyze Splice Tool (http://ibis.tau.ac.il/ssat/SpliceSiteFrame.htm). Colors correspond to the clusters defined in A-B. ** and *** indicate p values of … and …, respectively, in the difference between slice site strengths in the indicated clusters according to two sample t-test H. Distribution of alternative exon and flanking intron lengths for the four clusters of nucleosome profiles defined in A-B, represented similarly as in F. I. Distribution of GC content (in percentage) relative to the 3' splice site of alternative exons in the four classes of nucleosome profiles defined in A-B. Colors correspond to the clusters defined in A-B.

Even though the classification of exons in different nucleosome occupancy patterns is mainly determined by GC content distribution (Figure 4I), 30% of all the exons analyzed switch the nucleosome profile class upon hormone treatment, indicative of active modulation of nucleosome occupancy (Figure 5A). The proportion is higher for R5020-included exons than for R5020-skipped exons, suggesting that nucleosome remodeling may play a stronger role in the regulation of exon inclusion than in skipping. The proportion of exons from each class changing their profile ranged between 20 and 29% (Figure 5B). We then focused on R5020-regulated cassette exons, and observed that 35% of the included and 23% of the skipped exons changed their cluster class after induction (Figure 5A). Globally, the number of exons in the more populated A class are slightly reduced, while the number of exons in clusters C and D increase slightly (Figure 5C). R5020-included exons generally increase in classes B, C and D (all of them harboring strongly positioned nucleosomes), while they decrease in class A (harboring low nucleosome peaks) (Figures 5B and D). In contrast, skipped exons increase in A and D (and exons in these categories tend not to change class, as only 7% of cluster

A exons and 8% of cluster D exons do change class) and decrease in B and C (exons in these categories tend to switch, 35% for cluster B, 44% for cluster C) (Figures 5 B and E).

We then focused our analysis on those exons undergoing nucleosome profile changes upon R5020 treatment, and looked for frequent / infrequent transitions correlating with hormone-induced alternative splicing changes. Figures 5F and G summarize the relative frequency of transitions associated with included or skipped exons. The majority of included exons switch from cluster A (poorly positioned nucleosomes) to C or D (strongly positioned nucleosome on the exon or downstream from it), suggesting again that increased nucleosome positioning can facilitate alternative exon recognition. Concerning skipped exons, favored shifts are from C to D / A, and from B to C / A, i.e. from strongly positioned nucleosomes upstream or downstream of the alternative exon to other categories. Interestingly, skipped exons were rarely associated with nucleosome positioning upstream of the alternative exon. This and the results of Figure 2G are consistent with nucleosome positioning upstream of the alternative exon contributing to exon recognition. The picture is however complex, because transitions to a well-positioned exonic nucleosome (cluster D) appear to be favored for all regulated exons, regardless of the inclusion/skipping outcome.

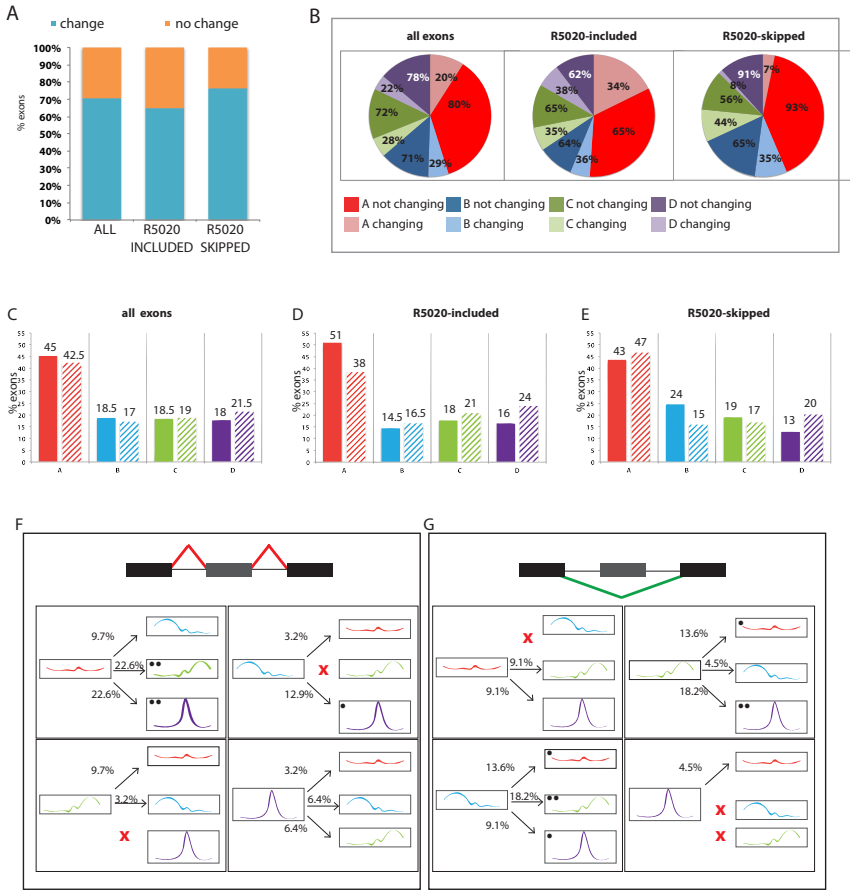


Figure 5. Switches in the distribution of nucleosome profile clusters upon hormone stimulation. A. Fraction of exons (total, R5020-induced and R5020 –skipped) switching or not between nucleosome profile clusters upon hormone stimulation. B. Fraction of exons in each class that switch (light colors) or not (dark colors) between nucleosome profile clusters upon hormone stimulation. Color codes are as in Figure 4. Values are represented for all, R5020-induced and R5020 –skipped exons. E-G. Distribution of exons in nucleosome profile clusters before (solid bars) and after (striped bars) progesterone treatment. Values are represented for all (C), E5020-induced (D) and E5020 –skipped (E) exons. Color codes representing each profile cluster are as in Figure 4. F-G. Distribution of nucleosome profile cluster switches for E5020-included (H) and E5020-skipped (I) exons. Classes are indicated by schemes of the profiles. The fraction of exons displaying each switch (out of the total number of exons switching profile) is indicated.

Next we analyzed profiles of RNA pol II accumulation over included and skipped exons, classified according to their nucleosome occupancy clusters before and after hormone induction. RNA pol II accumulation follows quite closely the profiles of nucleosome density, with polymerase peaks in the areas of higher nucleosome occupancy (Figure 6). The correlation is particularly clear in exons with strongly positioned nucleosomes upstream (cluster B) or in the alternative exon (cluster D). RNA Pol II accumulation is observed in these cases, however, regardless of the splicing regulation outcome. A peak of RNA pol II accumulation is particularly noticeable in R5020-included exons belonging to cluster C. As these regions harbor a nucleosome in the region downstream from the exon, upstream accumulation of RNA pol II may contribute to higher exon recognition.

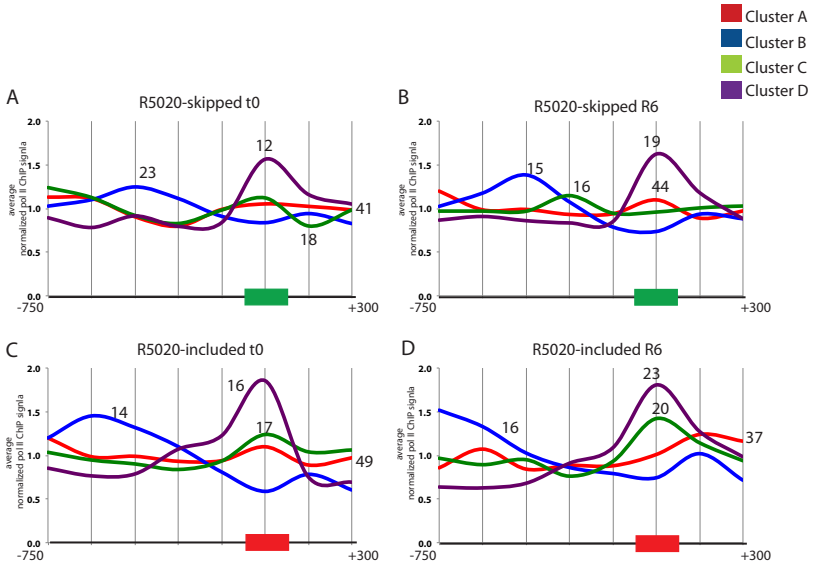


Figure 6. Distribution of RNA polymerase II in alternative exons and flanking introns. Average normalized RNA pol II ChIP-Seq reads are represented relative to the alternative exons and flanking introns in each

nucleosome profiling class (color code as in previous Figures), for R5020-skipped (A, B) or R5020-included exons (C,D), before (A,C) or 60 minutes after (B,D) hormone treatment.

Discussion

Regulation of alternative splicing of mRNA precursors is based upon the combinatorial effects of RNA motifs in the pre-mRNA and the activities of their cognate *trans*-acting factors (Smith and Valcarcel, 2000). It has become clear in recent years that chromatin structure and transcriptional dynamics play an important role in alternative splicing regulation (reviewed in Luco et al., 2011). A series of independent, concurrent studies put forward the concept that chromatin structure reflects the intron-exon structure, such that nucleosomes are especially enriched at exons compared to introns (Nahkuri et al., 2009; Tilgner et al., 2009, Schwartz et al., 2009, Spies et al., 2009). A straightforward and appealing hypothesis arising from these findings is that nucleosome density could play a role in the regulation of exon inclusion, serving either as “speed bump” for elongating polymerases and their associated splicing factors, or directly as a recruiting platform for splicing regulators. In this manuscript we wanted to test the conjecture that changes in nucleosome occupancy could contribute to alternative splicing regulation. One specific prediction was that increased nucleosome density would be associated with exon inclusion, while decreased nucleosome occupancy would be associated with exon skipping. While it has been reported that enrichment of specific histone modifications modulates alternative splicing regulation of some exon cassettes (Luco et al., 2010; Saint-Andre et al., 2011) it

remained unclear whether nucleosome positioning could indeed influence exon inclusion.

Our approach was to draw detailed maps of nucleosome density before and after hormone induction, and compare them with changes in alternative splicing induced by hormone. To simplify the picture we focused on internal cassette exons. Because high exposure to progesterone has been linked to mammary carcinogenesis, a deep understanding of the molecular mechanisms behind hormone-induced changes in gene expression, including alternative splicing, can be of significant interest. We found 248 events induced by progesterone stimulation, most of them independent of the effects of the hormone on transcription regulation.

We confirmed that internal exons display, on average, high nucleosome density compared to their flanking introns, but we could identify four different main classes of nucleosome profiles in internal expressed cassette exons, which closely follow the GC distribution in these regions. The most frequent class is characterized by rather small peaks of nucleosome density at exons compared with the flanking introns. Two subgroups of exons show higher nucleosome density in the intronic region upstream or downstream of the exon, while another class displays a strong distinct peak in the exon, which -similarly to what was previously reported- are frequently associated with weaker splice sites and longer flanking introns. Upon progesterone treatment, changes in nucleosome density profiles are observed in both regulated and not regulated exons.

Our results reveal correlations between nucleosome positioning upstream the alternative exons and increased exon inclusion. Contrary to our initial expectation, higher nucleosome density at

exons does not necessarily correlate with higher exon inclusion upon hormone treatment. The dynamics of RNA polymerase II transcription –which are strongly affected by nucleosome density- are known to influence alternative splicing decisions (de la Mata et al., 2003; de la Mata et al., 2010; Dujardin et al, 2014). R5020-included exons undergo more nucleosome changes and seem also to be more influenced by RNA pol II dynamics than R5020-skipped exons. These, in turn, seem less vulnerable to changes in chromatin and Pol II dynamics, as they generally display weaker nucleosome densities and lower Pol II accumulation.

It seems unlikely that every hormone-induced alternative splicing event is influenced by chromatin structure. Expression of a number of splicing factors and regulators is affected by progesterone stimulation, and this could contribute to the observed changes in alternative splicing. It also seems possible that chromatin architecture plays a role in hormone-controlled alternatively spliced exons whose nucleosome profiles change upon hormone stimulation.

Overall the picture is quite complicated, but there are some constant strong behaviors that we could describe. A direct correlation between high nucleosome positioning and RNA Polymerase II dynamics doesn't seem to be sufficient, unlike what predicted, to fully explain the interplay between exon inclusion and nucleosome positioning. Recent findings already overcame the classic view of the kinetic model (de la Mata et al, 2003), showing that slower elongating polymerase can also lead to increase exon skipping, favoring the recognition of a splicing silencer (Dujardin et al, 2014). Similarly, we can think that well positioned nucleosomes over weak exons (cluster D), lead to Pol II accumulation and

expose a number of *cis*-regulatory elements around the alternative exon.

The mechanism through which some nucleosome conformations (e.g. high nucleosome occupancy on upstream exons favoring inclusion) are influencing alternative splicing decision is still quite unclear. One possibility is that nucleosome re-arrangements, altering Pol II dynamics and thus recruitment of splicing factors, influence the timing in which these exons are processed, delaying them for post-transcriptional processing thus changing the range of regulators to which are exposed. Splicing inhibiting drugs impair, to some extent, the binding of splicing factor to histones (Convertini et al., 2014), and lead to changes of local levels of histone modifications (Kim et al., 2011). In depth study employing such drug treatments could help to dissect the mechanism of interplay between chromatin structure and splicing regulation, providing information regarding co-trascriptional recruitment of the spliceosome in time and space (along the gene) and help to deeper understand these still puzzling connections.

Materials and Methods

Cell Culture and Hormone Treatments

T47D-MTVL human breast cancer cells, carrying one stably integrated copy of the luciferase reporter gene driven by the MMTV promoter (Truss et al., 1995), were routinely grown as described (Vicent et al., 2011). Cells were plated in RPMI medium without phenol red, supplemented with 10% dextran-coated charcoal-treated fetal bovine serum (FBS). After 48 hr, the medium was replaced by fresh medium without serum. After 24 hr in serum-free

conditions, cells were incubated with 10 nM R5020 or EtOH for different lengths of time at 37°C.

PR binding site analysis

A Refseq "known genes" was assigned to any R5020-induced exon, to 243 random entries among genes displaying or not transcriptional changes upon R5020 treatment, and to genes not harboring R5020-regulated exons (in order to have the same number of genes containing R5020-regulated exons). Using the table browser, we intersected PR ChIP-Seq tracks (Ballare' et al, 2013) and the tracks of the coordinates of the genes corresponding to each category, including 3000 base pairs 5' of the TSS. The percentage of genes containing PR binding sites was subsequently determined for each category.

Chromatin Immunoprecipitation

RNA Polymerase II-ChIP was carried out as previously described (Paronetto, et al., 2010) with minor modifications of the protocol. T47D-MMTVluc cells (~ 2 x 106 cells/sample) were incubated with 1% (vol/vol) formaldehyde in culture medium for 10 minutes at room temperature. Cells were then washed in cold phosphate-buffered saline (PBS), harvested and lysed to isolate nuclei in a hypotonic buffer containing 5mM Pipes 8.0, 85 mM KCl and NP-40 0.5%. Nuclei were then resuspended, lysed in a buffer containing 1% SDS, 10 mM EDTA and 50 mM Tris/HCl pH 8.1, and sonicated in 15 ml tubes with Bioruptor UCD-200 Diagenode (ultrasonic wave output power 250W, 14 x 30 seconds) to yield chromatin sizes of 250-400 bp, and 100 µg of DNA/sample were used for immunoprecipitation with 2µg of anti-RNA polymerase II clone 4H8 antibodies (Millipore 05-623). Co-precipitated DNA and input DNA

were purified and subjected to deep sequencing using the Solexa Genome Analyzer (Illumina). The sequence reads were aligned to the human genome reference (assembly hg19), keeping only tags that mapped uniquely with up to two mismatches.

To investigate the accumulation of Pol II over exons, we downloaded from the UCSC table browser ChIP-Seq reads mapped to alternative exons, downstream exons, 300 base pairs downstream and 750 base pairs upstream the 3'ss. We then averaged the ChIP-seq signal over exons (alternative or downstream constitutive) in 5 consecutive bins of 150 bp upstream the 3' ss in a window of 750bp. These values were then normalized to local average for every exon and its upstream bins. Two-samples t-test was used to determine the associated p-Values.

Sequence processing

Single-end and paired-end Illumina sequencing data from MNase-digested chromatin was obtained from Ballare' et al (2013) T47D-MTVL breast cancer cells untreated or treated for 60 minutes with R5020. FASTQ files were aligned to the reference human genome GRCh37/hg19(Lander et al, 2001) using BWA (Li et al, 2010). At the time of writing, BWA does not fill all BAM fields completely, so "samtools fixmate" was run to fill in information about a paired-end read's mate. Resulting bam files {Li, 2009 #25564}[4] were then processed in the following way using custom software:

1. If 40% or more of a read's quality values were demarked by a '#' (Illumina 1.8 Phred quality), i.e. unknown quality value at that base, then the read was flagged with the 0x200 bit (the read does not pass quality controls), and tagged with ZL:Z:0.40.

2. If the read overlaps with a highly-duplicated region (HDR) {Pickrell, 2011 #25565}[5], the read is also flagged with the 0x200 bit and tagged with "ZR:Z:HDR".
3. If both the read and its mate's coordinates match one or more pair's coordinates, the read is tagged "ZD:Z:**x**.**y**", where in this case **x** indicates the total number of pairs duplicated, and **y** represents an index for one of the pairs in question, numbered from 1 to **x**.
4. If flags and tags are encountered on a read's mate, then they are added to the read.

The final BAM file is then indexed using "samtools index". Because no reads are ever removed at any stage, the original FASTQ files can be regenerated from the BAMs.

Nucleosome occupancy signal generation

Reads not flagged as unsuitable were gathered, and single-end reads were extended to be 147 bp long. Using custom software, genome-wide read depth profiles were then generated and subsequently normalized at each base using alternative read depth profiles where the reads were extended from their middle bases 10 kilobases upstream and downstream. In this way, the profiles are normalized to their 20 kb background, and are then acceptable for comparing between samples with different sequencing depths. In principle, the normalization also attenuates issues arising from the aneuploidity of the T47D genome.

RNA-sequencing

RNA-Seq Paired-end reads were mapped using GEM RNA Mapping Pipeline (v1.7) using last Gencode Annotation version 18, as follows:

1) Reads were mapped to genome reference index (including chromosome 1 to X) with maximum mismatch ratio of 0.06 and base quality threshold of 26.

2) Reads were mapped to a transcriptome index generated from previous reference index using the same options as step 1.

3) De-novo transcriptome index were created and the reads mapped again to this index using maximum mismatch ratio of 0.04, minimum split size of 15 bp and maximum of 500.000 bp.

4) Reads were mapped again to de-novo transcriptome index created using the same options as step 1.

5) The created mappings and pair alignments were merged using base quality threshold of 26 and maximum insertion size of 500.000 bp.

As a result, BAM alignment files were obtained and used to generate genome-wide normalized profiles using RSeQC software. Exon quantifications (summarized per-genes) were used for expression level determination, either as raw read counts or as reads per kilobase per million mapped reads (RPKM) using Flux Capacitor (http://liorpachter.wordpress.com/tag/flux-capacitor/).

Detecting hormone-dependent alternate splicing exons

Isoform expression and differentially spliced cassette exons in the T=1hr and T=6hr post treatment samples versus the untreated (T=0hr) samples were determined using the MISO pipeline (Katz et al 2010) using the "skipped exon" event type option. Briefly, PSI (“perecent spliced in”) values were estimated for each of the three timepoints using the MISO-provided reference exon-centric annotation (hg19 v1.0). Next a list of high confidence differentially expressed cassette exons for the two comparisons (T=1hr vs T=0hr and T=6hrs vs T=0hr) was compiled by applying the

following filters:

-An absolute difference of PSIs of at least 0.15 (|ΔPSI|>0.15).

-At least 3 reads supporting exon skipping in one of the two conditions.

-At least 10 reads coming from the flanking exons in both conditions.

-A p-value<0.01 for a different mean between the two posterior PSI distributions (two-tailed t-test)

-Only strictly internal cassette exons events are retained.

Clustering and aggregation of nucleosome signal

Skipped exons were extracted from all models in the MISO annotation to define a general set of internal exons (39,922 total). The coordinates of these exons were then projected ("lifted") from genome reference hg18 to hg19 using the liftOver tool from the UCSC Genome Browser website. 16 exons were discarded because did not lift correctly, leaving 39,206. Further exons were removed in the following order:

1. 182 were without ENSEMBL v68 gene IDs. (39,024 total).
2. 45 were on chrY. (38,979 total).
3. 11,748 exons were filtered where the 1 kb region surrounding the 3' splice site of the exon and the 1 kb regions surrounding the 3' splice sites of both flanking exons did not all have mappability above 90%. (27,231 total).

Using bwtool (Pohl et al, 2014), with the -starts option and 500 bases upstream and downstream, clusters of T0 normalized nucleosome occupancy signal were then obtained from the general internal exon set using a k-means clustering parameter of k=4. The averaged values of nucleosome signal at each base in the 1

kb surrounding the 3' splice site were then calculated (using bwtool) for each cluster, first at T0 and then at T60 of the MNase-Seq data.

Splice site scoring

All exons expressed in T47D at T0 were extracted from MISO annotation, as described in the previous section, and scored using "Analizer-Splice-Tool", attributing to each annotated exon the sum of its acceptor and its donor score. Two-samples t-test were used to assign p-Values.

Other data

50 bp mappability and 5 base-window GC percent data were downloaded from the UCSC Genome Browser.

References

Ballare, C., Castellano, G., Gaveglia, L., Althammer, S., Gonzalez-Vallinas, J., Eyras, E., Le Dily, F., Zaurin, R., Soronellas, D., Vicent, G.P., et al. (2013). Nucleosome-driven transcription factor binding and gene regulation. Molecular cell 49, 67-79.

Ballare, C., Zaurin, R., Vicent, G.P., and Beato, M. (2013b). More help than hindrance: nucleosomes aid transcriptional regulation. Nucleus 4, 189-194.

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.

Beato, M. (1996). Chromatin structure and the regulation of gene expression: remodeling at the MMTV promoter. J Mol Med (Berl) 74, 711-724.

Beato, M., and Klug, J. (2000). Steroid hormone receptors: an update. Human reproduction update 6, 225-236.

Beato, M., Herrlich, P., and Schutz, G. (1995). Steroid hormone receptors: many actors in search of a plot. Cell 83, 851-857.

Bird, G., Zorio, D.A., and Bentley, D.L. (2004). RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. Molecular and cellular biology 24, 8963-8969.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., et al. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell 122, 33-43.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nature reviews Molecular cell biology 10, 741-754.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469, 368-373.

Convertini, P., Shen, M., Potter, P.M., Palacios, G., Lagisetti, C., de la Grange, P., Horbinski, C., Fondufe-Mittendorf, Y.N., Webb, T.R., and Stamm, S. (2014). Sudemycin E influences alternative splicing and changes chromatin modifications. Nucleic acids research.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M.a., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. Molecular cell 12, 525-532.

de la Mata, M., Lafaille, C., and Kornblihtt, A.R. (2010). First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. RNA 16, 904-912.

Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., and Carroll, J.S. (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nature genetics 43, 27-33.

John, S., Sabo, P.J., Thurman, R.E., Sung, M.H., Biddie, S.C., Johnson, T.A., Hager, G.L., and Stamatoyannopoulos, J.A. (2011). Chromatin

accessibility pre-determines glucocorticoid receptor binding patterns. Nature genetics 43, 264-268.

Kaplan, N., Moore, I., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E. (2010). Nucleosome sequence preferences influence in vivo nucleosome organization. Nature structural & molecular biology 17, 918-920; author reply 920-912.

Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nature methods 7, 1009-1015.

Khodor, Y.L., Menet, J.S., Tolan, M., and Rosbash, M. (2012). Cotranscriptional splicing efficiency differs dramatically between Drosophila and mouse. RNA 18, 2174-2186.

Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.H., Marr, M.T., 2nd, and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes & development 25, 2502-2512.

Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D.L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. Proceedings of the National Academy of Sciences of the United States of America 108, 13564-13569.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell 128, 693-705.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science 339, 950-953.

Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921. doi:10.1038/35057062.

Lange, C.A. (2004). Making sense of cross-talk between steroid hormone receptors and intracellular signaling pathways: who will have the last word? Mol Endocrinol 18, 269-278.

Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–595. doi:10.1093/bioinformatics/btp698.

Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. Cell 144, 16-26.

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.
Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389, 251-260.

Migliaccio, A., Di Domenico, M., Castoria, G., de Falco, A., Bontempo, P., Nola, E., and Auricchio, F. (1996). Tyrosine kinase/p21ras/MAP-kinase pathway activation by estradiol-receptor complex in MCF-7 cells. The EMBO journal 15, 1292-1300.

Nahkuri, S., Taft, R.J., and Mattick, J.S. (2009). Nucleosomes are preferentially positioned at exons in somatic and sperm cells. Cell Cycle 8, 3420-3424.

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. Nature genetics 40, 1413-1415.

Pan, Q., Shai, O., Misquitta, C., Zhang, W., Saltzman, A.L., Mohammad, N., Babak, T., Siu, H., Hughes, T.R., Morris, Q.D., et al. (2004). Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. Molecular cell 16, 929-941.

Pickrell JK, Gaffney DJ, Gilad Y, Pritchard JK. False positive peaks in ChIP-seq and other sequencing-based functional assays caused by unannotated high copy number regions. Bioinformatics. 2011;27(15):2144–2146. doi:10.1093/bioinformatics/btr354.

Pohl A, Beato M. bwtool: A tool for bigWig files. Bioinformatics. 2014:btu056. doi:10.1093/bioinformatics/btu056.

Saint-Andre, V., Batsche, E., Rachez, C., and Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1gamma favor inclusion of alternative exons. Nature structural & molecular biology 18, 337-344.

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. Proceedings of the National Academy of Sciences of the United States of America 106, 4325-4330.

Schwartz, S., Gal-Mark, N., Kfir, N., Oren, R., Kim, E., and Ast, G. (2009a). Alu exonization events reveal features required for precise recognition of exons by the splicing machinery. PLoS computational biology 5, e1000300.

Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. Nature structural & molecular biology 16, 990-995.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K.,
Wang, J.P., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature 442, 772-778.

Singh, R.K., and Cooper, T.A. (2012). Pre-mRNA splicing in disease and therapeutics. Trends in molecular medicine 18, 472-482.

Smith, C.W., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. Trends in biochemical sciences 25, 381-388.

Spies, N., Nielsen, C.B., Padgett, R.A., and Burge, C.B. (2009). Biased chromatin signatures around polyadenylation sites and exons. Molecular cell 36, 245-254.
Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., and Guigo, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. Genome research 22, 1616-1625.

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcarcel, J., and Guigo, R. (2009). Nucleosome positioning as a

determinant of exon recognition. Nature structural & molecular biology 16, 996-1001.

Vicent, G.P., Ballare, C., Zaurin, R., Saragueta, P., and Beato, M. (2006). Chromatin remodeling and control of cell proliferation by progestins via cross talk of progesterone receptor with the estrogen receptors and kinase signaling pathways. Annals of the New York Academy of Sciences 1089, 59-72.

Vicent, G.P., Zaurin, R., Ballare, C., Nacht, A.S., and Beato, M. (2009a). Erk signaling and chromatin remodeling in MMTV promoter activation by progestins. Nuclear receptor signaling 7, e008.

Vicent, G.P., Zaurin, R., Nacht, A.S., Li, A., Font-Mateu, J., Le Dily, F., Vermeulen, M., Mann, M., and Beato, M. (2009b). Two chromatin remodeling activities cooperate during activation of hormone responsive promoters. PLoS genetics 5, e1000567.

Wang, Z., and Burge, C.B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA

Results-Part II

Manuscript:

Co-occurrence of cell type-specific splicing and cell type- specific chromatin arrangements

Hagen Tilgner[1,3,*], Joao Curado[1,4,*], Camilla Iannone[1,*], Anshul Kundaje[3], Ben Brown[5], Carrie Davis[6], Sudipto Chakrabortty[2], Michael Snyder[3], Thomas Gingeras[6], Juan Valcarcel[1], Roderic Guigó[1,2,+]

1 Centre for Genomic Regulation (CRG) and UPF, Dr. Aiguader, 88, 08003 Barcelona, Catalonia, Spain

2 Departament de Ciencies Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona.

3 Stanford University

4 GABBA Program, Universidade do Porto, Porto, Portugal

5 UC Berkeley

6 Cold Spring Harbor Laboratory

* equal contribution

+ To whom correspondence should be addressed

Summary

Increasing evidence supports a functional link between chromatin density or epigenetic modifications and pre-mRNA splicing. DNA methylation has also been implicated in alternative splicing regulation via the methylation-sensitive binding of the DNA binding protein CTCF. To investigate these relationships genome-wide, we have used ENCODE data from human cell lines to compare high-throughput genome wide data of chromatin features (histone modifications, Pol II and CTCF ChIP-Seq, DHS, ChIA-PET), with levels and isoforms of polyadenylated RNAs specifically monitored in the nucleus by RNASeq and Cap Analysis of Gene Expression (CAGE). Pairwise comparisons of cell lines uncover that three histone modifications (H3K4me3, H3K27ac and H3K9ac), usually marking promoter regions, co-occur in a subset of higher included exons, marking specifically the exon cassette, independently from gene expression regulation or activation of cryptic transcription start sites. These three histone marks are sufficient to predict differential inclusion levels in this subset of exons. Moreover, DNase I hypersensitivity sites, Pol II ChIP-Seq signals and long-range chromatin interaction signatures are higher in exons harboring H3K4me3, H3K9ac and H3K27ac, and correlate well with higher exon inclusion. Finally, we provide data compatible with accumulation of transcription factors in exons under conditions of high exon inclusion, including the SWI/SNF ATPase Brg1. These observations are suggestive for a functional role for 3-dimensional genome structure in the regulation of alternative splicing.

Introduction

Whether an alternative exon is included or excluded in a mature RNA is considered a matter of combinatorial control, involving splice sites, additional binding sites and the factors that recognize them (Smith and Valcárcel, 2000). Recent evidence strongly suggests that chromatin status and transcription dynamics also contribute to this decision. First, splicing can occur co-transcriptionally (Beyer and Osheim, 1988; Kornblihtt et al., 2004), and this has been demonstrated to be widespread in yeast (Carrillo Oesterreich et al., 2010), drosophila (Khodor et al., 2011) and human (Ameur et al., 2011; Tilgner et al., 2012). Second, some splicing factors are also known to interact with modified histone tails, and intragenic histone modifications have been shown to be involved in alternative splicing decisions on individual genes (Luco et al., 2010). Third, RNA Polymerase II elongation dynamics is known to influence exon inclusion (de la Mata et al., 2003; Roberts et al., 1998), which, in this context, can be promoted by CTCF binding (Shukla et al., 2011). And lastly, a number of independent studies showed that chromatin structure correlates with exon-intron structure genome-wide (Andersson et al., 2009; Hon et al., 2009; Kolasinska-Zwierz et al., 2009; Nahkuri et al., 2009; Schwartz et al., 2009; Spies et al., 2009; Tilgner et al., 2009).

While it is now clear that links between chromatin and splicing exist, the details of this relationship are likely to be complex and still not well described. The potential genome-wide role of histone modifications in the regulation of alternative splicing remains therefore to be firmly established. In particular, to what extent cell-

type specific chromatin organization can contribute to cell type-specific splicing patterns.

We use data generated within the ENCODE project (consortium, 2012; Myers et al., 2011) to show that differential levels of a number of histone modifications in the vicinity of exons do indeed correlate with the differential inclusion level of alternative exons in different cell types. We use a statistical framework to identify differentially used exons between cell lines using RNASeq and validate some of these predictions by qPCR. Using ChIPSeq-datasets we compute the chromatin profiles around these exons and identify histone modifications whose levels change significantly when inclusion levels change. By partitioning the alternative exon set we identify exons whose cell-type specific inclusion level is correlated with cell-type specific chromatin arrangements and we investigate the presence of DNAse I hypersensitive sites (DHS), chromatin loops to the promoter, RNA polymerase II kinetics and binding of transcription factors (TF) within the exons and in their neighborhood.

Our data are consistent with distinct contributions of particular chromatin marks to alternative exon regulation.

Results

R1. Alternatively included exons in pair-wise cell type comparisons

We identified sets of internal exons that are differentially included when comparing the transcriptomes of two different cell lines (Figure 1a). We applied a computational method similar to that published by Wang and co-workers (Wang et al., 2008; Methods, Figure S1) to nuclear polyA+ RNASeq samples from five cell lines (K562, Gm12878, Hepg2, Huvec, Helas3) (Djebali et al., 2012). RNA in this cellular fraction has mostly completed splicing (Tilgner et al., 2012), but it is unlikely to have gone through NMD and thus likely presents a more direct view of splicing outcomes than whole-cell or cytoplasmic fractions, which may biased by retention of some RNA-molecules in the nucleus and are effected also by the rates of cytoplasmic RNA decay. The ENCODE tier-1 cell lines K562 and Gm12878 (the first a myelogenous leukemia and the second a lymphoblastoid cell line) were used first, and a total of 1688 exons were initially detected as differentially alternatively skipped.

To obtain a highly reliable set of alternative exons that can be compared to chromatin structure, we applied additional filtering (see Methods). We excluded exons from genes having large expression differences between cell lines, as measured by CAGE data (see Methods), because gene expression changes are clearly linked to chromatin re-organization (Barski et al., 2007). Finally, we obtained a total of 510 exons that show significantly different inclusion in Gm12878 as compared to K562 (283 more highly included in Gm12878 - referred to as “more included exons” - and

227 more highly included in K562 – referred to as “less included exons”) (Figure 1b).

These differentially spliced exons had characteristic properties of alternative exons, as previously described in the literature (Magen and Ast, 2005; Zheng et al., 2005): their splice sites are associated with lower splice site strength (Figure 1c), the exons tend to be shorter (Figure 1d) and, for coding exons, their length is more often divisible by three than for exons not selected as alternatively spliced (Figure 1e). Moreover, we did not find significant differences in the levels of expression of the genes with alternatively included exons between K562 and Gm12878 (Figure 1f). In order to independently validate these alternative exon inclusion calls, we selected a total of 12 of these exons for validation (Figure 1g). Using exon-junction oligonucleotides, we quantified by qPCR the ratio of inclusion/skipping isoform in one cell line compared to the other and we obtained a validation rate of 92% (11 out of 12 selected cases, Fig 1g). This method of quantification provides an assessment of differences in the relative inclusion of the exons regardless of possible differences in gene expression between the cell lines.

Given the high validation rate for the K562 versus Gm12878 comparison, we repeated the same analysis for the remaining 14 pair-wise cell type comparisons. In order to avoid biases, we restricted our further analysis to the seven pairwise cell type comparisons in which exons detected as differentially spliced passed all the above criteria.

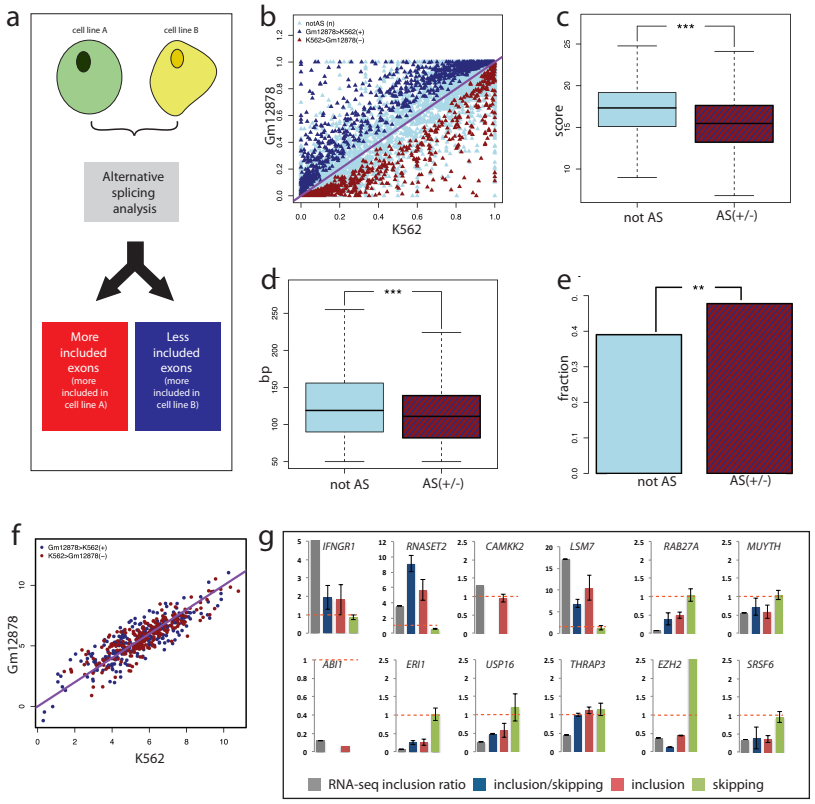


Figure 1: **calling AS exons (K562 vs. Gm12878)**. **(a)** schematic representation of classification of "more included" and "lower included" exons **(b)** Estimated inclusion ratio in the K562 cell line (x-axis) and in the Gm12878 cell line (y-axis) of exons whose inclusion is (i) significantly higher in Gm12878 (dark blue), (ii) significantly higher in the K562 cell line (dark red), (iii) whose inclusion does not change significantly between the two cell lines (light blue). (**c**) Comparison of combined splice site strength between exons without significant inclusion changes (left boxplot) and differentially included exons (right boxplot). (**d**) Comparison of the same exon sets as in figure 1c, but for exon length. (**e**) Assessment of divisibility by 3: Only exons that were entirely coding were considered. Furthermore for differentially included exons we considered only those for which no other exon in the same gene was differentially included. (**f**) Dot-plot of CAGE gene expression for genes of exons, whose inclusion is (i) significantly higher in Gm12878 (dark blue), (ii) significantly higher in the K562 cell line (dark red). X-axis: Log2(cage RPM) in K562; Y-axis Log2(cage RPM) in Gm12878. This shows that there is no significant difference in differential gene expression between the two exon sets. **(g)** Experimental validation: RT-qPCR analyses of RNA extracted from K562 and Gm12878. Histograms represent: ratio of inclusion level obtained for

each exon between Gm12878 and K562 through the analysis of the RNA-seq data (ENCODE) (grey); average of three independent experiments of the ratio between Gm12878 and K562 of qPCR amplification signals for exon junctions of inclusion (red), skipping (green) or ratio of inclusion/skipping (blue), normalized to the ratios corresponding to a constitutive exon. Error bars represent standard deviations of three independent experiments. 11 out of 12 exons tested show the same tendency of RNA-seq of being more included in Gm12878 or K562 (*THRAP3* is not validated). Skipping junction primers didn't work for ABI1 and CAMKK2, which have been validated also employing primers specific for the alternative exons (data not shown).

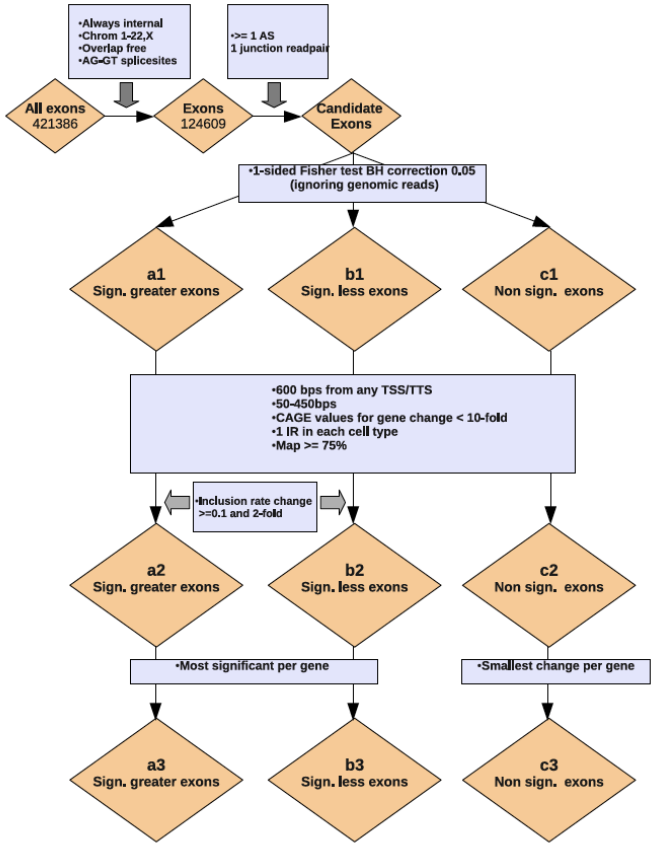


Figure S1: Flowchart of the pipeline employed to define differentially skipped exons between two RNAseq data sets. Here RNAseq datasets originate from cell types investigated within the ENCODE project. See text for details.

R2. Co-occurrence of differences in histone modifications with alternative exon inclusion

For the 7 “valid” cell pairs, we pooled separately all “more included exons” and “less included exons”, and compared the chromatin marks between both groups. From normalized chromatin-signal tracks we compared the levels of 9 histone marks, the insulator protein CTCF and input DNA on more and less exons (consortium, 2012; Hoffman, 2012; Kundaje, 2011). The “differential signal” for any chromatin feature was defined as the difference of the average signal over the exon in the 2nd cell type (e.g. Gm12878 for the comparison K562 vs Gm12878) and the average signal over the exon in the first cell type (e.g. K562 for the same comparison). This differential exonic signal was computed for all exons and in all seven pairwise cell comparisons separately, and then pooled together (see Methods) to produce a single composite comparison for each monitored histone modification.

Our analysis shows that CTCF, H3K27ac, H3K4me3 and H3K9ac have a positive correlation with exon inclusion (pvalue < 0.01; wilcoxon signed-rank test with bonferroni correction) (Figure 2a). Surprisingly, the input control experiment, which is sonicated DNA, did not behave entirely neutrally but shows a negative correlation with exon inclusion – that is, it shows the opposite behavior of the above ChIP-Seq datasets. Since input samples are cross-linked and sonicated, these differences might reflect measure of chromatin compaction state, consistent with a general correlation between open chromatin and lower exon inclusion. Therefore we didn’t use this signal to normalize the other marks.

We investigated if the histone modifications behavior is local and specific to the studied exons. Thus, we selected, for each differentially included alternative exon, the closest upstream and

the closest downstream exon that our method did not identify as alternative in the particular pairwise cell-type comparison under study (Methods). This approach creates a “constitutive-alternative-constitutive” exon triplet. We calculated differential chromatin levels in 800bp windows from the exon-center for all exons in both types of triplets (higher inclusion in cell type 1 and higher inclusion in cell type 2).

The significant difference for the 4 ChIP-Seq datasets between “more:” and “less” exons did not extend to the flanking exons (Figures 2b-e). We conclude that a significant positive association can be detected between exon inclusion and the local levels of H3K9ac, H3K27ac, H3K4me3 and CTCF binding.

For H3K9ac, this observation did not fit with previously proposed models linking closed chromatin and slower transcription elongation to exon inclusion (Alló et al., 2009; de la Mata et al., 2003; Saint-André et al., 2011; Schor et al., 2009b), because acetylation is typically associated with open chromatin and consequently faster RNA pol II elongation rates. We performed ChIP-qPCR with antibodies specific for H3K9ac and primers specific for the target exons and one of the closest constitutive exons to validate the differences between K562 and Gm12878 (See Methods). We used an antibody against total H3 for normalization, which also serves as a control to exclude that the effects observed are caused by differences in nucleosome occupancy. In all four alternative exons used we could detect a clear difference in H3K9ac abundance between the two cell lines – a difference positively correlated with differential exon usage. Flanking constitutive exons showed in general smaller (sometimes even opposite) differences and higher variability among replicates

(Figure 2f), confirming that the effect of H3K9ac was local and specific for the alternative exon.

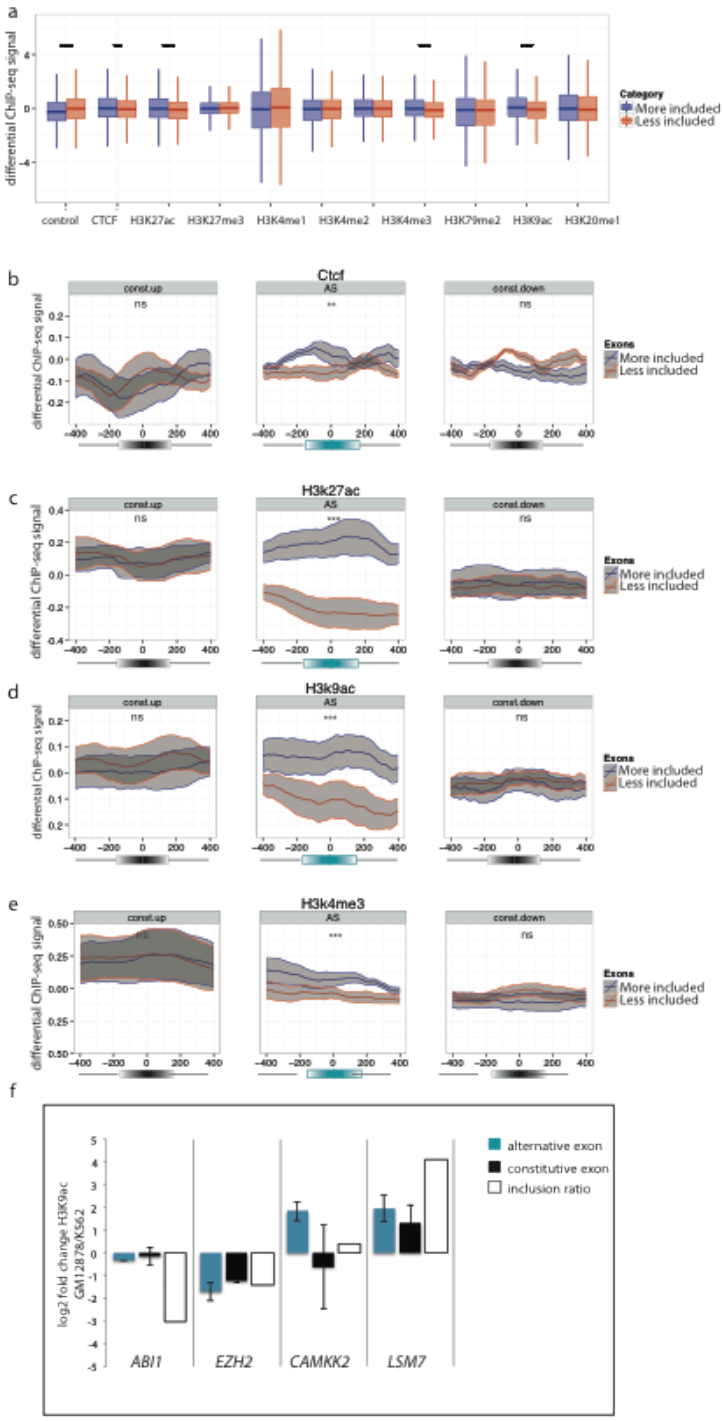

a
differential ChIP-seq signal
Category
More included
Less included
control
CTCF
H3K27ac
H3K27me3
H3K4me1
H3K4me2
H3K4me3
H3K79me2
H3K9ac
H3K20me1
b
Ctcf
const.up
AS
const.down
ns
**
Exons
More included
Less included
c
H3k27ac
***
d
H3k9ac
e
H3k4me3
f
log2 fold change H3K9ac GM12878/K562
alternative exon
constitutive exon
inclusion ratio
ABI1
EZH2
CAMKK2
LSM7

Figure 2: focused chromatin investigation on alternative exons. (a) We pooled alternative "more" (blue) or "less" (red) exons from seven cell type comparisons, and depict the difference of the log2 of differential signal (Y-axis) from 11 ChIP-Seq datasets on these exon. The boxplots represent the average differential signal over the alternative exon. We calculated Wilcoxon rank-sum-tests for the two distributions, significance levels are indicated using * (0.05>p>0.01), ** (0.01>p>0.001), *** (0.001>p); ns=non significant. CTCF, H3K9ac, H3K4me3 and H3K27ac have significantly higher signal in "more" compared to "lower" exons. The control shows the opposite behavior. **(b-e)** We pooled constitutive-alternative-constitutive triplets of "more included exons" (blue) and "less included exons" (red), from our seven cell type comparisons and depict the average differential signal of ChIP seq of CTCF, H3K9ac, H3K4me3 and H3K27ac in one cell line minus the other for each comparison, as well as its standard error of the mean (SEM), in a 800bp-window around the acceptor site of the alternative exon (middle) and around the acceptor sites of the control non alternative upstream (left) and downstream (right) exons. As one can observe, only alternative exons show difference in accumulation of these marks. **(f)** H3K9ac levels quantified by Chromatin Immunoprecipitations performed in Gm12878 and K562 using antibodies against H3K9ac and total H3. Histograms represent the average and standard deviations of the log2 ratio between Gm12878 and K562 of the fraction of input detected in the H3k9ac precipitates of three independent experiments, calculated by Real Time PCR and normalized by the same ratio obtained for H3 precipitates for each exon. Blue bars represent alternative exons (same analyzed as in figure 1g and 1h), black bars are representative of one of the flanking constitutive exons; white bars represent the ratio between inclusion levels in RNA-seq (log2 of grey bars in figure 1f). Error bars represent standard deviations of three independent experiments. We observed a general correlation between exon inclusion and higher levels of H3K9 acetylation for the tested alternative exons. Abi1, and Ezh2, are more included in K562 than in Gm12878, while Camkk2, and Lsm7, are more included in K562 than in GM12878. The correlation could not be extended for two cases to the flanking constitutive exons, which showed in general smaller changes and higher variability among replicates.

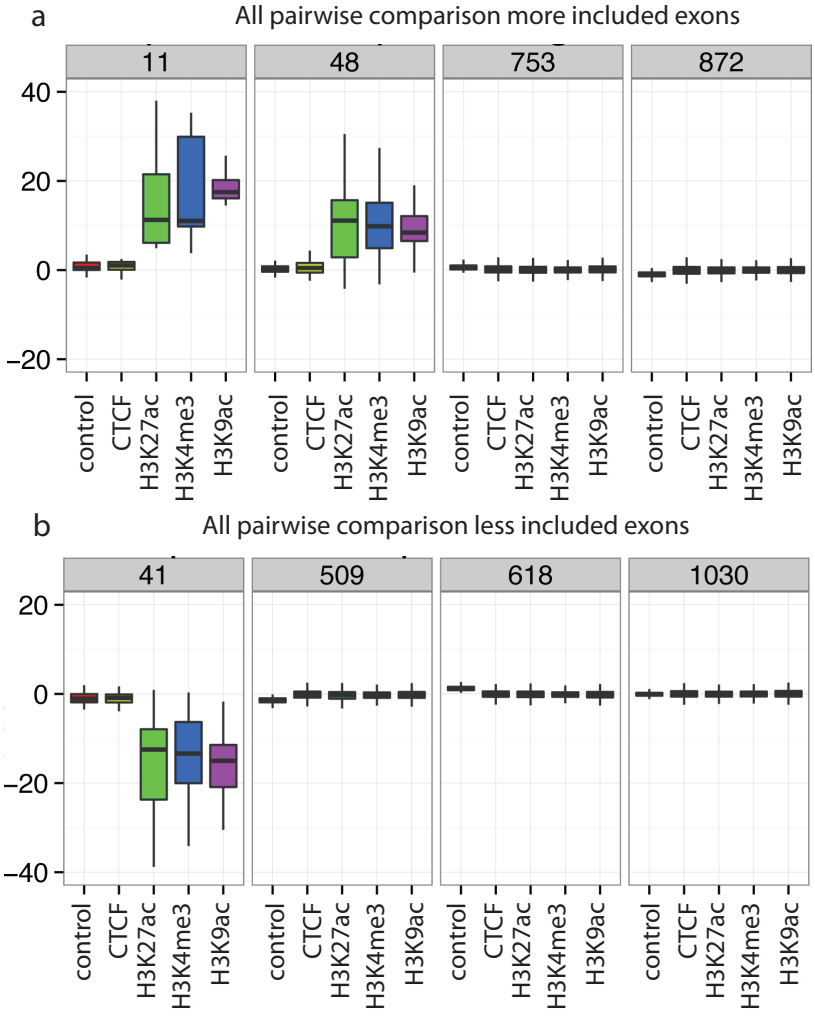
R3. Promoter-like histone marks and exon inclusion

It is well established that histone modifications can function in combinatorial fashion. To address this issue, we investigated the possible interplay between histone marks and splicing using a clustering approach. The 5 signal tracks that had a significant difference between “more” and “less” exons, were classified using k-means clustering analysis. The data was portioned in 4 clusters of different signal intensity, and each exon was assigned to the cluster with the nearest mean (Figure 3a,b). Since the definition of “more” and “less” exons is in this case arbitrary, as expected we found that both categories generate similar but mirrored clusters. The majority of differentially spliced exons do not display differences in the levels of the monitored histone modification compared to the Control ChIP. These exons are probably regulated by other factors, including classical proteins of the SR and hnRNP families whose expression or activity differs between cell lines. We observed instances of co-occurrence of changes of H3K9ac, K27ac and K4me3 in the same direction. These included 59 exons from the more group and 41 exons from the less-regulated group. These 100 exons were merged and classified as “higher” or “lower” depending on the frequency of exon inclusion (Figure 3c).

To better understand the nature of the correlation between H3K9ac, K27ac and K4me3 and exon inclusion, we first confirmed that “higher” exons have, on average, significantly higher inclusion levels compared to “lower” exons (Figure 3d). By analyzing the average distance from the TSS, we found that clustered exons are closer to TSS when compared to non-clustered exons (Figure 3e), an observation likely related to most of the clustered exons being

second exons in the transcript, with their upstream exon being close to the the TSS. H3K4me3, H3K9ac and H3K27ac are known hallmarks of promoters and their levels are associated with transcription activation. To exclude the possibility that those chromatin signatures are marking novel TSS overlapping the clustered exons, we looked for CAGE signals overlapping these exons. Even though “higher” exons have significantly more CAGE signal than “lower” exons, the signal coming from these sequences is almost non-existent when compared to the signal coming from the annotated TSS (Figure 3f). Moreover, there is no significant difference between gene expression levels in transcripts containing “higher” or “lower” exons (Figure 3f). This indicates that differences in chromatin accumulation cannot be explained as a consequence of transcription regulation or usage of novel non-annotated TSS.

In order to cross-validate our observations, we use differences in H3K9ac, K27ac and K4me3 levels to predict exon inclusion changes in a cell line that had not been previously used for our analysis. We measured the histone modification levels in NHEK for 21 exons obtained from pairwise comparisons with K562 (data from ENCODE consortium), and calculated the total difference of H3K9ac, K27ac and K4me3 levels in NHEK versus K562 to predict if the exon inclusion was higher or lower. We could predict exon inclusion differences with an accuracy of 75% for higher and 55% for lower included exons. Focusing on exons with inclusion changes larger than 0.1, we increased our correct predictions to 100% for higher and 75% for lower exons (Figure 3g). These results argue that co-occurrence of three chromatin marks plays a role in defining the levels of inclusion of an exon in a statistically predictable fashion.


a
All pairwise comparison more included exons
11
48
753
872
40
20
0
−20
control
CTCF
H3K27ac
H3K4me3
H3K9ac
b
All pairwise comparison less included exons
41
509
618
1030
20
0
−20
−40

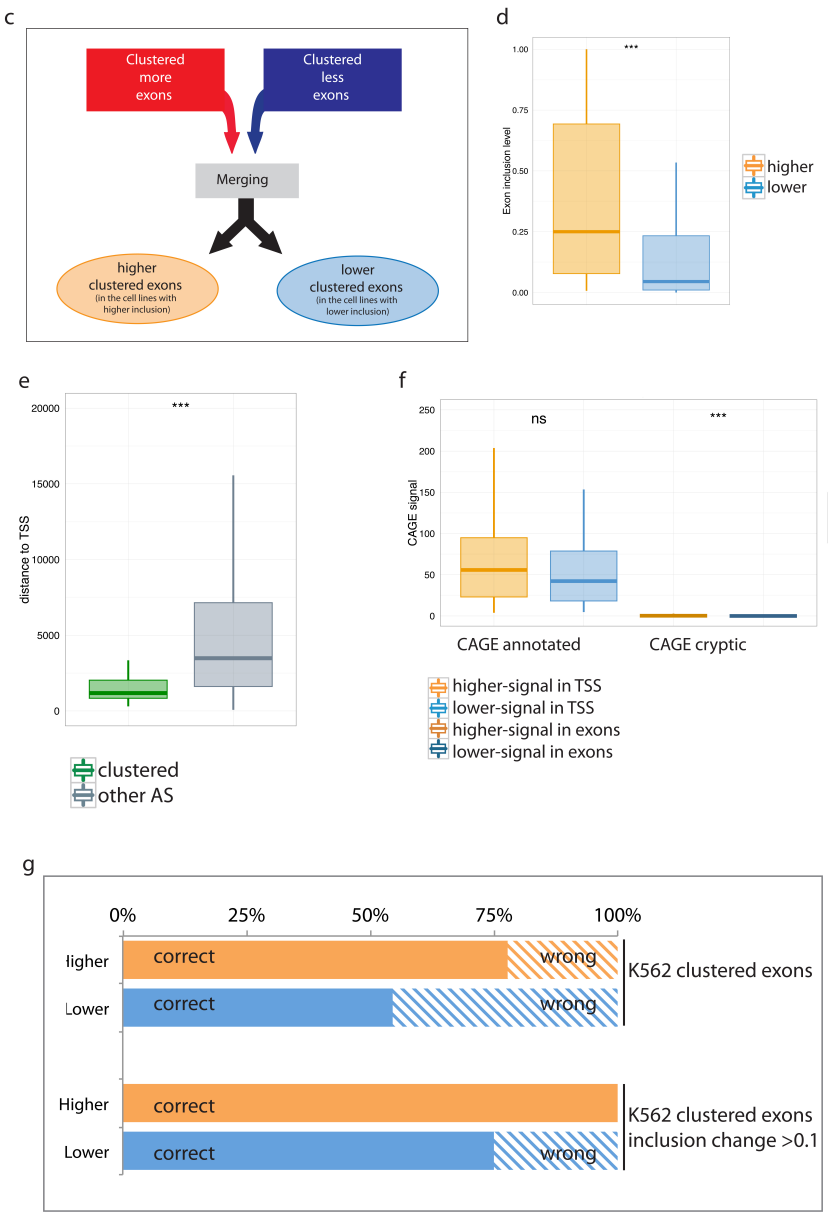


Fig3: identification and characterization of exons showing co-occurrence H3K4me3, H3K27ac and H3K9ac and higher inclusion. (a, b) boxplot represent differential ChIP-seq value for (a) more and (b) less exons. Each square represents one cluster of exons identified through K-means clustering (see text), classified according correlation of histone modification occurrence and splicing behavior. Number of exons present in each cluster is indicated above each square **(c)** Schematic representation of how exons are then defined as “higher” and “lower” included. Briefly, exons in the first two clusters of (a) and in the first of (b)

were merged, and then classified as “higher” in the cell line of the comparison in which they were more frequently included, and “lower” in the cell line of the comparison in which they were less frequently included. **(d)** Boxplots represent inclusion levels of higher (yellow) and lower (blue) exons. **(e)** Boxplots represent distance from the TSS of exons in the clusters mentioned above (green) and of all the other alternative exons identified (figure 1). **(f)** Boxplots represent CAGE signal of TSS of higher exons or higher exons themselves (yellow) or TSS of lower exons or lower exons themselves. **(g)** Number of correctly/wrongly predicted inclusion differences in K562 clustered exons in K562 versus NHEK cells.
Significance levels are indicated using * (0.05>p>0.01), ** (0.01>p>0.001), *** (0.001>p); ns=non significant

We further analyzed chromatin features on these exons and found that “higher” exons are enriched in DNAse I hypersensitive sites (DHS) compared to “lower” clustered exons. DHS are indicative of open, more accessible chromatin; they usually mark *cis*-regulatory elements, including promoter and enhancer sequences. Because many of our “clustered” exons are close to TSS (Figure 3e), we found that upstream exons display higher DHS signal compared to alternative or downstream exons, without significant differences between “higher” and “lower” flanking constitutive exons (Figure 4a).

RNA polymerase II kinetics can influence alternative splicing regulation and slow elongation rates typically favor inclusion of weak exons (de la Mata et al., 2010). Analysis of ChIP-Seq data obtained with an antibody that recognizes RNA Polymerase II show a clear enrichment of RNA Pol II signal in “higher” compared to “lower” exons (Figure 4d). As for DHS, upstream exons display a higher signal of RNA Pol II compared to downstream and alternative exons, an observation possibly related to stalling of RNA Pol II at the promoter itself.

Classic ChIP experiments involve crosslinking that can lead to immunoprecipitation of genomic sequences not directly bound by the protein recognized by the antibody, but rather by higher order complexes that bring the protein in the neighborhood of other sequences. It was indeed recently shown that some internal exons loop and physically interact with promoter and enhancers, and for this reason display “promoter-like” or “enhancer-like” chromatin marks (Mercer et al., 2013). We analyzed ChiA-pet sequencing data (a technique developed to identify chromatin loops and resolve 3-dimensional genome organization-Fullwood et al., 2009), and found an enriched signal (indicative of an increase in looping) in “higher” compared to “lower” exons (Figure 4b). As a control, ChIA-PET signatures in all the alternatively spliced exons identified in our analysis did not show differences between “higher” and “lower” included exons (Figure S2). Our results suggest that exons organized in a 3-dimensional loop with promoters display increased inclusion levels.

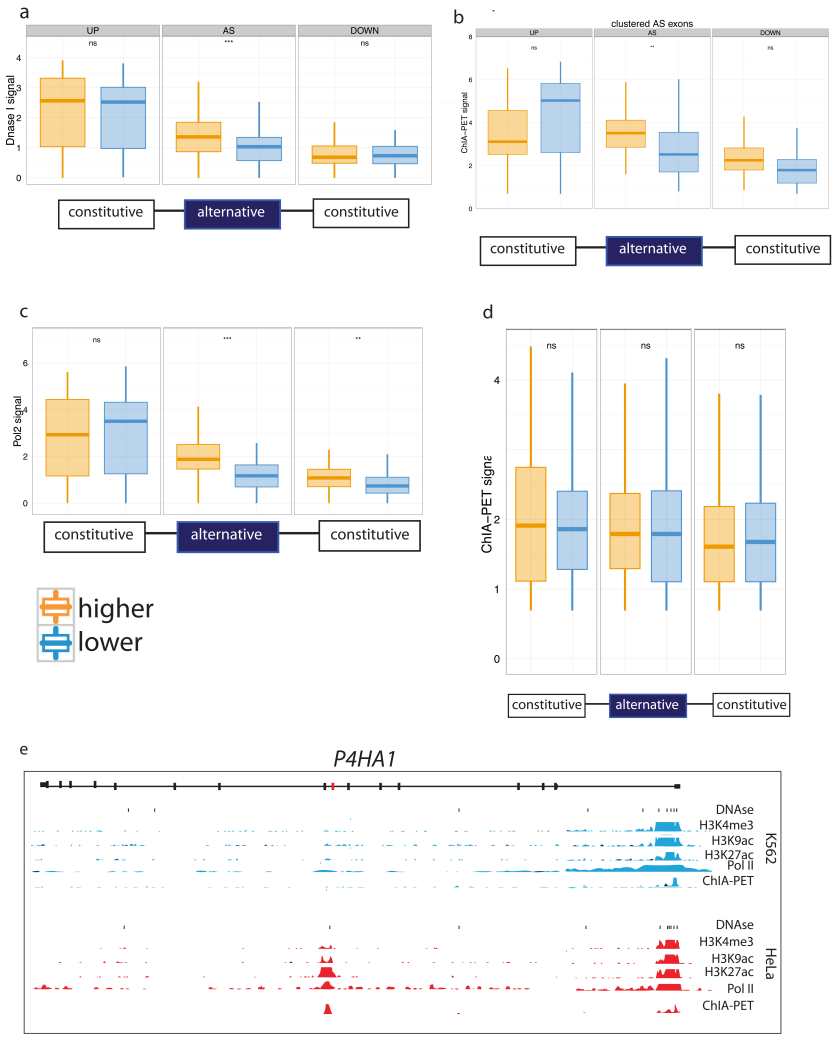


Figure 4
Boxplots represent **(a)** DNAseI signal, **(b)** ChIA-PET signal and **(c)** Pol II signal in higher (yellow) or lower (blue) exons, and in their upstream and downstream constitutive exons. **(d)** Briefly, exons in the last two clusters of 3a and in the three last of 3b were merged, and then classified as "higher" in the cell line of the comparison in which they were more frequently included, and "lower" in the cell line of the comparison in which they were less frequently included. Boxplots represent ChIA-PET signal in higher (yellow) or lower (blue) exons, and in their upstream and downstream constitutive exons. **(e)** A representation of UCSC genome browser tracks of DHS, ChIA-PET, ChIP-seq of Pol II, H3K9ac, H3K27ac and H3K4me3 in K562 and HeLa. In red it is marked the exon differentially regulated between the two cell lines, which in concomitance with enrichment of these chromatin features is higher included in HeLa.

Significance levels are indicated using * (0.05>p>0.01), ** (0.01>p>0.001), *** (0.001>p); ns=non significant.

Roles for transcription factors and chromatin-remodeling complexes on splicing have been documented (reviewed by Saint-Andre et al., 2011; Luco et al., 2011; Kornblihtt et al., 2013; Iannone and Valcarcel, 2013). Their binding is generally enhanced in DHS (Ballare et al., 2013). We thus analyzed Transcription Factors ChIP-Seq data available in ENCODE and found that 10 out of the 30 TF analyzed display significantly more accumulation on higher versus lower exons, while none displays the opposite trend (Figure 5a). Among the transcription factors enriched in "higher" exons we found Brg1 (Figure 5b), which together with Brm is one of the two ATPase of the chromatin remodeling complex SWI/SNF (Wang, 2003). Both Brg1 and Brm co-purify with the nuclear co-repressor complex N-coR, which in turns interacts with the splicing factors SAP130 and SF3a120 (Underhill et al., 2000). While it has already been shown that Brm plays an active role in alternative splicing regulation (Tyagi et al., 2009; Batsche et al., 2006), it is not clear whether Brg1 is also actively modulating pre-mRNA processing.

We depleted Brg1 from HeLa cells by RNAi and tested if this resulted in decreased inclusion levels of "clustered" exons. In parallel, we knocked down Brm. One of the exons tested was a negative control; out of the others, two show decreased exon inclusion upon Brg1 depletion (Figure 5e), but only for one of them, exon 3 of *SRSF6,* the phenomena seem Brg1-specific (Figure 5c,e). Interestingly, the inclusion of this alternative exon on *SRSF6* introduces a stop codon, thus potentially influencing the levels of

the protein. For *USP16*, a gene without Brg1 Chip signal, no effects on splicing upon Brg-1 knock down were observed (Figure 5d,e).
We conclude that for alternative exons whose inclusion correlates with epigenetic marks, correlations can be found between their association with components of chromatin remodeling complexes and changes in alternative splicing upon knock down of these factors.

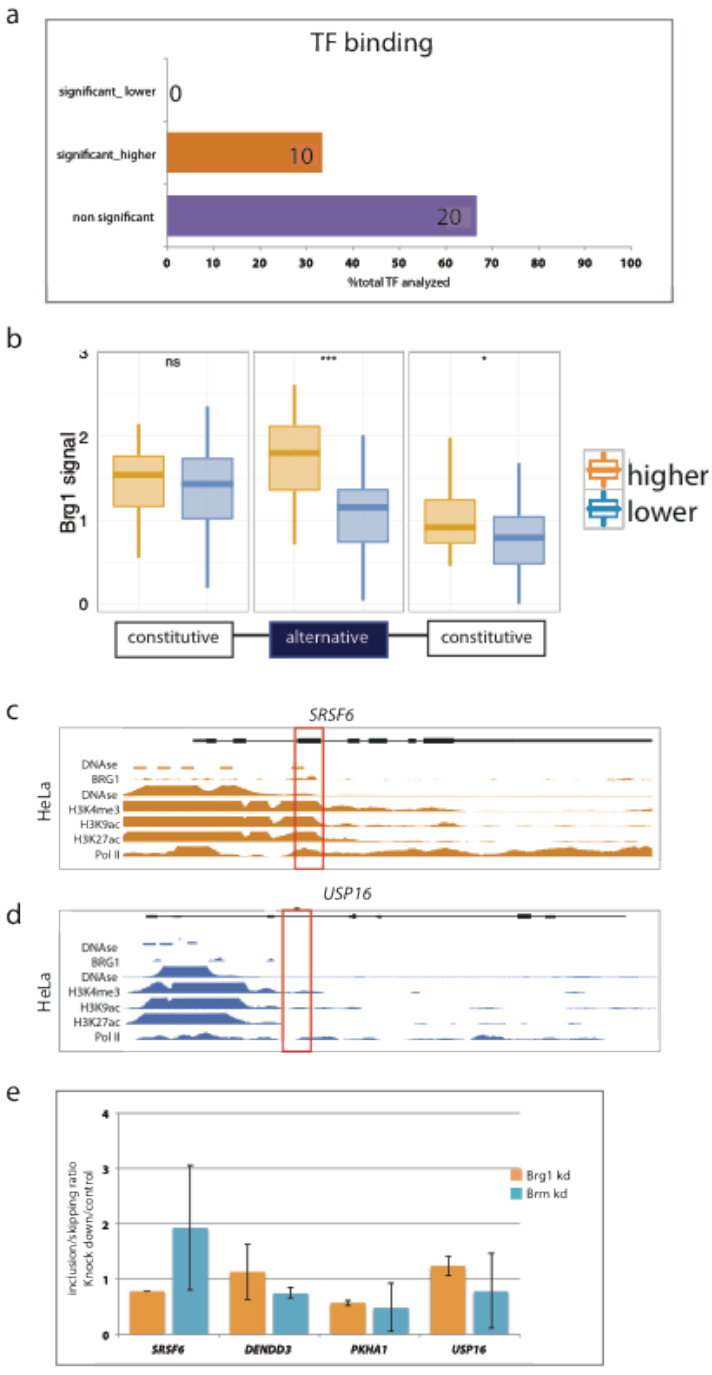

a
TF binding
significant_ lower
0
significant_higher
10
non significant
20
0 10 20 30 40 50 60 70 80 90 100
%total TF analyzed
b
Brg1 signal
ns
***
*
higher
lower
constitutive
alternative
constitutive
c
SRSF6
HeLa
DNAse
BRG1
DNAse
H3K4me3
H3K9ac
H3K27ac
Pol II
d
USP16
HeLa
DNAse
BRG1
DNAse
H3K4me3
H3K9ac
H3K27ac
Pol II
e
inclusion/skipping ratio
Knock down/control
Brg1 kd
Brm kd
SRSF6
DENDD3
PKHA1
USP16

Figure 5
(a) ChIP-Seq signal density of 30 transcription factors (TF) was analyzed over “higher” and “lower” exons. Histograms represent % of TF showing non-significant enrichment between the two categories (violet), significantly enriched in “higher” exons (orange). No TF was found to be significantly enriched in lower exons. (Significant p>0.05) **(b)** Boxplot represent BRG1 ChIP-seq signal in higher (yellow) or lower (blue) exons, and in their upstream and downstream constitutive exons. Significance levels are indicated using * (0.05>p>0.01), ** (0.01>p>0.001), *** (0.001>p); ns=non significant **(c,d)** representation of the UCSC genome browser tracks of DNAse (peaks and signal), BRG1, ChIA-PET, and ChIP-seq of Pol II, H3K9ac, H3K27ac and H3K4me3 in HeLa. Red boxes surround alternatively spiced exons. *SRSF6* alternative exon is “higher” included in HeLa versus K562, while *USP16* has the opposite trend. Brg1 enrichment is only present in *SRSF6* **(e)** q-PCR analysis of exon inclusion regulation upon Brg1 (yellow bars) and Brm (blue bars) knock-down (KD). Histograms represent ratio of inclusion/skipping isoform in KD versus control cells. Error bars represent standard deviations of three independent experiments.

Discussion

Recent results have unveiled that pre-mRNA splicing occurs predominantly co-transcriptionally, thus providing a framework in which chromatin and transcription-related factors interact with pre-mRNA processing machinery. In this work we expand the functional connections between chromatin structure and pre-mRNA processing, establishing associations between epigenetic marks and differential exon inclusion and postulating a role for 3-dimensional genome structure in the regulation of alternative splicing.

To investigate genome-wide connections between splicing and chromatin, we combined polyA+ nuclear RNA-Seq data with high throughput data describing different chromatin features (histone modifications, RNA Polymerase II, CTCF, DHS and ChIA-PET) across five cell lines. We find that H3K4me3, H3K9ac, H3K27ac and CTCF accumulation positively correlate with higher exon inclusion, and further characterized a subgroup of exons for which the co-occurrence of these histone marks strongly correlates with – as is predictive of- higher exon usage. We also provide results that suggest that the organization of these sequences in genomic loops influence splicing regulation. Finally, we provide data compatible with accumulation of transcription factors in exons under conditions of high exon inclusion, including the SWI/SNF ATPase Brg1, with potential consequences on the levels of the encoded proteins like SRSF6, itself a regulator of alternative splicing.

Nuclear polyadenylated RNA is an appropriate choice to study aspects of co-transcriptional RNA processing because it captures recently synthesized-and significantly spliced (Tilgner et al., 2012),

transcripts closer in time and space to the DNA template, allowing also the control for gene expression changes. The strict filtering and high validation rate shows that our method of calling alternative exon cassettes is very efficient.

Several previous studies revealed links between accumulation of specific histone marks and splicing outcome documented (reviewed by Saint-Andre et al., 2011; Luco et al., 2011; Kornblihtt et al., 2013; Iannone and Valcarcel, 2013). Histone modification can influence splicing outcomes directly, through recruitment of splicing factors, or indirectly, through modulation of RNA Pol II dynamics. For example, higher levels of H3K36me3 or of H3K27me3 along the *FGFR2* gene modulate the recruitment of a splicing factor and correlate with the regulation of a mutually exclusive alternative splicing event in a cell specific fashion (Luco et al., 2010). High H3K9ac in the *NCAM* gene, instead, result in more open chromatin and fast elongating RNA Pol II, which leads to skipping of a specific exon (Schor et al., 2009a). In the case of NCAM or FGFR2, changes in histone modifications affect the whole gene. Here, analyzing triplets of “constitutive-alternative-constitutive” exons, we see that positive correlation of H3K9ac, H3K27ac, H3K4me3 and CTCF with inclusion is local and limited to the alternative exon cassette. We identify a subset of 100 cases for which the relationship between inclusion and co-occurrence of these histone modifications is consistent and strong across cell lines. We also see that, for these exons, these three variables alone are enough to predict splicing changes between cell lines, including cell lines that were not part of the training set.

Recent work exploiting the three-dimensional structure of the genome reported physical association of internal exons with associated promoter or enhancers (Mercer et al., 2013). Notably, for a subset of exons, we also report a significant enrichment in DHS and ChIA-PET signals in higher included compared to lower included exons, suggesting the existence of long-range interactions between the exonic genomic regions and promoter regions. These results argue for an interplay between higher order of chromatin structure and alternative splicing regulation and warrant the systematic analysis of these associations in future studies using conformation capture technologies (Dekker et al., 2013).

The correlation between exon inclusion and accumulation of RNA Pol II found in highly included exons can be partially explained by the classic kinetic model (de la Mata et al., 2003). However, both our data on the accumulation of chromatin marks and on the presence of DNase hypersensitive sites suggest that the subset of higher included exons is in an open chromatin context that would typically lead to faster RNA Pol II elongation rates that would typically favor exon skipping rather than inclusion. We propose two, not mutually exclusive scenarios to explain this apparent discrepancy. First, if exons in higher included conditions are engaged in a loop with far away sequences (promoters), the ChIP-Seq signal we see in the exons could actually originate from the high RNA Pol II levels at promoters. Second, some other factors, facilitated by the opening of chromatin, could bind on these exonic regions and slow down the elongation rate of Pol II, a mechanism resembling how CTCF promotes exon inclusion (Shukla et al., 2011). The accumulation of transcription factors in higher inclusion exons could in fact contribute to the two scenarios.

Our results provide new information about the interplay of chromatin structure and pre-mRNA splicing regulation, not only establishing novel correlations between the accumulation of specific combinations of histone modifications and exon inclusion, but also arguing for a role of 3-dimentional organization of the genome in pre-mRNA splicing. Further work will be needed to validate these findings and work out their underlying molecular mechanisms.

Material and Methods

Inclusion level calculation

IDR (Irreproducible Discovery Rate) is a measure widely used within the ENCODE project to assess reproducibility between replicates(Li et al., 2011).

Splice site strength measure

For each exon we used maxEnt (Yeo and Burge, 2004) in order to calculate an acceptor score and a donor score and represented the "exon strength" by the sum of these two scores.

Alternatively skipped exon calling

Using the gencode V15 annotation we determined all exons that are

1. internal in all transcripts they appeared in, in both gene annotations

2. not overlapped by any non-identical exon in both annotations. Identity of exons is defined by their location (chromosome, start, end strand)
3. between 50 and 450bps long
4. surrounded by AG-GT splice sites
5. located on chromosomes 1-22 and X

Then for a given pairwise cell type comparison (here we use K562 and Gm12878 as an example) only exons with minimal AS evidence were retained, that it (JIR_K562(ex)>=1 AND JER_Gm12878>=1) OR (JER_K562(ex)>=1 AND JIR_Gm12878>=1).

For the remaining exons a two by two table was constructed containing junction inclusion reads and junction exclusion reads in the two cell types. Two one-sided fisher tests were run and corrected for multiple testing for the N tests in the Benjamini-Hochberg sense, resulting in three disjoint sets of exons:

1. exons that are significantly more included in Gm12878 (which will be referred to as “more included exons”, even though the choice of the direction from K562 to Gm12878 is clearly arbitrary)
2. exons that are significantly less included in Gm12878 (which will be referred to as “less included exons”)
3. exons whose inclusion is not significantly changed between the two cell types (which will be referred to as “non-AS exons” for the sake of conciseness and clarity although “non-significant AS exons” would be more correct.)

We then chose a subset of the more and less included exons for comparison with chromatin changes meeting the following criteria:

i. at least 600nts away from the respective annotated TSS or TTS

ii. CAGE values of the associated gene did not change more than 10fold between the two cell types
iii. JIR(ex)>=1 or EIR(ex)>=1
iv. at least 75% of all positions in a 900bp window around the acceptor were uniquely mappable for 36mers.
v. inclusion changed by at least 0.1 or two-fold.

Genes frequently contained more than one alternatively spliced exon thus defined. In order to avoid gene specific characteristics that might be introduced into these sets by genes that contribute many alternative exons (as for example the TTN gene where 212 exons passed the fisher test), we chose only one more and up to one less included exon per gene: The exon with the lowest p-value among all exons for the gene in question.
For non-AS exons a similar procedure was carried out, removing however the “inclusion changed by at least 0.1 or two-fold”-criterion and choosing the exon per gene whose estimated inclusion change was minimal among all non-AS exons of that gene (instead of the exon with the smallest p-value). Figure S1 illustrates this approach.

Exon triplets

For each alternative exon and each cell type comparison, we defined two non-alternative exons: The closest up- and downstream exons that

- appeared in a transcript together with the alternative exon
- that showed Benjamini-Hochberg-corrected p-values of 0.05 or more.

Gene expression calculation

We employed CAGE-clusters filtered by a Hidden Markov Model algorithm to differentiate between 5' capped termini of PolII transcripts and recapping events(Djebali et al., 2012) with an IDR value of 0.01 or lower, and summed CAGE-cluster-values over both strands for each nucleotide.

For each annotated transcript we have considered its first nucleotide the TSS of that transcript. For each gene we have defined gene expression as the sum of all cage reads in clusters overlaping the gene's TSS, normalized by the total number of reads sequenced.

Mappability calculation

Mappability for the hg19 genome was calculated using the GEM-mapper for 36bp and 75bp reads. For each acceptor in the genome, mappability was calculated in the direction of transcription.

Exon selection for validation experiments

Out of the exons that had cell type specific H3k9ac peaks that co-occurred with high exon inclusion, we selected a total of 12 exons for validation by RT-PCR and 4 for ChIP.

Cell culture, RNA extraction and RT-PCR analysis

K562 and HeLa cells were grown in Dulbecco's modified Eagle's medium (Gibco BRL) supplemented with 10% of fetal bovine serum (FBS), then penicillin and streptomycin. Gm12878 cells were grown in RPMI (Gibco BRL), supplemented with 15%FBS (Gibco BRL), glycine, penicillin and streptomycin. For RNAi experiments, $2x10^5$ HeLa cells were plated the day before transfection in six wells

plates and transfected with 100 nM of either Brg1 (Dharmacon, L-010431-00-0005), Brm (L-017253-00-0005) or scrambled siRNA oligonucleotides using Lipofectamine RNAiMAX (Invitrogen).
Total RNA was isolated using Qiagen RNeasy mini kit and resuspended in RNAse-free water (Ambion). DNA digestion was performed using RNase-free DNase (Promega). DNA-free total RNA (500 ng) was used for RT−PCR using SuperScript III reverse transcriptase (Invitrogen) using random hexamers and oligo dT. 5% of the reaction was used for Real Time PCR (Applied Biosystem) together with the primers (Table SX) following manufacturer's instructions.

Chromatin immunoprecipitation (ChIP).
Cells were plated at a density of $2X10^5$ and after 48h of culture incubated with 1% (vol/vol) formaldehyde in culture medium for 10 minutes at room temperature. Cells were then washed in cold phosphate-buffered saline (PBS), harvested and lysed in a buffer containing 1% SDS, 10 mM EDTA and 50 mM Tris/HCl pH 8.1, and sonicated in 15ml tubes with Bioruptor UCD-200 Diagenode (ultrasonic wave output power 250W, 30" on-30"off, 4X10') to yield chromatin size of 150-300 bp. 100 µg of DNA/sample were used for immunoprecipitation with 5µg of anti-H3K9ac rabbit (ab4441), anti-H3 rabbit (ab1791) or control rabbit IgGs (Sigma-Aldrich). Co-precipitated DNA was then analyzed by Quantitative real time PCR performed with Sybr Green mix (Applied Biosystem) according to manufacturer's instructions.

References

Alló, M., Buggiano, V., Fededa, J.P., Petrillo, E., Schor, I., de la Mata, M., Agirre, E., Plass, M., Eyras, E., Elela, S.A., *et al.* (2009). Control of alternative splicing through siRNA-mediated transcriptional gene silencing. Nat Struct Mol Biol *16*, 717-724.

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L., and Feuk, L. (2011). Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat Struct Mol Biol *18*, 1435-1440.

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.

Ballare, C., Castellano, G., Gaveglia, L., Althammer, S., Gonzalez-Vallinas, J., Eyras, E., Le Dily, F., Zaurin, R., Soronellas, D., Vicent, G.P., *et al.* (2013). Nucleosome-driven transcription factor binding and gene regulation. Molecular cell *49*, 67-79.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Batsche, E., Yaniv, M., and Muchardt, C. (2006). The human SWI/SNF subunit Brm is a regulator of alternative splicing. Nature structural & molecular biology *13*, 22-29.

Beyer, A.L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes Dev *2*, 754-765.

Carrillo Oesterreich, F., Preibisch, S., and Neugebauer, K.M. (2010). Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. Mol Cell *40*, 571-581.

consortium, T.E.p. (2012). An Integrated Encyclopedia of DNA Elements in the Human Genome. Nature *XXX*.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow

RNA polymerase II affects alternative splicing in vivo. Mol Cell *12*, 525-532.

de la Mata, M., Lafaille, C., and Kornblihtt, A.R. (2010). First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. RNA *16*, 904-912.

Dekker, J., Marti-Renom, M.A., and Mirny, L.A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nature reviews Genetics *14*, 390-403.
Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A.M., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F.*, et al.* (2012). Landscape of transcription in human cells. Nature *submitted*.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H.*, et al.* (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature *462*, 58-64.

Hoffman, M.e.a. (2012). Integrative annotation of chromatin elements from encode data. Genome research *XXX*.

Hon, G., Wang, W., and Ren, B. (2009). Discovery and annotation of functional chromatin signatures in the human genome. PLoS Comput Biol *5*, e1000566.

Iannone, C., and Valcarcel, J. (2013). Chromatin's thread to alternative splicing regulation. Chromosoma *122*, 465-474.

Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.H., Marr, M.T., and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev *25*, 2502-2512.

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.

Kornblihtt, A.R., de la Mata, M., Fededa, J.P., Munoz, M.J., and Nogues, G. (2004). Multiple links between transcription and splicing. RNA *10*, 1489-1498.

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. Nature reviews Molecular cell biology *14*, 153-165.
Kundaje, A. (2011).
https://sites.google.com/site/anshulkundaje/projects/wiggler.

Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. Annals of Applied Statistics *5*, 1752-1779.

Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. Cell *144*, 16-26.

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.

Magen, A., and Ast, G. (2005). The importance of being divisible by three in alternative splicing. Nucleic Acids Res *33*, 5574-5582.

Mercer, T.R., Edwards, S.L., Clark, M.B., Neph, S.J., Wang, H., Stergachis, A.B., John, S., Sandstrom, R., Li, G., Sandhu, K.S.*, et al.* (2013). DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. Nature genetics *45*, 852-859.

Myers, R.M., Stamatoyannopoulos, J., Snyder, M., Dunham, I., Hardison, R.C., Bernstein, B.E., Gingeras, T.R., Kent, W.J., Birney, E., Wold, B.*, et al.* (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol *9*, e1001046.

Nahkuri, S., Taft, R.J., and Mattick, J.S. (2009). Nucleosomes are preferentially positioned at exons in somatic and sperm cells. Cell Cycle *8*, 3420-3424.

Roberts, G.C., Gooding, C., Mak, H.Y., Proudfoot, N.J., and Smith, C.W. (1998). Co-transcriptional commitment to alternative splice site selection. Nucleic Acids Res *26*, 5568-5572.

Saint-André, V., Batsché, E., Rachez, C., and Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1γ favor inclusion of alternative exons. Nat Struct Mol Biol *18*, 337-344.

Schor, I.E., Rascovan, N., Pelisch, F., Allo, M., and Kornblihtt, A.R. (2009a). Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. Proceedings of the National Academy of
Sciences of the United States of America *106*, 4325-4330.

Schor, I.E., Rascovan, N., Pelisch, F., Alló, M., and Kornblihtt, A.R. (2009b). Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. Proc Natl Acad Sci U S A *106*, 4325-4330.

Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. Nat Struct Mol Biol *16*, 990-995.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74-79.

Smith, C.W., and Valcárcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem Sci *25*, 381-388.

Spies, N., Nielsen, C.B., Padgett, R.A., and Burge, C.B. (2009). Biased chromatin signatures around polyadenylation sites and exons. Mol Cell *36*, 245-254.

Tilgner, H., Knowles, D.G., Johnson, R., Davis, C., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., and Guigó, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. Genome research *XXX*.

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J., and Guigó, R. (2009). Nucleosome positioning as a determinant of exon recognition. Nat Struct Mol Biol *16*, 996-1001.
Tyagi, A., Ryme, J., Brodin, D., Ostlund Farrants, A.K., and Visa, N. (2009). SWI/SNF associates with nascent pre-mRNPs and regulates alternative pre-mRNA processing. PLoS genetics *5*, e1000470.

Underhill, C., Qutob, M.S., Yee, S.P., and Torchia, J. (2000). A novel nuclear receptor corepressor complex, N-CoR, contains components of the mammalian SWI/SNF complex and the corepressor KAP-1. The Journal of biological chemistry *275*, 40463-40470.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470-476.

Wang, W. (2003). The SWI/SNF family of ATP-dependent chromatin remodelers: similar mechanisms for diverse functions. Current topics in microbiology and immunology *274*, 143-169.

Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol *11*, 377-394.

Zheng, C.L., Fu, X.D., and Gribskov, M. (2005). Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. RNA *11*, 1777-

Discussion

In 1988 Beyer and Osheim (Beyer and Osheim, 1988) put forward for the first time the idea that splicing can occur co-transcriptionally, and since then it has become clear that processing of RNA happens largely while the pre-mRNA is still tethered to the transcribing RNA polymerase.

The emerging concept that chromatin dynamics and epigenetic modifications can influence and forecast alternative splicing decisions, shaping exon-intron architectures during evolution, is a powerful complement to predictive models based on regulatory sequence motifs identifiable in pre-mRNAs (Barash et al., 2010). It has been suggested that the chromatin epigenetic code and RNA polymerase transcription rates are part of a feed-forward loop that stabilizes inputs leading to alternative splicing changes, in analogy to transcriptional networks in *E. coli* (de Almeida and Carmo-Fonseca, 2012). Evidence, however, remains largely correlative in nature and rigorous mechanistic proof that recruitment of splicing factors through chromatin modifications can be rate-limiting for splicing decisions is not yet generally available.

A series of independent, concurrent studies put forward the concept that chromatin follows an intron-exon structure, and showed that nucleosomes are especially enriched at exons compared to introns (Nahkuri et al., 2009; Schwartz et al., 2009; Spies et al., 2009; Tilgner et al., 2009). A straightforward and appealing hypothesis arising from these findings is that nucleosome density could play a role in regulation of exon

inclusion, serving either as “speed bump” for elongating polymerases or as a landing platform for splicing factor. With this mindset one possible conjecture was that high exonic nucleosome occupancy could favor exon inclusion and, conversely, nucleosome-depleted exons would be mainly skipped.

In the frame of this thesis work, we aimed to better understand connections and interplay between chromatin structure and alternative splicing regulation using two distinct approaches. One of the approaches investigated links between progesterone-induced regulation of splicing and nucleosome occupancy. The second approach took advantage of recent massive datasets comparing epigenetic marks and RNA profiling in a variety of cell lines (The ENCODE consortium, 2012). The first approach tries to evaluate links between chromatin and splicing in a highly controlled manner, with progesterone as the trigger for changes in both. The second approach is of a more holistic nature and tries to capitalize on many comparisons between differentiated cell types and capture commonalities across many biological samples spanning multiple biological processes.

Nucleosome occupancy and splicing regulation

To study the effect of progesterone, we correlated genome-wide analysis through different high-throughput techniques: we draw detailed maps of nucleosome density, before and after hormone induction, and at the same time we described progesterone-induced regulation of exon inclusion of all alternative cassettes. We intersected these datasets, extrapolated common and recurring

behaviors and tried to deduce and model the way nucleosome density changes could influence splicing outcomes. Our results suggest a rich and complex scenario that remains to be fully understood. We analyzed the average nucleosome profile of all internal cassette exons and saw, confirming previous reports, that exons display higher occupancy compared to flanking introns. Using this approach, however, we mask possible distinct profiles associated with particular subclasses of exons. We thus classified exons according to their nucleosome occupancy profile in different clusters, and could distinguish four different nucleosome density patterns mainly determined by GC content distribution in the DNA sequence. In two out of the four classes identified, upstream (cluster B) or downstream (cluster C) introns show high nucleosome accumulation compared to the exons. As others reported, we also see that a conformation of high exonic nucleosome density is associated with exons with weaker splice sites and longer introns. Even though GC content seems to be the main determinant of nucleosome distribution, we see that a percentage of exons change nucleosome profile upon hormone treatment.

In contrast with current models, increases in exon inclusion are not necessarily associated with increases in nucleosome density. Strong nucleosome positioning on exons with weak splice sites and long introns correlates well with both hormone-included and hormone-skipped exons. This suggests that nucleosome conformation contributes to alternative splicing regulation in concert with a number of other features.

RNA pol II ChIP-Seq profiles closely follow nucleosome density, and in exons with well-positioned nucleosomes we see high accumulation over the alternative exon, independent of the splicing outcome. It is possible that a mechanism involving changes in Pol II dynamics and exposure of a number of *cis*-regulatory elements influences regulation of these exons. Our results also suggest that higher nucleosome occupancy upstream of the alternative exon favors exon inclusion rather than skipping.

Nucleosome re-arrangements, altering RNA Pol II dynamics, influence the timing in which exons are processed, possibly shifting their processing from co- to post-transcriptional processing or vice versa, thus exposing them to different regulators. Recent findings show that splicing inhibiting drugs impair to some extent the binding of splicing factor to histones (Convertini et al., 2014) and lead to changes of local levels of histone modifications (Kim et al., 2011) Such drug treatments could help to dissect the mechanism of interplay between chromatin structure and splicing regulation, providing information regarding co-trascriptional recruitment of the spliceosome in time and space (along the gene).

Studies focused on understanding how specific splicing factors modulate inclusion of concrete alternative exons exploit simple, alterable *in vitro* systems (Chen and Manley, 2009). Unfortunately it remains technically challenging to modulate nucleosome density or epigenetic marks locally and at will and consequently our results remain correlative. The predictive power of some of these correlations (e.g. combination of certain epigenetic marks and exon inclusion), however, argues in favor of causal effects.

Co-occurrence of promoter-like histone modifications and splicing regulation

In the second part of this work we carried out pairwise comparison between different cell lines and identified a subset of exons whose inclusion is influenced by epigenetic signatures. Several previous studies revealed links between accumulation of specific histone marks and splicing outcome. Histone modifications can influence splicing outcomes directly, through recruitment of splicing factors, or indirectly, through modulation of RNA Pol II dynamics (reviewed in Luco et al., 2011; Iannone and Valcarcel, 2013). Exploiting ENCODE data, we combined polyA+ nuclear RNA-Seq with ChIP-Seq from different histone modifications across five cell lines and we found positive correlation between accumulation of H3K9ac, H3K27ac, H3K4me3 and CTCF and higher exon inclusion. Analyzing each exon as a triplet of "constitutive-alternative-constitutive" exons, we see that for these histone modifications the accumulation is local and limited to the alternative exon cassette. Moreover, we identify a subset of 100 exons for which H3K9ac, H3K27ac, H3K4me3 co-occur and are sufficient to predict splicing differences between cell lines. Since these histone marks are characteristic of promoter regions, we first excluded the possibility that these 100 exons were un-annotated promoters, and used as transcription start sites.

Recent work reported that some internal exons are marked by enhancer-like and promoter-like signature, due to physical association of internal exons with cognate promoters or enhancers (Mercer et al., 2013). We find that co-occurrence of H3K9ac, H3K27ac and H3K4me3 on cassette exons associates with

significant enrichment in DHS and ChIA-PET signals, suggesting the existence of long-range interactions between the exons and promoters.

The accumulation of H3K4me3, H3K27ac and H3K9ac together with DNase hypersensitive sites implies that exons displaying higher inclusion are in a region of open chromatin, which would typically lead to faster RNA Pol II elongation rates and also typically to higher levels of exon skipping. We do detect, however, accumulation of total Pol II in these exons. If indeed efficiently included alternative exons are engaged in a loop with promoters, accumulation of RNA Pol II signal could actually correspond to high levels of Polymerase bound to transcription initiation sites. On the other hand, the accumulation over the alternative exons could be due to some other factors binding in these DHS marked exons.

Our results argue for a role of three-dimensional organization of the genome in alternative splicing regulation, perhaps operating also at other steps of pre-mRNA processing. These observations should be further validated and complemented with analysis of conformation capture technologies to understand the underlying molecular mechanisms.

Conclusions

Part I

1. Progesterone regulates 243 internal exon cassettes, mostly independently from transcriptional changes

2. Nucleosomes display higher density in exons compared to introns, and their distribution around exons follows four distinct patterns

3. DNA sequence is the main determinant of nucleosome positioning, but progesterone induction results in nucleosome rearrangement on a subset of both regulated and not regulated exons

4. Progesterone-skipped exons tend to undergo less nucleosome rearrangements than induced exons, and display less pronounced nucleosome peaks

5. RNA-Polymerase II accumulation is highly influenced by nucleosome density

6. High nucleosome density upstream of alternative exons correlates with higher levels of exon inclusion

Part II

1. H3K4me3, H3K9ac, H3K27ac and CTCF are enriched in exons that exhibit higher levels of inclusion in pairwise comparison between a variety of cell lines

2. H3K4me3, H3K9ac and H3K27ac co-occur in a subset of exons with higher inclusion ratios, and their relative levels are sufficient to predict inclusion differences between cell lines

3. DNase I hypersensitivity sites, Pol II ChIP-Seq signals and long-range chromatin interaction signatures are higher in exons harboring H3K4me3, H3K9ac and H3K27ac, and correlate well with higher exon inclusion

4. Transcription factor ChIP-Seq signals are also higher in exons enriched in the previously mentioned chromatin features

5. These observations suggest a functional role for 3-dimensional genome structure in the regulation of alternative splicing.

Bibliography

Allo, M., Buggiano, V., Fededa, J.P., Petrillo, E., Schor, I., de la Mata, M., Agirre, E., Plass, M., Eyras, E., Elela, S.A., et al. (2009). Control of alternative splicing through siRNA-mediated transcriptional gene silencing. Nature structural & molecular biology 16, 717-724.

Alexander, R.D., Innocente, S.A., Barrass, J.D., and Beggs, J.D. (2010). Splicing-dependent RNA polymerase pausing in yeast. Molecular cell 40, 582-593.

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L., and Feuk, L. (2011). Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nature structural & molecular biology 18, 1435-1440.

Ameyar-Zazoua, M., Rachez, C., Souidi, M., Robin, P., Fritsch, L., Young, R., Morozova, N., Fenouil, R., Descostes, N., Andrau, J.C., et al. (2012). Argonaute proteins couple chromatin silencing to alternative splicing. Nature structural & molecular biology 19, 998-1004.

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 research 19, 1732-1741.

Ballare, C., Castellano, G., Gaveglia, L., Althammer, S., Gonzalez-Vallinas, J., Eyras, E., Le Dily, F., Zaurin, R., Soronellas, D., Vicent, G.P., et al. (2013). Nucleosome-driven transcription factor binding and gene regulation. Molecular cell 49, 67-79.

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.

Barboric, M., Lenasi, T., Chen, H., Johansen, E.B., Guo, S., and Peterlin, B.M. (2009). 7SK snRNP/P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development. Proceedings of the National Academy of Sciences of the United States of America 106, 7798-7803.

Batsche, E., Yaniv, M., and Muchardt, C. (2006). The human SWI/SNF subunit Brm is a regulator of alternative splicing. Nature structural & molecular biology 13, 22-29.

Beato, M. (1996). Chromatin structure and the regulation of gene expression: remodeling at the MMTV promoter. J Mol Med (Berl) 74, 711-724.

Beato, M. (2008). The time course of transmitter at glycinergic synapses onto motoneurons. The Journal of neuroscience : the official journal of the Society for Neuroscience 28, 7412-7425.

Beato, M., and Klug, J. (2000). Steroid hormone receptors: an update. Human reproduction update 6, 225-236.

Beckmann, J.S., and Trifonov, E.N. (1991). Splice junctions follow a 205-base ladder. Proceedings of the National Academy of Sciences of the United States of America 88, 2380-2383.

Beleut, M., Rajaram, R.D., Caikovski, M., Ayyanan, A., Germano, D., Choi, Y., Schneider, P., and Brisken, C. (2010). Two distinct mechanisms underlie progesterone-induced proliferation in the mammary gland. Proceedings of the National Academy of Sciences of the United States of America 107, 2989-2994.

Beyer, A.L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. Genes & Development 2, 754-765.

Bieberstein, N.I., Carrillo Oesterreich, F., Straube, K., and Neugebauer, K.M. (2012). First exon length controls active chromatin signatures and transcription. Cell reports 2, 62-68.

Bonnal, S., Vigevani, L., and Valcarcel, J. (2012). The spliceosome as a target of novel antitumour drugs. Nature reviews Drug discovery 11, 847-859.

Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., and Blencowe, B.J. (2013). Dynamic Integration of Splicing within Gene Regulatory Pathways. Cell 152, 1252-1269.

Brisken, C. (2013). Progesterone signalling in breast cancer: a neglected hormone coming into the limelight. Nature reviews Cancer 13, 385-396.

Brody, Y., Neufeld, N., Bieberstein, N., Causse, S.Z., Bohnlein, E.M., Neugebauer, K.M., Darzacq, X., and Shav-Tal, Y. (2011). The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. PLoS biology 9, e1000573.

Carrillo Oesterreich, F., Preibisch, S., and Neugebauer, K.M. (2010). Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. Molecular cell 40, 571-581.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., et al. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell 122, 33-43.

Chen, L., Xiao, S., and Manley, N.R. (2009). Foxn1 is required to maintain the postnatal thymic microenvironment in a dosage-sensitive manner. Blood 113, 567-574.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nature reviews Molecular cell biology 10, 741-754.

Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J., et al. (2010). Relationship between nucleosome positioning and DNA methylation. Nature 466, 388-392.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469, 368-373.

Clarke, R.B., Howell, A., Potten, C.S., and Anderson, E. (1997). Dissociation between steroid receptor expression and cell proliferation in the human breast. Cancer research 57, 4987-4991.

Close, P., East, P., Dirac-Svejstrup, A.B., Hartmann, H., Heron, M., Maslen, S., Chariot, A., Soding, J., Skehel, M., and Svejstrup, J.Q. (2012). DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. Nature 484, 386-389.

Convertini, P., Shen, M., Potter, P.M., Palacios, G., Lagisetti, C., de la Grange, P., Horbinski, C., Fondufe-Mittendorf, Y.N., Webb, T.R., and

Stamm, S. (2014). Sudemycin E influences alternative splicing and changes chromatin modifications. Nucleic acids research.

Darnell, J.E., Jr. (2013). Reflections on the history of pre-mRNA processing and highlights of current knowledge: A unified picture. RNA 19, 443-460.

de Almeida, S.F., and Carmo-Fonseca, M. (2012). Design principles of interconnections between chromatin and pre-mRNA splicing. Trends in biochemical sciences 37, 248-253.

de Almeida, S.F., Grosso, A.R., Koch, F., Fenouil, R., Carvalho, S., Andrade, J., Levezinho, H., Gut, M., Eick, D., Gut, I., et al. (2011). Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. Nature structural & molecular biology 18, 977-983.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M.a., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. Molecular cell 12, 525-532.

de la Mata, M., Lafaille, C., and Kornblihtt, A.R. (2010). First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. RNA 16, 904-912.

Dekker J, Marti-Renom MA, Mirny LA. (2013) Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nature reviews Genetics. 14, 390-403

Dhami, P., Saffrey, P., Bruce, A.W., Dillon, S.C., Chiang, K., Bonhoure, N., Koch, C.M., Bye, J., James, K., Foad, N.S., et al. (2010). Complex exon-intron marking by histone modifications is not determined solely by nucleosome distribution. PloS one 5, e12339.

Dye, M.J., Gromak, N., and Proudfoot, N.J. (2006). Exon tethering in transcription by RNA polymerase II. Molecular cell 21, 849-859.

Filipowicz, W., Jaskiewicz, L., Kolb, F.A., and Pillai, R.S. (2005). Post-transcriptional gene silencing by siRNAs and miRNAs. Current opinion in structural biology 15, 331-341.

Fong, Y.W., and Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. Nature 414, 929-933.

Gelfman, S., Cohen, N., Yearim, A., and Ast, G. (2013). DNA-methylation effect on co-transcriptional splicing is dependent on GC-architecture of the exon-intron structure. Genome research.

Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K., and Luhrmann, R. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. Nature communications 3, 994.

Gunderson, F.Q., and Johnson, T.L. (2009). Acetylation by the transcriptional coactivator Gcn5 plays a novel role in co-transcriptional spliceosome assembly. PLoS genetics 5, e1000682.

Gunderson, F.Q., Merkhofer, E.C., and Johnson, T.L. (2011). Dynamic histone acetylation is critical for cotranscriptional spliceosome assembly and spliceosomal rearrangements. Proceedings of the National Academy of Sciences of the United States of America 108, 2004-2009.

Hnilicova, J., Hozeifi, S., Duskova, E., Icha, J., Tomankova, T., and Stanek, D. (2011). Histone deacetylase activity modulates alternative splicing. PloS one 6, e16727.

Hodges, E., Smith, A.D., Kendall, J., Xuan, Z., Ravi, K., Rooks, M., Zhang, M.Q., Ye, K., Bhattacharjee, A., Brizuela, L., et al. (2009). High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. Genome research 19, 1593-1605.

Howe KJ, Kane CM, Ares M Jr.(2003) Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae. RNA. 2003 8, 993-1006

Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., and Carroll, J.S. (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nature genetics 43, 27-33.

Iannone, C., and Valcarcel, J. (2013). Chromatin's thread to alternative splicing regulation. Chromosoma 122, 465-474.

Ip, J.Y., Schmidt, D., Pan, Q., Ramani, A.K., Fraser, A.G., Odom, D.T., and Blencowe, B.J. (2011). Global impact of RNA polymerase II

elongation inhibition on alternative splicing regulation. Genome research 21, 390-401.

Ji X, Zhou Y, Pandit S, Huang J, Li H, Lin CY, Xiao R, Burge CB, Fu XD. (2013) SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. Cell 153: 855-68

John, S., Sabo, P.J., Thurman, R.E., Sung, M.H., Biddie, S.C., Johnson, T.A., Hager, G.L., and Stamatoyannopoulos, J.A. (2011). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nature genetics 43, 264-268.

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., and Stamm, S. (2013). Function of alternative splicing. Gene 514, 1-30.

Keren-Shaul, H., Lev-Maor, G., and Ast, G. (2013). Pre-mRNA splicing is a determinant of nucleosome organization. PloS one 8, e53506.

Khare, T., Pai, S., Koncevicius, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., Irimia, M., Jia, P., Ptak, C., Xia, M., et al. (2012). 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. Nature structural & molecular biology 19, 1037-1043.

Khodor, Y.L., Menet, J.S., Tolan, M., and Rosbash, M. (2012). Cotranscriptional splicing efficiency differs dramatically between Drosophila and mouse. RNA 18, 2174-2186.

Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.H., Marr, M.T., 2nd, and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes & development 25, 2502-2512.

Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D.L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. Proceedings of the National Academy of Sciences of the United States of America 108, 13564-13569.

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. Nature genetics 41, 376-381.

Kornblihtt, A.R. (2007). Coupling transcription and alternative splicing. Advances in experimental medicine and biology 623, 175-189.

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. Nature reviews Molecular cell biology 14, 153-165.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell 128, 693-705.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science 339, 950-953.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell 128, 707-719.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. Nature structural & molecular biology 15, 819-826.

Listerman, I., Sapra, A.K., and Neugebauer, K.M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nature structural & molecular biology 13, 815-822.

Loomis RJ, Naoe Y, Parker JB, Savic V, Bozovsky MR, Macfarlan T, Manley JL, Chakravarti D. (2009) Chromatin binding of SRp20 and ASF/SF2 and dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation. Molecular Cell 33, 450-61

Lorincz, M.C., Dickerson, D.R., Schmitt, M., and Groudine, M. (2004). Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nature structural & molecular biology 11, 1068-1075.

Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. Cell 144, 16-26.

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.

Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., and Maleszka, R. (2010). The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS biology 8, e1000506.

Martins, S.B., Rino, J., Carvalho, T., Carvalho, C., Yoshida, M., Klose, J.M., de Almeida, S.F., and Carmo-Fonseca, M. (2011). Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3' end of human genes. Nature structural & molecular biology 18, 1115-1123.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature 385, 357-361.

Mercer, T.R., Edwards, S. L., Clark, M.B., Neph, S.J., Wang, H., Stergachis, A.B., John, S., Sandstrom, R., Li, G., Sandhu, K.S., Ruan, Y., Nilesen, L.K., Mattick, J.S., and Stamatoyannopoulos, J.A. (2013). DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. Nature genetics, advanced online publication.

Montano, M.M., Muller, V., Trobaugh, A., and Katzenellenbogen, B.S. (1995). The carboxy-terminal F domain of the human estrogen receptor: role in the transcriptional activity of the receptor and the effectiveness of antiestrogens as estrogen antagonists. Mol Endocrinol 9, 814-825.

Munoz, M.J., de la Mata, M., and Kornblihtt, A.R. (2010). The carboxy terminal domain of RNA polymerase II and alternative splicing. Trends in biochemical sciences 35, 497-504.
Munoz, M.J., Perez Santangelo, M.S., Paronetto, M.P., de la Mata, M., Pelisch, F., Boireau, S., Glover-Cutter, K., Ben-Dov, C., Blaustein, M., Lozano, J.J., et al. (2009). DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. Cell 137, 708-720.

Nahkuri, S., Taft, R.J., and Mattick, J.S. (2009). Nucleosomes are preferentially positioned at exons in somatic and sperm cells. Cell Cycle 8, 3420-3424.

Nandi, S. (1958). Endocrine control of mammarygland development and function in the C3H/ He Crgl mouse. Journal of the National Cancer Institute 21, 1039-1063.

Nogues, G., Kadener, S., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2002). Transcriptional activators differ in their abilities to control

alternative splicing. The Journal of biological chemistry 277, 43110-43114.

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. Nature genetics 40, 1413-1415.

Pandya-Jones, A., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons. RNA 15, 1896-1908.

Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R., and Bickmore, W.A. (2012). Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. PLoS genetics 8, e1002717.

Rauch, T.A., Wu, X., Zhong, X., Riggs, A.D., and Pfeifer, G.P. (2009). A human B cell methylome at 100-base pair resolution. Proceedings of the National Academy of Sciences of the United States of America 106, 671-678.

Saint-Andre, V., Batsche, E., Rachez, C., and Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1gamma favor inclusion of alternative exons. Nature structural & molecular biology 18, 337-344.

Sanchez, G., Bittencourt, D., Laud, K., Barbier, J., Delattre, O., Auboeuf, D., and Dutertre, M. (2008). Alteration of cyclin D1 transcript elongation by a mutated transcription factor up-regulates the oncogenic D1b splice isoform in cancer. Proceedings of the National Academy of Sciences of the United States of America 105, 6004-6009.

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. Proceedings of the National Academy of Sciences of the United States of America 106, 4325-4330.

Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. Nature structural & molecular biology 16, 990-995.

Shieh, G.S., Pan, C.H., Wu, J.H., Sun, Y.J., Wang, C.C., Hsiao, W.C., Lin, C.Y., Tung, L., Chang, T.H., Fleming, A.B., et al. (2011). H2B ubiquitylation is part of chromatin architecture that marks exon-intron structure in budding yeast. BMC genomics 12, 627.

Shukla, S., and Oberdoerffer, S. (2012). Co-transcriptional regulation of alternative pre-mRNA splicing. Biochimica et biophysica acta 1819, 673-683.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 479, 74-79.

Sims, R.J., 3rd, Millhouse, S., Chen, C.F., Lewis, B.A., Erdjument-Bromage, H., Tempst, P., Manley, J.L., and Reinberg, D. (2007). Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. Molecular cell 28, 665-676.

Smallwood A, Hon GC, Jin F, Henry RE, Espinosa JM, Ren B. (2012) CBX3 regulates efficient RNA processing genome-wide. Genome Research 22, 1426-36.

Spies, N., Nielsen, C.B., Padgett, R.A., and Burge, C.B. (2009). Biased chromatin signatures around polyadenylation sites and exons. Molecular cell 36, 245-254.

Taliaferro, J.M., Aspden, J.L., Bradley, T., Marwha, D., Blanchette, M., and Rio, D.C. (2013). Two new and distinct roles for Drosophila Argonaute-2 in the nucleus: alternative pre-mRNA splicing and transcriptional repression. Genes & development 27, 378-389.

The ENCODE consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57-74.

Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., and Guigo, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. Genome research 22, 1616-1625.

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcarcel, J., and Guigo, R. (2009). Nucleosome positioning as a determinant of exon recognition. Nature structural & molecular biology 16, 996-1001.

Torchia, J., Rose, D.W., Inostroza, J., Kamei, Y., Westin, S., Glass, C.K., and Rosenfeld, M.G. (1997). The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function. Nature 387, 677-684.

Vargas, D.Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S.A., Schedl, P., and Tyagi, S. (2011). Single-molecule imaging of transcriptionally coupled and uncoupled splicing. Cell 147, 1054-1065.

Vicent, G.P., Ballare, C., Nacht, A.S., Clausell, J., Subtil-Rodriguez, A., Quiles, I., Jordan, A., and Beato, M. (2006). Induction of progesterone target genes requires activation of Erk and Msk kinases and phosphorylation of histone H3. Molecular cell 24, 367-381.

Vicent, G.P., Nacht, A.S., Smith, C.L., Peterson, C.L., Dimitrov, S., and Beato, M. (2004). DNA instructed displacement of histones H2A and H2B at an inducible promoter. Molecular cell 16, 439-452. Vicent, G.P., Nacht, A.S., Zaurin, R., Ballare, C., Clausell, J., and Beato, M. (2010). Minireview: role of kinases and chromatin remodeling in progesterone signaling to chromatin. Mol Endocrinol 24, 2088-2098.

Voegel, J.J., Heine, M.J., Tini, M., Vivat, V., Chambon, P., and Gronemeyer, H. (1998). The coactivator TIF2 contains three nuclear receptor-binding motifs and mediates transactivation through CBP binding-dependent and -independent pathways. The EMBO journal 17, 507-519.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell 136, 701-718.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470-476.

Wang, Z., and Burge, C.B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA 14, 802-813.

Zhong, X.Y., Wang, P., Han, J., Rosenfeld, M.G., and Fu, X.D. (2009). SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. Molecular cell 35, 1-10.

Zhou, B., Yan, H., Li, Y., Wang, R., Chen, K., Zhou, Z., and Sun, X. (2012). PNAS-4 expression and its relationship to p53 in colorectal cancer. Molecular biology reports 39, 243-249.

Zhu, H., Hasman, R.A., Barron, V.A., Luo, G., and Lou, H. (2006). A nuclear function of Hu proteins as neuron-specific alternative RNA