Charles University in Prague
Faculty of Science

Molecular and Cell Biology, Genetics and Virology

Zuzana Krchňáková

Determinants of the splice site selection in protein-coding and long non-coding RNAs

Klíčové faktory při výběru sestřihových míst v kódujících a v dlouhých nekódujících RNA

Doctoral thesis

Supervisor: Doc. David Staněk, PhD.
Laboratory of RNA Biology
Institute of Molecular Genetics, Czech Academy of Science

Prague, 2018

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

Praha, 2018 Zuzana Krchňáková

I hereby declare that I wrote the thesis independently and I cited all informational sources. This work or a substantial part of it was not presented to obtain another academic degree or equivalent.

Prague, 2018 Zuzana Krchňáková

Acknowledgements

Here, I would like to thank my supervisor, David Staněk, for the great leading and support through my PhD study. I would also like to thank Michaela Krausová for helping me with experimental designs and preparation of CRISPR/Cas9 cells, as well as Prasoon Kumar Thakur for doing all the bioinformatic analyses which further improved the project. I am also very thankful that I could work in a great team of RNA Biology laboratory. Finally, my great thanks also go to my husband for his endless support throughout the whole of my studies.

Table of Contents

Abstract

In my thesis, I focused on several underexplored areas of RNA splicing regulation. In the first part, I analyzed how chromatin and transcription regulatory elements change pre-mRNA splicing. In the second part, I studied why long non-coding RNAs (lncRNAs) are spliced less efficiently than protein-coding mRNAs. Finally, I was testing the importance of intron for the activating function of lncRNAs.

It has been shown that chromatin and promoter identity modulate alternative splicing decisions. Here, I tested whether local chromatin and distant genomic elements that influence transcription can also modulate splicing. Using the chromatin modifying enzymes directly targeted to *FOSL1* gene by TALE technology, I showed that changes in histone H3K9 methylation affect constitutive splicing. Furthermore, I provide evidence that deletion of transcription enhancer located several kilobases upstream of an alternative exons changes splicing pattern of the alternative exon.

Many nascent lncRNAs undergo the same maturation steps as pre-mRNAs of protein-coding genes (PCGs), but they are often poorly spliced. To identify the underlying mechanisms for this phenomenon, we searched for putative splicing inhibitory sequences. Genome-wide analysis of intergenic lncRNAs (lincRNAs) revealed that, in general, they do not contain more splicing inhibitory sequences compared to PCGs. Using *ncRNA-a2* as a model, we provide evidence that its inefficient splicing is independent of chromatin or promoter sequence. On the contrary, we show that the intron sequence of ncRNA-a2 is a major determinant of its inefficient splicing. Additionally, we provide experimental evidence that the strengthening of the 5' splice site and increasing the thymidine content in polypyrimidine tract significantly enhance lincRNA splicing. We further show that lincRNA exons contain less putative binding sites for SR proteins and are bound to a much lower extent by SR proteins than expression-matched PCGs. We propose that lincRNAs lack the cooperative interaction network that enhances splicing, which renders their splicing outcome more dependent on the optimality of splice sites.

Finally, we removed intron from *ncRNA-a2* and tested whether the splicing process is important for the function of an enhancer-like lncRNA. My results suggest the functionality of DNA element of *ncRNA-a2* locus rather the RNA product itself in the promoting transcription of neighboring genes. However, we could not distinguish between

these two possibilities thus future experiments have to be done to provide a definite answer.

Abstrakt

Ve své dizertační práci jsem se zaměřila na několik málo proskoumaných oblastí regulace sestřihu RNA. V první části jsem analyzovala, jak chromatinové a transkripční regulační elementy mění sestřih pre-mRNA. Ve druhé části jsem studovala, proč jsou dlouhé nekódující RNA méně efektivně stříhané než mRNA kódující proteiny. Nakonec jsem zkoumala důležitost intronu pro aktivační funkci dlouhých nekódujících RNA.

Bylo ukázáno, že chromatin a promotor mění alternativní sestřih. Zde jsem testovala, zda lokální chromatinové a vzdálené genomové prvky, které ovlivňují transkripci, mohou také modulovat sestřih. Použila jsem enzymy modifikující histony a pomocí TALE technologie je navedla na specifické oblasti ve FOSL1 genu. Pomocí tohoto přístupu jsem ukázala, že změny metylace v lyzínu 9 histónu H3 ovlivňují konstitutivní sestřih. Navíc podávám důkaz, že odstranení transkripčního zesilovače vzdáleného několik kilobází od alternativního exonu mění sestřih alternativního exonu.

Mnohé dlouhé nekódujíci RNA podléhají stejnému mechanizmu zpracování jako pre-mRNA genů kódujících proteiny, ale často jsou neefektivně sestřiženy. Abychom identifikovali základní mechanismy tohoto jevu, hledali jsme možné inhibiční sekvence sestřihu u těchto dlouhých nekódujících RNA. Celogenomová analýza ukázala, že obecně dlouhé nekódujíci RNA neobsahují více inhibičních sekvencí sestřihu ve srovnání s geny kódující proteiny. Abych identifikovala sekvence inhibující sestřih nekódujících RNA, použila jsem ncRNA-a2 jako modelovou nekódující RNA a ukázala jsem, že neefektivní sestřih je nezávislý na sekvenci chromatinu nebo promotoru. Naopak, moje výsledky ukazují, že intronová sekvence ncRNA-a2 je hlavním určujícím činitelem neefektivního sestřihu. Dále poskytujeme experimentální důkazy, že zesílení 5'ss a zvýšení obsahu tymidinů v polypyrimidínové oblasti významně zlepšují sestřih dlouhých nekódujících RNA. Dále ukazujeme, že exony dlouhých nekódujících RNA obsahují méně vazebných míst pro SR proteiny a jsou vázány v mnohem menším rozsahu SR proteiny než mRNA kódující proteiny. Na základě našich výsledků navrhujeme, že dlouhým nekódujícím RNA schází komplexní interakční síť, která zlepšuje sestřih, což způsobuje, že výsledek jejich sestřihu je více závislý na optimální sekvenci sestřihových míst, které interagují přímo se sestřihovým komplexem.

Nakonec jsme odstranili intron z ncRNA-a2 a testovali jsme, zda je proces sestřihu důležitý pro funkci této dlouhé nekódující RNA. Moje výsledky naznačují funkci DNA elementu spíše než samotného RNA produktu při podpoře transkripce sousedních genů. Bohužel jsme však nemohli rozlišit mezi těmito dvěma možnostmi, proto jsou třeba provést další experimenty s cílem poskytnout jednoznačnou odpověď.

Summary

The mutual influence of transcription, chromatin and splicing on each other has been known for several years. Previously, most studies about the influence of chromatin marks on splicing have used a global approach to perturb histone modifications genome-wide. However, the secondary effects of a globally affected transcriptional program of the cell this way cannot be fully excluded. Because of that, we used histone modification domains targeted locally at a selected target exon by TALE approach to alter H3K36 and H3K9 methylation and observed changes in alternative splicing. Together with global enrichment of H3K9me3 around internal exons, we hypothesize this histone mark at nucleosomes plays a general role in exon recognition. Moreover, there is evidence showing the transcription enhancer and promoter sequences can influence splicing independently of transcription regulation at the minigene background. However, the question if such regulation also occurs at a great distance at endogenous loci is not answered. In the first project, using CRISPR/Cas9 method, we endogenously cut out the DNA enhancer element and showed that enhancer element located several tens of thousands of bases upstream of alternatively spliced exon can influence its splicing. One of the proposed mechanisms of such splicing regulation is modulating chromatin modifications, mainly over the region of the altered exon.

Furthermore, we have also looked at the splicing of long non-coding RNAs (lncRNAs). Many studies have recently shown that nascent lncRNAs undergo the same maturation steps as pre-mRNAs of protein-coding genes (PCGs), but they are often poorly spliced when compared to PCGs. In the second project, we have decided to identify the underlying mechanisms for this phenomenon, firstly by searching for putative splicing inhibitory sequences. Genome-wide analysis of intergenic lncRNAs (lincRNAs) revealed that, in general, they do not contain more splicing inhibitory sequences compared to PCGs suggesting that prevalence of splicing inhibitors is not the major cause of the inefficient splicing of lincRNAs. Using ncRNA-a2 as a model, we suggest that its inefficient splicing is independent on chromatin or promoter sequence since its transient expression from a plasmid under the CMV promoter did not influence its splicing efficiency. On the contrary, we provide evidence that the intron sequence of ncRNA-a2 is a major determinant of its inefficient splicing because intron of a PCG exhibited efficient splicing when placed between ncRNA-a2 exons and at the same time, ncRNA-a2 intron was not spliced when

placed between exons of a PCG. Surprisingly, by extensive mutagenesis, we found that middle region of ncRNA-a2 intron harbors G-rich intronic splicing enhancers which are likely regulated by hnRNP H protein since its siRNA-mediated knock-down increased unspliced transcripts of ncRNA-a2. Additionally, we provided experimental evidence that strengthening of the 5'ss and increasing the thymidine content in polypyrimidine tract significantly enhance lincRNA splicing. We further show that lincRNA exons contain less putative binding sites for SR proteins than PCGs and are bound to a much lower extent by SR proteins than expression-matched PCGs. From these results, we propose that lincRNAs lack the cooperative interaction network that enhances splicing, which renders their splicing outcome more dependent on the optimality of splice sites.

The third project was focused on the activating function of lncRNAs and specifically on the importance of the splicing process on the function of enhancer-like lncRNAs. At the beginning, we confirmed previously shown results that *ncRNA-a2* and *ncRNA-a5* act as transcription enhancers because their depletion by RNAi decreased the expression of adjacent PCGs. Moreover, we found that *ncRNA-a2* seems to act in *cis* because its overexpression from a CMV-driven plasmid did not increase expression of its target PCG. In addition, we have utilized newly emerged techniques for endogenous transcription activation or repression (CRISPRa and CRISPRi). By the targeting transcriptional activation or repression domain to the promoter of *ncRNA-a2*, we modulated the expression of *ncRNA-a2* as well as its neighboring genes. These experiments suggest that *ncRNA-a2* expression does not have a direct impact on the transcription of neighboring genes. We hypothesise that the *ncRNA-a2* promoter rather than the RNA product is important for the expression of neighboring PCGs. Finally, we analyzed whether ncRNA-a2 intron is important for its activating function. Using CRISPR/Cas9, we cut out intron sequence from *ncRNA-a2* and showed that the splicing process is not important for the function of enhancer-like lncRNA supporting the *ncRNA-a2* promoter rather than the RNA product promotes transcription of neighboring genes.

Aims of the Study

The main goal of this work is to further elucidate the regulation of the splicing process. I focused on three major topics: 1. A role of chromatin and enhancers in splicing regulation, 2. Molecular explanation why long non-coding RNAs (lncRNAs) are in general less efficiently spliced than protein-coding pre-mRNAs, and 3. Importance of intron for the enhancer function of activating lncRNA.

The main aims of my thesis:

- Assay the role of chromatin on alternative splicing by local targeting of histone modification domains.
- Assess the possible role of a transcription enhancer element in the regulation of splicing.
- Determine which sequence(s) is responsible for inefficient splicing of lncRNAs.
- Define the role of 5'ss and polypyrimidine tract in splicing of lncRNAs.
- Evaluate the contribution of several splicing factors for splicing of lncRNAs.
- Describe the role of introns in the function of lncRNAs.

Literature Review

Process of Splicing

Most of the genes in phylogenetically younger eukaryotes are split into coding (exons) and non-coding regions (introns). These genes are transcribed into pre-mRNA transcripts which have to be processed to their maturated form. Such RNA processing usually involves 5'capping, 3'tailing and splicing during of which introns are spliced out, and remaining exons are joined together. For the first time, the process of splicing was observed in mRNAs of adenoviruses (Berget et al. 1977; Chow et al. 1977) and mRNA of the mouse β-globin gene (Konkel et al. 1978).

The exon-intron boundaries are called 5' or 3' splice site (ss) and are represented by conserved consensus sequences (**Figure 1**). At 5'end, the exon-intron boundary consists of AG|GURAGU sequence (slash represents exon-intron boundary), 3'end consists of YAG|G sequence (Brent and Guigó 2004; Carmel et al. 2004; Padgett 2012). Additional conserved sequences playing an important role in the intron recognition are branch point (BP) and polypyrimidine tract (PPT) both localized near 3'end of introns.

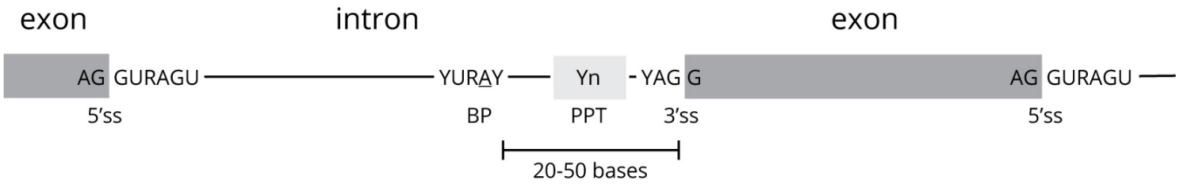

Figure 1. Conserved sequences at exon/intron boundaries. These sequences are conserved throughout evolution since there are recognized by regulatory RNAs/proteins. 5'ss and 3'ss – splice sites, R – purine, Y – pyrimidine, BP – branch point, subscribed A represents the base mediating branching of intron lariat, and PPT – polypyrimidine tract.

Splicing is mediated by ribonucleoprotein (RNP) complex called spliceosome composed of five U-rich small nuclear RNAs (U snRNAs) and numerous proteins. Together, they form the small nuclear ribonucleoprotein (snRNP) particles (reviewed in Will and Lührmann 2011; Matera and Wang 2014). The major spliceosome is assembled from U1, U2, U4, U5 and U6 snRNAs associated with U snRNA-specific Sm proteins together with non-snRNP proteins (**Figure 2**).

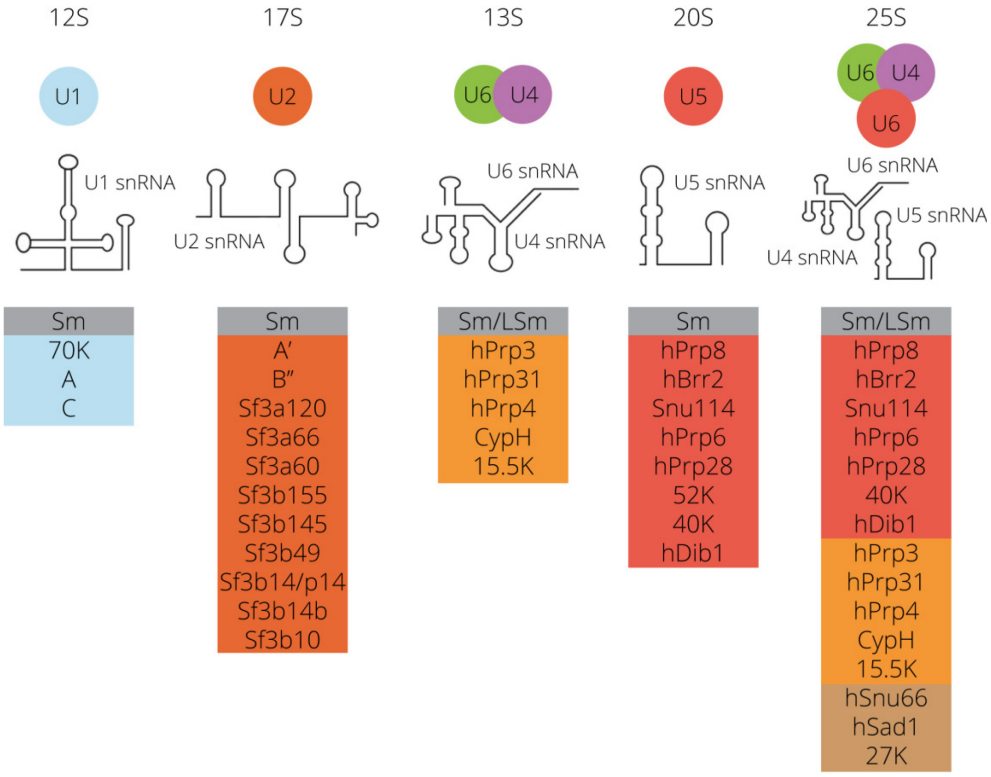

Figure 2. The components of the major human spliceosome. Individual snRNP particles are composed of one U snRNA (of two in the case of U4/U6 and three in the case of U4/U6.U5 tri-snRNP), a common core of seven Sm/LSm proteins (B/B', D3, D2, D1, E, F, and G) and variable number of U snRNA's specifically-bound proteins. U snRNAs are showed in their most probable secondary structures (adapted from Will and Lührmann 2011).

In most eukaryotes, two different types of spliceosomes are found. The more prevalent is the canonical U2-dependent spliceosome which contains the U2 snRNP. On the other hand, the second and minor form of spliceosome contains the U12 snRNP which only accounts for less than 0.5% of all introns (reviewed in Patel and Steitz 2003; Turunen et al. 2013). However, besides the U12 snRNP, the minor spliceosome consists of U11, U4atac and U6atac snRNPs. Moreover, splice sites of the minor introns are sometimes deviated from the canonical ones and are consisted of AT-AC dinucleotides at the 5' and 3'end of introns. However, the more common subtype of U12-type introns has still major GT-AG dinucleotides (Dietrich et al. 1997). Thus, the defining features of U12-type introns are the 5′ss and BP, which are more conserved than in most other introns (Burge et al. 1998). U12-type introns are spliced less efficiently than the major introns, and it is believed that this limits the expression of the genes containing such introns (reviewed in Turunen et al. 2013).

The process of splicing occurs via two *trans*-esterification reactions which result in the excision of intron sequences and joining the exons. In the first reaction, 2'OH group of BP (adenosine) located near 3'end of the intron attacks 5'ss. The formation of the 2'-5'phosphodiester bond between 5'end of intron and attacking adenosine results in the branched lariat, and 2'OH group of the 5'exon becomes free. The second step is consisting of attacking the 3'ss by 2'OH group at 5'exon resulting in ligation of exons and excision of intron lariat.

The whole process of splicing occurs in a stepwise manner (Will and Lührmann 2011). At the beginning, the intron boundaries have to be identified (**Figure 3**). This process is tightly regulated and initiated by the binding of U1 snRNA together with U1C protein to the 5'ss. The earliest spliceosome complex is formed (complex E). The 3'end of intron, including BP, PPT, and the YAG motif of 3'ss, is bound by U2 snRNP-associated factors (splicing factor 1 - SF1, U2 auxiliary factors 1 and 2 - U2AF1/2), which subsequently recruit the U2 snRNP to the BP (reviewed in Chou et al. 1999; De Conti et al. 2013; Matera and Wang 2014). This 5' and 3'ss recognition results in the bulging of adenosine and enabling its 2'OH group to initiate the first catalytic step of splicing. This spliceosome complex is called pre-spliceosome (complex A) (**Figure 4**).

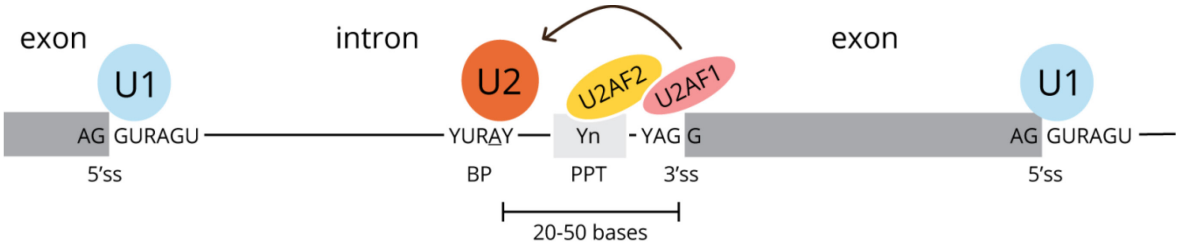

Figure 3. Recognition of intron boundaries. 5'and 3'ss – 5'and 3'splice sites, BP – branch point, subscribed A represents the base mediating branching of intron lariat, PPT – polypyrimidine tract, R – purine, Y – pyrimidine, U1 – U1 snRNP, U2 – U2 snRNP, U2AF1/2 – U2 snRNP auxiliary factor 1/2.

In the next step, pre-catalytic spliceosome (complex B) is created. Pre-assembled U4-U6.U5 tri-snRNP particle associates with the complex. Then conformational changes are initiated by RNA helicase Brr2 resulting in new extensive base pairing of U2 and U6 snRNAs, and the pairing of U5 snRNA with exonic sequences near the 5'ss. These compositional and conformational changes lead to dissociation of U4 snRNA from U6 which can subsequently bind to 5'ss and displace U1 snRNA. All these rearrangements

bring together 5'ss and BP. After the dissociation of U1 and U4 snRNPs from spliceosome, the spliceosome consisting of only U2, U5 and U6 snRNPs is activated (complex B^{act}). However, it has to be catalytically activated (complex B^{*}) to catalyze the first *trans*-esterification reaction of the splicing process by additional rearrangements. To mediate the catalytic event itself, the formation of intramolecular stem-loop in U6 snRNA is needed (Matera and Wang 2014). This leads to the forming of the C complex which then catalyzes the second reaction. In this step, the interaction of U6 snRNP and 5'ss is disrupted (Konarska et al. 2006), and U5 snRNP mediates juxtaposing the 5'and 3'exons. After the second step, post-spliceosomal complex is formed. The remaining U snRNPs are dissociated, and intron is released in the form of the lariat. For another round of splicing, spliceosome has to be assembled *de novo*, and spliceosome components are recycled by extensive remodeling.

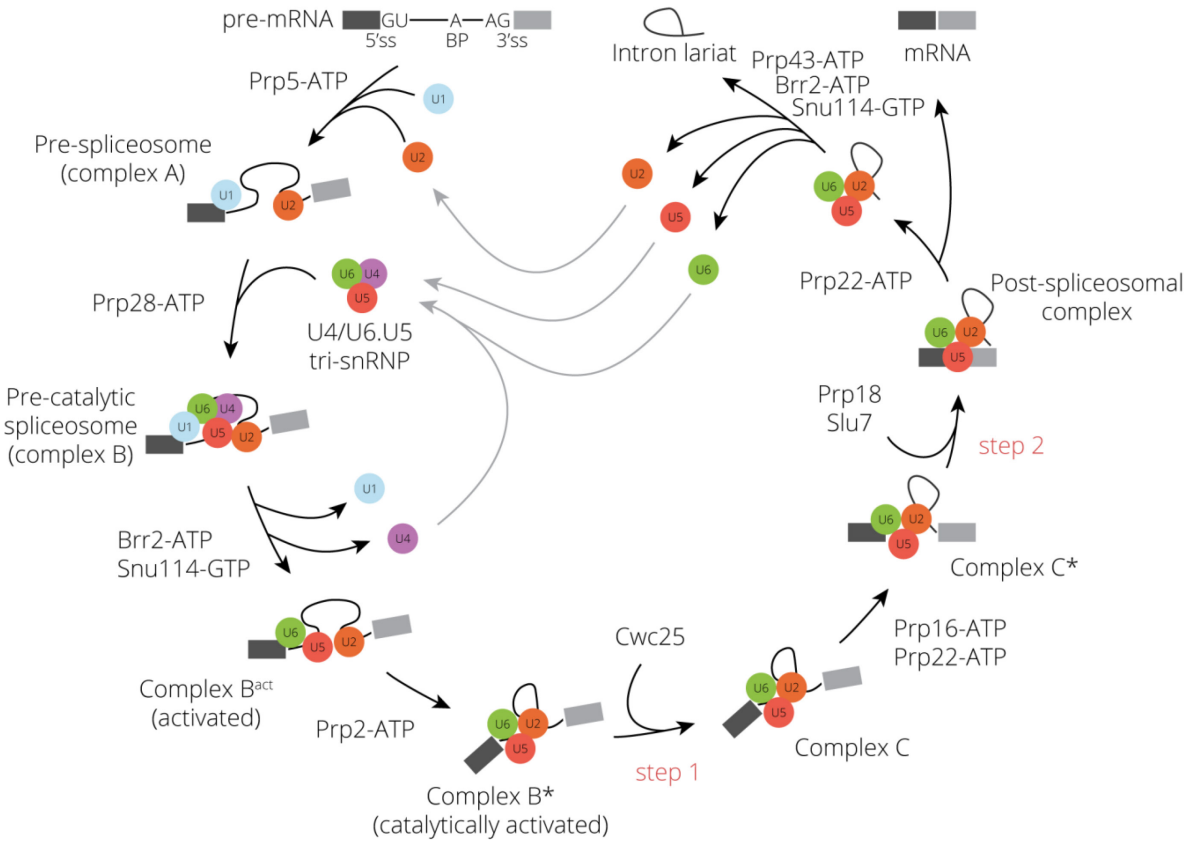

Figure 4. Splicing process of major spliceosome in metazoans. U snRNPs are depicted as colored bubbles. Exon and intron sequences are indicated by boxes and lines, respectively. Multiple DExH/D-box RNA ATPases/helicases needed throughout splicing are indicated (adapted from Will and Lührmann 2011).

The two splicing *trans*-esterification reactions are mediated by the network of RNA-RNA interactions between two snRNAs or snRNAs and particular sequences of pre-mRNA. Throughout the splicing, several rearrangements of these RNA-RNA interactions are broken and formed to finally create the catalytic center of the spliceosome (**Figure 5**). Because this catalytic center is composed of snRNAs (Fica et al. 2013), spliceosome is a ribozyme (reviewed in Lee and Rio 2015). Even though, the catalytic activity is mediated solely by RNA there are multiple proteins involved, which catalyze the remodeling of snRNAs. The most common are ATPases from the DEAH-box, DEAD-box and Ski2-like helicase families that can act at multiple stages of the splicing reaction (**Figure 3**) (Cordin et al. 2012). The vast majority of energy used during splicing is utilized by these ATPases since only minimal energy is required for the *trans*-esterification reactions. This elevated uptake of energy in the form of ATP by these helicases increases the fidelity of splicing reactions while decreasing the time needed for splicing reactions to be done.

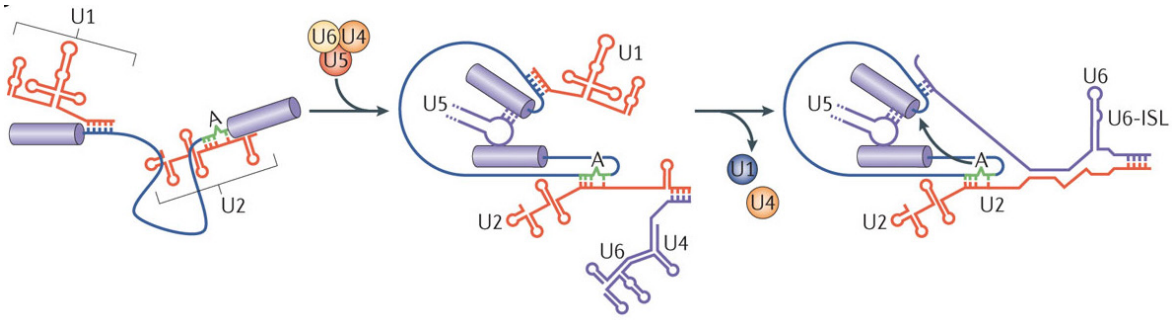

Figure 5. Extensive conformational rearrangements during splicing. During spliceosome assembly, there are several alterations in RNA-RNA base pairing including disruption of U1 snRNA binding to 5'ss replaced by U6 snRNA. To create a catalytic center of the spliceosome, U4-U6 snRNA binding has to be disrupted to allow to form an internal loop (U6-ISL) within U6 snRNA creating a metal-binding platform. U6 snRNA together with U2 snRNA forms the active site of the spliceosome. Because this step is remarkably similar to the formation of self-splicing group II introns, the evolutionary model of spliceosomal splicing originating from these self-splicing introns has been proposed (taken from Matera and Wang 2014).

Regulation of Splicing

Because sequences resembling splice sites (aberrant or cryptic splice sites) are relatively abundant throughout the transcribed regions (Nakai and Sakamoto 1994; Roca et al. 2003; Královicová et al. 2005), the recognition of the correct exon boundaries is a crucial step during the splicing process. High fidelity of splice site recognition is mediated throughout a network of interactions of *cis*-acting elements localized in pre-mRNA and *trans*-acting factors recognizing these elements. The most conserved interactions include snRNA base-pairing with sequences around 5'and 3'ss and the binding of numerous splicing regulatory proteins such as U2AFs to BP and PPT (**Figure 3**) (reviewed in De Conti et al. 2013). However, in many cases, these basic splicing signals are not strong enough to ensure their efficient recognition. Therefore, there are several additional *cis*-acting conserved sequences called splicing regulatory elements (SREs). They can be localized to either exon or intron and can either enhance or silence the splicing (**Figure 6**). Typically, they consist of hexamer motifs which are bound by proteins that can stimulate or repress the spliceosome assembly onto exon-intron boundaries (reviewed in Wang and Burge 2008).

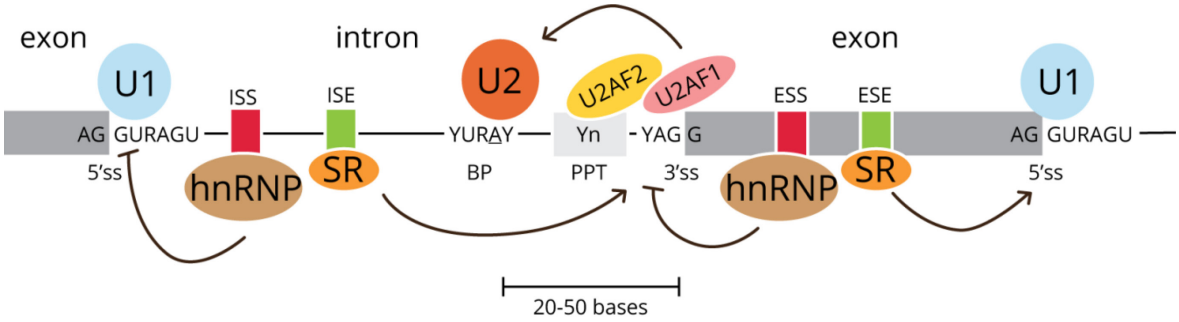

Figure 6. The interaction network during recognition of introns. 5'and 3'ss – 5'and 3'splice sites, BP – branch point, subscribed A represents the base mediating branching of intron lariat, PPT – polypyrimidine tract, ISS/ISE – intron splicing silencer/enhancer, ESS/ESE – exon splicing silencer/enhancer, R – purine, Y – pyrimidine, U1 – U1 snRNP, U2 – U2 snRNP, U2AF1/2 – U2 snRNP auxiliary factor 1/2, hnRNP – hnRNP proteins, SR – SR proteins.

The best studied *trans*-factors binding to SREs are SR (serine-arginine) proteins and hnRNP (heterogeneous nuclear ribonucleoproteins) proteins (reviewed in Busch and Hertel 2012). Generally, SR proteins promote splicing, while hnRNPs most frequently act as splicing silencers. However, there were also showed cases when hnRNP proteins function as splicing enhancers (Schaub et al. 2007; Wang et al. 2011a; Wang et al. 2012).

Moreover, the activities of these regulatory elements are often context-dependent, and they can activate or repress splicing according to their location within the transcript (Ule et al. 2006; Wang and Burge 2008; Erkelenz et al. 2013; Fu and Ares Jr 2014; Brillen et al. 2017; Rot et al. 2017). It was shown, that SR proteins bound to exonic regions stimulate splicing. However, intronic SR binding suppresses splicing (Kanopka et al. 1996; Ibrahim et al. 2005; Hicks et al. 2010; Giudice and Cooper 2014). Nevertheless, one of the key steps in the initial stages of intron recognition and spliceosome assembly is the correct positioning of snRNPs on a pre-mRNA through cooperative interactions with non-snRNP factors.

SR Proteins

SR proteins are a family of approximately twelve evolutionary conserved polypeptides (SRSF1-12 - serine/arginine-rich splicing factor 1–12) (reviewed in Graveley 2000; Long and Caceres 2009; Busch and Hertel 2012). All twelve prototypical SR proteins likely have a single ancient origin among RNA recognition motifs-containing proteins (Califice et al. 2012). They all have a common structure including one or two RNA recognition motifs (RRMs) and one or two arginine-serine (RS) domains (**Figure 7**). RRMs in the N-terminal regions are important for RNA-binding specificity (Wu and Maniatis 1993; Kohtz et al. 1994). RS-domains that are rich in repetitive arginine-serine dipeptides of variable lengths are located at C-termini. They can promote splicing via protein-protein interactions that facilitate the recruitment of spliceosomal components (Wu and Maniatis 1993; Kohtz et al. 1994) such as recruitment and stabilization of interaction between U1 snRNP at 5'ss and U2AF2 at 3'ss (Graveley et al. 2001). Additionally, some SR proteins can prevent exon skipping by acting as a barrier to ensure the correct order of exons in spliced mRNA (Ibrahim et al. 2005) or they can inhibit the negative role of hnRNP proteins (Zhu et al. 2001). Interestingly, RS domains were also shown to interact with pre-mRNA directly via BP and 5'ss (Shen et al. 2004; Shen and Green 2006). Serines in RS domains undergo extensive phosphorylation and dephosphorylation cycles. Such changing of phosphorylation status was shown to be important for regulation of activity and subcellular localization of SR proteins and protein-protein interactions (reviewed in Lin and Fu 2007). Interestingly, in some cases, phosphorylation leads to splicing promotion (Xiao and Manley 1997), whereas in some cases, splicing is catalyzed by the dephosphorylated SR proteins (Tazi et al. 1993; Cao et al. 1997). Furthermore, RS domain acts also as nuclear

localization signal (NLS) that can affect the subcellular localization of SR proteins (Cáceres et al. 1997).

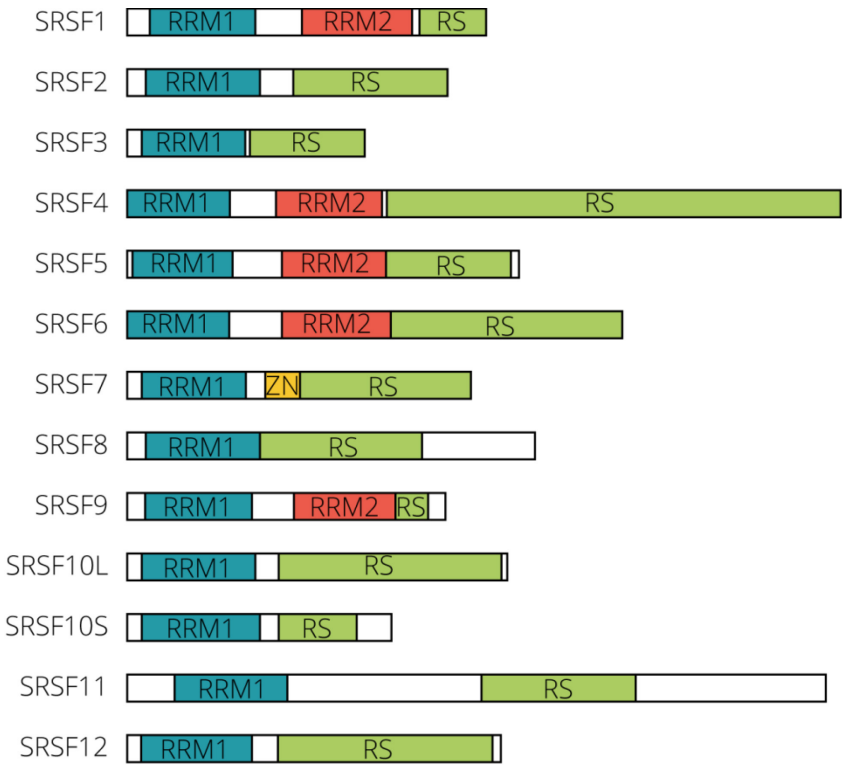

Figure 7. Domain structure of human SR proteins. All 12 members of the canonical SR protein splicing family contain one or two N-terminal RNA recognition motifs (RRMs) followed by a downstream arginine/serine (RS) domain. Additionally, SRSF7 contains also zinc-finger (ZN) domain. The RRM is responsible for RNA binding, while the RS domain mediates protein/protein interactions (adapted from Mahiet and Swanson 2016).

Target sequences of SR proteins were widely studied throughout the years by various approaches, e.g. SELEX (selected evolution of ligands through exponential enrichment) techniques (Tacke and Manley 1995; Tacke et al. 1997; Cavaloc et al. 1999), computational approaches such as RESCUE - (relative enhancer and silencer classification by unanimous enrichment ESE predictions) (Fairbrother et al. 2002), ESEfinder (Cartegni et al. 2003) and others (Akerman et al. 2009; Paz et al. 2010), ChIP (chromatin immunoprecipitation) and RIP (RNA immunoprecipitation) (Niranjanakumari et al. 2002) assays, and recently by CLIP (cross-linking and immunoprecipitation) (Ule et al. 2006; Sanford et al. 2008) method and its derivates (Änkö et al. 2012; Müller-McNicoll et al. 2016). Even though there is the different target-site preference of different SR proteins and different SR proteins have specific functions, it seems there is, at least partial, redundancy between some SR proteins in different species. Moreover, the few SR proteins with

essential functions appear to be tissue- or developmental stage-specific (reviewed in Long and Caceres 2009). And on top of that, several cases showed that specificity of binding of some SR proteins *in vivo* is not fully dependent only on sequence recognition and RS domain but also other factors and RNA binding domains (RBDs) are involved (Tacke and Manley 1995; Liu et al. 1998).

In addition to the relatively homogenous family of twelve SR proteins, there are a large number of RS domain-containing proteins also known as SR-like or SR-related proteins. They can contain but it is not obligatory RRM and have a role in splicing regulation, as well as in chromatin remodeling, transcription and cell cycle (Boucher et al. 2001). Altogether, it was proposed that any protein containing at least one RRM and RS domain irrespective of their positions within the protein and functioning in splicing regulation belongs to SR proteins (Lin and Fu 2007). Into this category belongs for example proteins such as U2AF1 and U2AF2, U1-70K and others (Blencowe et al. 1999).

HnRNP Proteins

HnRNPs are proteins originally described as a group of abundant proteins associated with heterogeneous nuclear RNAs (hnRNAs) present in a high concentration within the nucleus (nearly as abundant as histones) (reviewed in Martínez-Contreras et al. 2007). Currently, mammalian hnRNP proteins represent a heterogenous set of ~20 distinct proteins of different structures and functions, associating with nascent pre-mRNAs as an only common feature (Piñol-Roma et al. 1988). Although hnRNPs were shown to play a role in telomere biogenesis, polyadenylation, translation, RNA editing and mRNA stability (reviewed in Krecic and Swanson 1999), the most studied function is their regulation of both constitutive and alternative splicing (Wang and Burge 2008). Similarly to SR proteins, hnRNPs are modular containing two or more RRMs or RRM-related domains (qRRMs – quasi-RRMs or KH – K homology motif) which are complemented with regions rich mainly in arginine and glycine. The mammalian hnRNP proteins are grouped into 10 families according to their global protein identity including sequence and RRM similarity, cellular localization, and RNA binding site (**Figure 8**). Moreover, multicellular organisms contain more families and family members of hnRNPs than unicellular organisms which was proposed to be a result of successive gene duplication events (Busch and Hertel 2012).

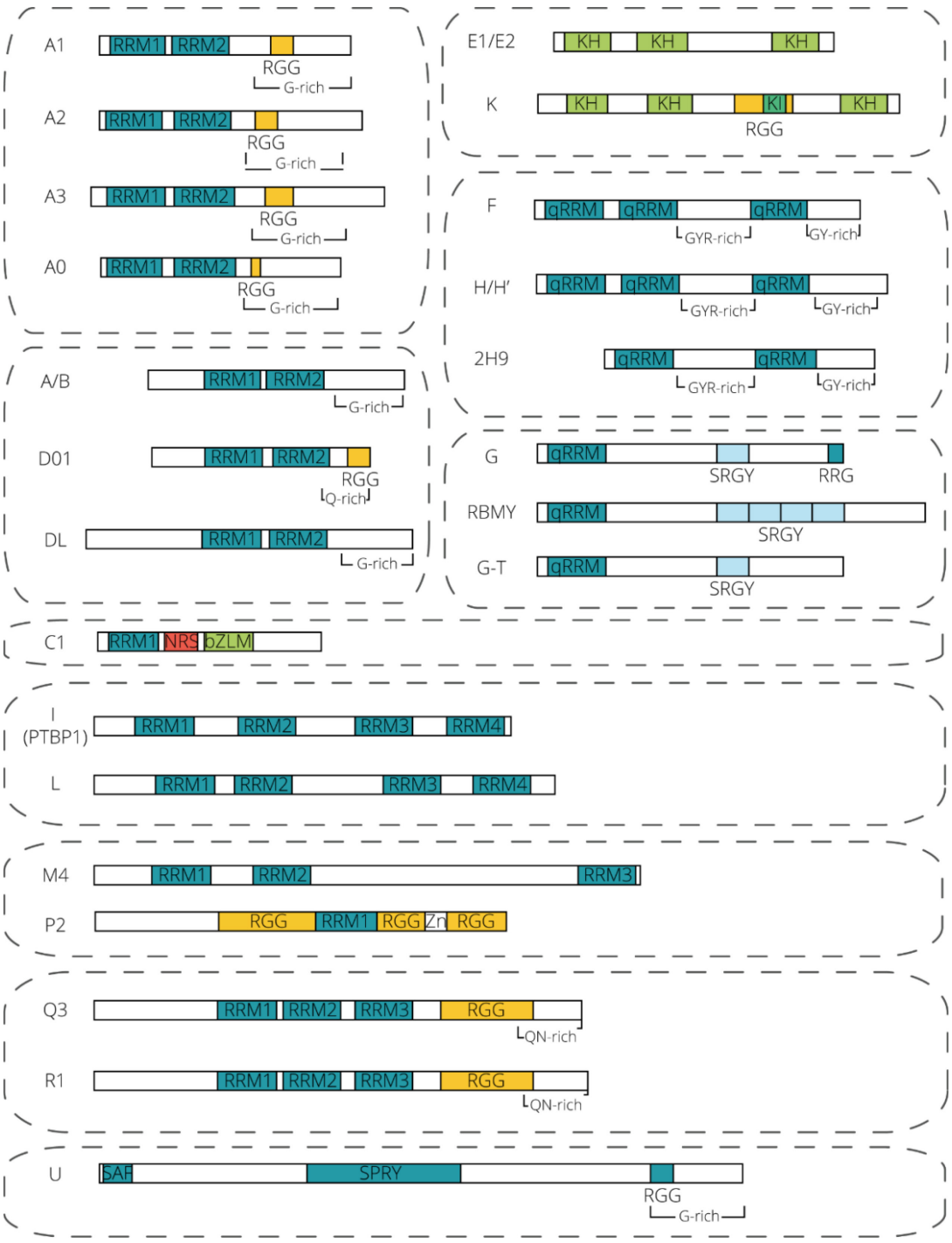

Figure 8. The mammalian hnRNP proteins are grouped into 10 families. The type of RNA recognition motif (RRM, quasi (q)RRM or KH) is indicated as well as the presence of auxiliary domains, such as: G - the glycine-rich domain, bZLM - basic leucine zipper (bZIP)-like motif, NRS - nuclear retention signal, RGG – a region rich in arginine and glycine, GYR - glycine, tyrosine and arginine-rich domain, GY- glycine- and tyrosine-rich domain, SRGY - motif enriched in serine, arginine, glycine and tyrosine (also found in some SR proteins), Zn - RNA binding seems to be mediated by the zinc finger domain, QN - glutamine- and asparagine-rich domain, SPRY- SP1a and ryanodine receptor (SPRY) homology domain of unknown function, SAF - scaffold-associated region (SAR)-specific bipartite DNA binding domain capable of binding specific DNA sequences (reviewed in Martínez-Contreras et al. 2007).

The mode of function of hnRNPs in splicing regulation varies between specific hnRNP proteins. However, there are some similar principles that can be applied. As it was mentioned earlier, even though there are few cases when hnRNPs act as splicing enhancers (Schaub et al. 2007; Wang et al. 2011a; Wang et al. 2012), more often hnRNPs bind to pre-mRNAs to sterically occlude the interaction of components of splicing machinery to splice sites (Tange et al. 2001; House and Lynch 2006). It was shown that hnRNP A1 interfere with the binding of SR proteins, hnRNP H with the binding of U1 snRNP, and hnRNP I (PTBP1 - polypyrimidine tract binding protein 1) with U2AF binding (reviewed in Martínez-Contreras et al. 2007). Finally, hnRNPs can influence splicing regulation co-transcriptionally via binding TBP (TATA-binding protein) at promoter sequences (Moumen et al. 2005; Wei et al. 2006), binding chromatin remodeling complexes (Mahajan et al. 2005), or associate with transcription factors (Mattern et al. 1997) and phosphorylated C-terminal domain (CTD) of RNA polymerase II (Hager et al. 2004).

In addition, as it was shown for SR proteins, hnRNPs can be phosphorylated and can shuttle between the nucleus and the cytoplasm (Piñol-Roma and Dreyfuss 1992; Matthew Michael et al. 1995). However, the specific impact of such modification on splicing regulation has not been studied to a greater extent.

Alternative Splicing

Alternative splicing is characterized as a diverse usage of splice sites resulting in the formation of several different mRNA isoforms (**Figure 9**) (reviewed in Lee and Rio 2015). As it was discussed above, the splice site recognition is mediated by the base-pairing of U1 snRNA with 5'ss, and U2 snRNA with BP region located near 3'ss. Such bindings are enabled through sequence complementarity. However, in metazoans, the splice sites are degenerated to some extent (Sheth et al. 2006) resulting in the certain flexibility of splice site recognition which can lead to alternative splicing. It was estimated that humans have more than 9,000 5'ss sequence variants (Roca et al. 2012). However, the process of exon-intron boundary recognition seems to be much more complicated. There were found 5'ss sequences that matched 5'ss consensus as well as or better than the real one but were not used (termed as pseudo-5'ss) (Roca et al. 2013). Additionally, there are 5'ss that resemble the 5'ss consensus but are used only when real ones are inactivated (termed as cryptic 5'ss) (Wieringa et al. 1983) together suggesting that 5'ss sequence seems not to be the only

determinant of use of such 5'ss as real splice site. Finally, there are cases when two 5'ss are in the competition for use (termed as alternative 5'ss) depending on the sequences of such sites (Montell et al. 1982). Therefore, SREs mentioned above bound by SR proteins or hnRNPs are necessary to provide additional information which splice site should be used.

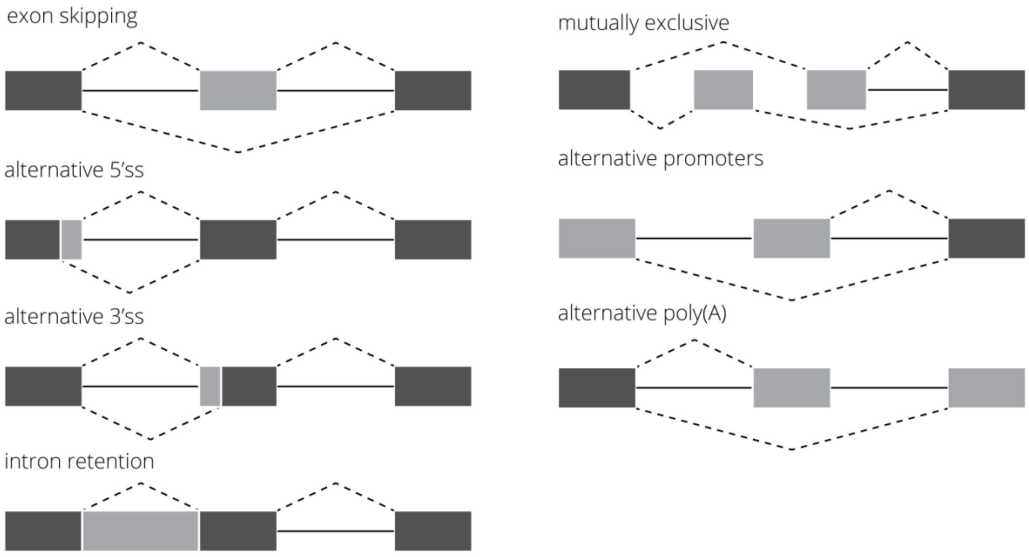

Figure 9. Seven major types of alternative splicing events. The prevalence of the specific type of alternative splicing varies among species. Exon skipping is more abundant gradually with eukaryotic tree. On the other hand, intron retention is most common in lower metazoans, fungi and protozoa. Therefore, alternative 5'ss and 3'ss as a subfamily of exon skipping are considered to be an intermediate evolutionary stage. Interestingly, alternative splicing seems do not have a substantial role in plants because of its low level in these organisms. Plants are probably using whole-genome duplication to enhance their transcriptomic and proteomic diversity. Constitutive exons are shown in dark gray and alternatively spliced regions in light gray, introns are represented by solid lines, and dashed lines indicate splicing activities (adapted from Ast 2004; Keren et al. 2010).

There are additional characteristics that distinguish alternative exons from constitutive ones. Many conserved alternative exons (77%) are flanked by introns containing long (~100 nt) conserved intronic sequences that are rarely found in constitutive exons (17%) (Sorek and Ast 2003). This suggests that there might be a regulatory mechanism common to many alternative spliced exons using flanking conserved intronic sequences. Since their length usually exceeds 50 nt, they can regulate splicing through multiple factors including transcription machinery and chromatin remodeling complex, or even by the formation of secondary structures. In addition, conserved alternative exons tend to be smaller, and their length is divisible by three otherwise they would disrupt the

reading frame (Resch et al. 2004; Sorek et al. 2004b). However, both alternative and constitutive genes have a similar amino-acid usage (Zhuang et al. 2003).

In addition to sequences, alternative usage of splice sites can also be controlled by RNA secondary, and tertiary structures (Buratti and Baralle 2004; McManus and Graveley 2011) since nascent pre-mRNAs synthesized by RNA polymerase II are folded and spliced co-transcriptionally (see below). One mode of action is when the formed secondary structure makes a sterical constraint for binding of splicing regulatory proteins (such as SR proteins or hnRNP proteins) (Buratti et al. 2004) or suppress pseudo splice sites recognition (Buratti et al. 2007). One of the best studied examples of RNA secondary structure's importance on alternative splicing is the *DSCAM* (Down syndrome cell adhesion molecule) gene in *Drosophila* which can combinatorically generate more than 36,000 distinct spliced variants (Graveley 2005) from four clusters of alternative exons used in a mutually exclusive manner (Olson et al. 2007). Furthermore, riboswitches represent another mechanism how to modify splicing. They are RNA aptamers that regulate the expression of numerous bacterial metabolic genes in response to small molecule ligands in the absence of proteins (Nudler and Mironov 2004; Tucker and Breaker 2005). The most characterized representative of such regulation is an intron of *NMT1* gene of a eukaryote, *Neurospora crassa* (Cheah et al. 2007) which employs the metabolite (thiamine pyrophosphate – TPP) to alter the availability of alternative splice site and the BP, and thus modify usage of splice sites.

The definite origin of alternative splicing is not known. However, early eukaryotic ancestors seem to have high intron densities in their genes (Fedorov et al. 2002; Rogozin et al. 2003; Nguyen et al. 2005; Roy and Gilbert 2005; Carmel et al. 2007; Csűrös et al. 2008; Roy and Irimia 2009) suggesting they were similar to mammalian eukaryotes. This is also supported by the presence of complex spliceosome in phylogenetically older eukaryotes (Fedorov et al. 2002; Rogozin et al. 2003; Roy and Irimia 2009), the presence of non-sense mediated decay (NMD) in animals, fungi, plants, excavates and chromalveolates (Jaillon et al. 2008; Kerényi et al. 2008; Roy and Irimia 2009), and the homology of splicing factors in different species (Barbosa-Morais et al. 2006; Plass et al. 2008; Schwartz et al. 2008; Gal-Mark et al. 2009). So the alternative splicing could have existed early in a eukaryotic lineage which is supported by the similar alternative splicing pattern among 12 analyzed eukaryotic genomes (Irimia et al. 2007).

There are currently three different evolutionary mechanisms how the alternative spliced exons could have evolved: exonization of introns, exon shuffling, and transition of a constitutive exon to an alternative one (reviewed in Keren et al. 2010). Exonization of introns is a process observed to occur in various species including humans (Makałowski et al. 1994; Nekrutenko and Li 2001; Sorek et al. 2002), and other vertebrates (Wang and Kirkness 2005; Wang et al. 2005; Alekseyenko et al. 2007), *Drosophila* (Kandul and Noor 2009) and other insects (Kent and Robertson 2009), and plants (Fu et al. 2009). Almost half of the human genome is derived from transposable elements (International Human Genome Sequencing et al. 2001), with *Alu* elements as the most abundant (>10% of the human genome) (International Human Genome Sequencing et al. 2001; Sela et al. 2007). About 4% of human genes contain transposable elements in their coding regions suggesting an exonization mechanism (Nekrutenko and Li 2001; Sorek et al. 2002). *Alu* elements are short (~300 nt), primate-specific retrotransposons mainly located in introns containing right and left arm joined by an A-rich linker and a poly(A) tail-like region (Lev-Maor et al. 2003; Sorek et al. 2004a). They belong to long terminal repeat (LTR) group of transposable elements with the mechanism of copying themselves within genome via an RNA intermediate. At first, they are transcribed by RNA polymerase II followed by reverse transcription and then they can be re-integrated into the host genome, usually in the antisense orientation. So upon such integration into intron sequences, poly(A) tract in the antisense orientation creates a strong PPT. To create a new exon, a few mutations, usually in the right arm (Sorek et al. 2002; Sela et al. 2007), have to occur to create a functional 5'ss and 3'ss (Lev-Maor et al. 2003; Sorek et al. 2004a; Krull et al. 2005). 3'ss is selected downstream of newly developed PPT, while 5'ss is recognized further downstream (~120 nt) (Gal-Mark et al. 2009). Alternatively, new splice sites can also be created by RNA editing mechanism using ADAR (adenosine deaminase acting on RNA) enzyme. In this case, the two adjacent *Alu* elements in the opposite orientations create a dsRNA secondary structure that serves as a template for the ADAR enzyme which converts adenosine to inosine by deamination. By most biological machinery, inosine is recognized as guanosine (Bass 2002) since both are structurally similar. This can then result in the creation of a functional 3'ss (conversion of AA to AG) (Athanasiadis et al. 2004).

Additionally, it was suggested that transposable elements also considered as "junk DNA sequences" can acquire a novel function in the genome (reviewed in de Souza et al. 2013; Göke and Ng 2016; Thompson et al. 2016) by providing new promoters, enhancers,

and chromatin barriers (Medstrand et al. 2005; Franke et al. 2017). It was even proposed that transposable element insertions are responsible for an extremely rapid transcriptional rewiring in vertebrates (Feschotte 2008; Bourque 2009; Franke et al. 2017). Indeed, it was shown that more than 5% of the alternatively spliced internal exons in human are derived from *Alu* elements (Sorek et al. 2002). Finally, the only selective pressure for the intronic element to become an alternative exon is the creation of weak splice site flanking the alternatively spliced *Alu* exon (Sorek et al. 2002; Lev-Maor et al. 2003; Sorek et al. 2004a).

Another mechanism of creation of alternatively spliced exons is exon shuffling. It is a process of insertion of a new exon into an existing gene or duplication of an exon within the same gene resulting in the creation of new chimeric proteins that could be evolutionary advantageous (Gilbert 1978; Kondrashov and Koonin 2001; Kondrashov and Koonin 2003). The mechanism includes intronic illegitimate recombination (IR) between two non-homologous sequences or between short homologous sequences that induce genomic rearrangements (van Rijk and Bloemendal 2003; Babushok et al. 2007). It is thought that over 30% of such recombination occurs through crossovers between *Alu* elements (Babushok et al. 2007). Consistent with exon shuffling theory is an observed correlation between borders of exons and protein domains (Doolittle 1995; Kolkman and Stemmer 2001; Liu and Grigoriev 2004) which includes the insertion of introns at positions that correspond to the boundaries of a protein domain. This means that introns are at the same phase (position relative to the reading frame) at both their 5'ss and 3'ss otherwise it would result in a shift in the reading frame and thus protein disruption (Patthy 1987; Patthy 1996; Kolkman and Stemmer 2001). Moreover, this border-domain correlation is stronger in the more complex eukaryotes suggesting that in eukaryotes, exon shuffling has contributed substantially to the complexity of their proteome. It was suggested that exon shuffling became important mainly for multicellular organisms since most proteins believed to be created by this mechanism are associated with multicellularity (such as the extracellular membrane-associated proteins) (Patthy 1999). Also, exon shuffling is considered to contribute to the rapid metazoan radiation (Patthy 1999). One of the known examples of such exon duplication is above mentioned *DSCAM* gene in *Drosophila* with extensive, mutually exclusive alternative splicing (Schmucker et al. 2000; Graveley 2005). Furthermore, it was proposed that alternative splicing precedes tandem duplication rather than it is propagated by it (Peng and Li 2009) since the newly duplicated

exons tend to preserve the splicing status of their original exons (Letunic et al. 2002). Tandem duplication was found to date back to the radiation of vertebrate classes (Kondrashov and Koonin 2001) and it was shown that ~10% of the genes in humans, flies and worms contain tandemly duplicated exons of which 60% was likely to be alternatively spliced in a mutually exclusive manner (Letunic et al. 2002).

The last proposed evolutionary mechanism of creating alternatively spliced exons is the transition process. On the contrary to the first two mechanisms when alternatively spliced exons are generated *de novo*, during transition alternative exons are derived from constitutive ones. There are two ways how a constitutive exon can become an alternative exon. First option includes accumulation of mutations in splice sites (reduction in the affinity of U1 snRNA binding to 5'ss) or in splicing regulatory elements (disruption of splicing enhancers or creating splicing silencers) leading to the suboptimal recognition of the exon which result in an exon skipping (Lev-Maor et al. 2007; Ke et al. 2008). This is supported by the findings that alternative splice sites are weaker than constitutive ones (Lear et al. 1990; Stamm et al. 1994; Carmel et al. 2004; Sorek et al. 2004a). Alternatively, two *Alu* elements in opposite orientation to each other (one *de novo* inserted into an intron in close proximity to the adjacent exon) can form dsRNA secondary structure. Such double-stranded region can act as intronic splicing silencer and thus influence the downstream exon recognition leading to changes in splicing outcome, usually to exon skipping (Mola et al. 2007; Lev-Maor et al. 2008; Tappino et al. 2008). This mechanism is also supported by the finding that introns flanking alternatively spliced exons tend to contain more *Alu* sequences than constitutively spliced ones. Similarly, this is also true for exons that have changed their mode of splicing from constitutive to alternative during human evolution (Lev-Maor et al. 2008).

One of the proposed functions of alternative splicing is to significantly expand proteome taken that in human there are ~20,000 protein-coding genes (PCGs)[1] but the number of protein-coding transcripts is estimated to be more than 83,000[1] (Modrek and Lee 2002; Nilsen and Graveley 2010; Irimia and Blencowe 2012; Braunschweig et al. 2013). However, a new mRNA isoform does not have to lead only to the generation of a novel protein, but it can have a regulatory role. In this case, it can balance the levels of mRNAs that produce functional proteins and mRNAs producing non-functional proteins.

[1] GENCODE Release (version 29). Retrieved November 11, 2018, from https://www.gencodegenes.org/human/stats.html

Additionally, the new splicing isoform does not have to have a function right from the beginning, it can be the result of the merely stochastic noise of splicing machinery (Melamud and Moult 2009). If it had a deleterious effect, it would be eliminated through purifying selection. However, if it does not harm, it can be tolerated because of its low abundance in the cell. With time, it might accumulate additional mutations and acquire a new function. Indeed, exons of low inclusion level are associated with increased evolutionary changes (Modrek and Lee 2003). Furthermore, it was observed that more morphologically and behaviorally complex organisms have a similar number of genes as low-complex metazoan species (Lee and Rio 2015). The increased role and extent of alternative splicing in the more complex organisms could explain the imbalance between a number of genes and organismal complexity (Kim et al. 2004; Kim et al. 2007). Only 25% of multiexonic PCGs of *Caenorhabditis elegans* undergo alternative splicing, whereas in *Drosophila* it is up to 45% and in mice 63%, in humans even more than 88% (Lee and Rio 2015). It is currently estimated that more than 95% of human genes generate at least two alternative mRNA isoforms (Pan et al. 2008; Wang et al. 2008).

Exon versus Intron Definition

The recognition of *bona fide* splice sites among numerous pseudo sites by spliceosome is not always straightforward. One of the major roles in such recognition plays the relative strength of splice sites meaning how close the sequence is to the consensus one. However, there are exons and nearby intron sequences (~50 nt from the exon boundaries) (Zhang et al. 2005) containing many regulatory elements providing additional information about the position of real splice boundaries. Even though there is no mechanistic difference in the spliceosomal complex assembly, two general modes how the spliceosome separate exons from introns exist: intron definition and exon definition (**Figure 10**) (reviewed in Ast 2004; De Conti et al. 2013). During the intron definition, spliceosome complex is assembled over the intron sequences. However, there is an upper limit for the length of introns recognized by this mechanism, so these introns are evolutionarily constrained to remain short. Introns of phylogenetically older eukaryotes are generally short (~100 nt) (Hawkins 1988; Goguel and Rosbash 1993), so intron definition is probably their dominant mode of splicing (Berget 1995; Xiao et al. 2007). This was confirmed when small introns in yeast and *Drosophila* were experimentally extended which lead to splicing defects such as intron retention (Guo et al. 1993; Talerico and Berget 1994). Moreover, *Drosophila*

exons flanked by very long introns have been observed to preferentially undergo alternative splicing compared to exons flanked by short introns (Fox-Walsh et al. 2005) suggesting that predominant mode of spliceosome formation in *Drosophila* is occurring over introns.

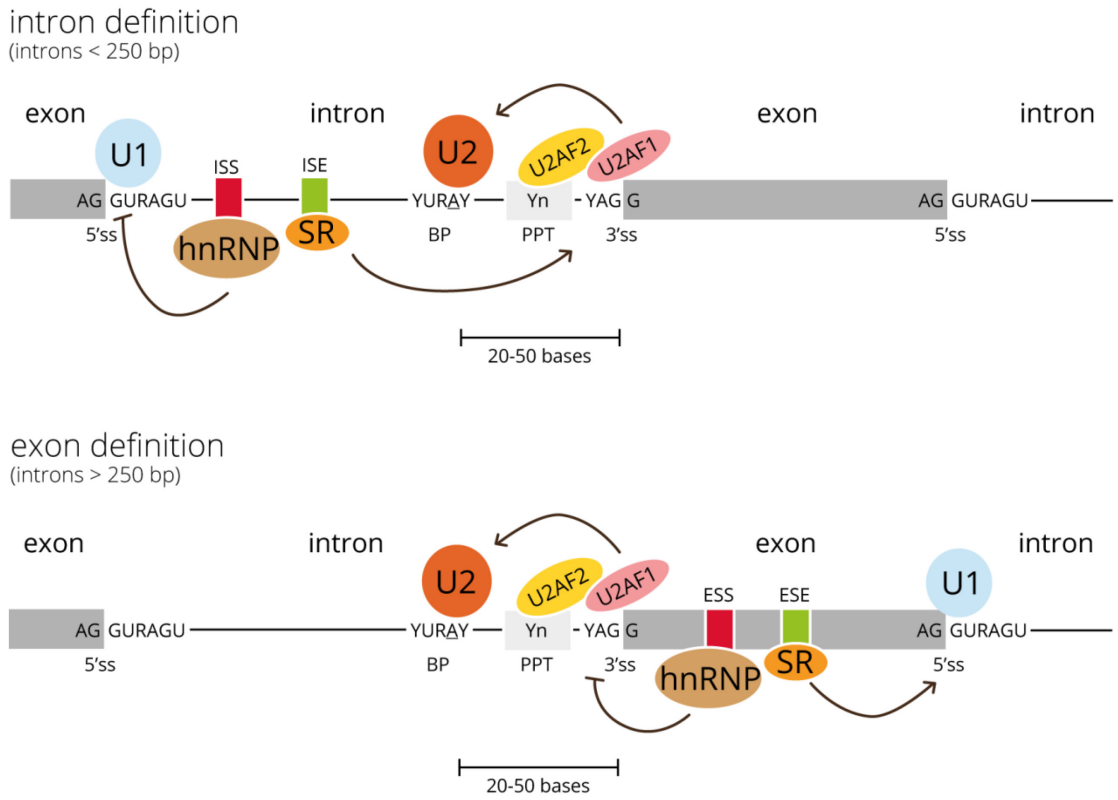

Figure 10. Exon and intron definition models. During intron definition, spliceosome complex is assembled primarily over intron sequences with the help of mainly intronic splicing regulators since introns are usually shorter than 250 bp. On the other hand, in exon definition, the spliceosome assembly occurs over exon sequences with the help of mainly exon splicing regulators since introns are usually several hundred or thousand nucleotides long. It should be mentioned that this is oversimplified version; the real exon-intron boundary recognition is usually much more complicated and regulated by a cooperated network of numerous interactions. 5'and 3'ss – 5'and 3'splice sites, BP – branch point, subscribed A represents the base mediating branching of intron lariat, PPT – polypyrimidine tract, ISS/ISE – intron splicing silencer/enhancer, ESS/ESE – exon splicing silencer/enhancer, R – purine, Y – pyrimidine.

On the other hand, the exon definition mode (Robberson et al. 1990; Berget 1995) is more pronounced in phylogenetically younger eukaryotes (especially in mammals) (Xiao et al. 2007) since their introns are usually longer (from several hundred to several thousand nucleotides) (Zhang 1998; Deutsch and Long 1999; Sakharkar et al. 2005), splicing machinery had to adapt to be able to recognize short exons (average size ~185 nt) (Lagarde

et al. 2017) among long introns. One of the possible signals to identify short exons among long introns is their higher GC content relative to their flanking introns (Schwartz et al. 2009). However, such a mechanism seems to be unnecessary in the case of exons flanked by short introns in the case of intron definition (Amit et al. 2012). Additionally, it is proposed that it is easier to juxtapose short exons instead very long introns (Robberson et al. 1990; Fox-Walsh et al. 2005). There seems to be lower and upper limit for the length of exons since the spliceosome formation was strongly inhibited if exon lengths were expanded artificially to over 300 nt (Sterner et al. 1996) or reduced to less than 50 nt (Dominski and Kole 1991). Experimentally was exon definition mode of spliceosome assembly supported when *in vitro* splicing efficiency was enhanced upon 5'ss addition to its 3'end (Robberson et al. 1990). Additionally, it was observed that in humans the activation of alternative splice sites is less likely if the flanking introns are long (Fox-Walsh et al. 2005) suggesting the predominant mode of splicing to be the exon definition.

Co-transcriptional Splicing and Processing

The splicing of many pre-mRNAs occurs during transcription on nascent transcripts while are still connected to their genomic DNA via transcriptional machinery (**Figure 11**) (Lacadie et al. 2006; Pandya-Jones and Black 2009; Perales and Bentley 2009; Carrillo Oesterreich et al. 2010; Khodor et al. 2011; Bhatt et al. 2012; Khodor et al. 2012; Rodriguez et al. 2012; Brugiolo et al. 2013; Pandya-Jones et al. 2013). One of the first experiments suggesting that splicing is coupled with transcription was the electron microscopy observation of RNA loops in the nascent transcripts in *Drosophila* embryos (Beyer and Osheim 1988) followed by the different splicing outcome for a gene spliced *in vivo* (during ongoing transcription) or *in vitro* (on a premade pre-mRNA template) (Eperon et al. 1988). Later, it has been observed that different RNA polymerase II promoters placed upstream of the same gene influence its splicing (Cramer et al. 1997; Cramer et al. 1999) as well as mutations in promoter sequence were shown to alter splicing (Dušková et al. 2014). An additional indication of co-transcriptional splicing was finding that snRNPs are recruited to actively transcribed genes (Görnemann et al. 2005; Lacadie and Rosbash 2005; Listerman et al. 2006). Even though, not all the pre-mRNAs seem to be spliced before the gene is finished transcribing (Girard et al. 2012; Khodor et al. 2012), and such transcripts stay associated with chromatin at the gene locus until all introns have been removed (Bhatt et al. 2012), which is especially true for long genes with large introns, most introns are

spliced in the order of their transcription. In addition, some spliced and polyadenylated transcripts stay associated with chromatin at the gene locus before they are exported to the cytoplasm (Brody et al. 2011).

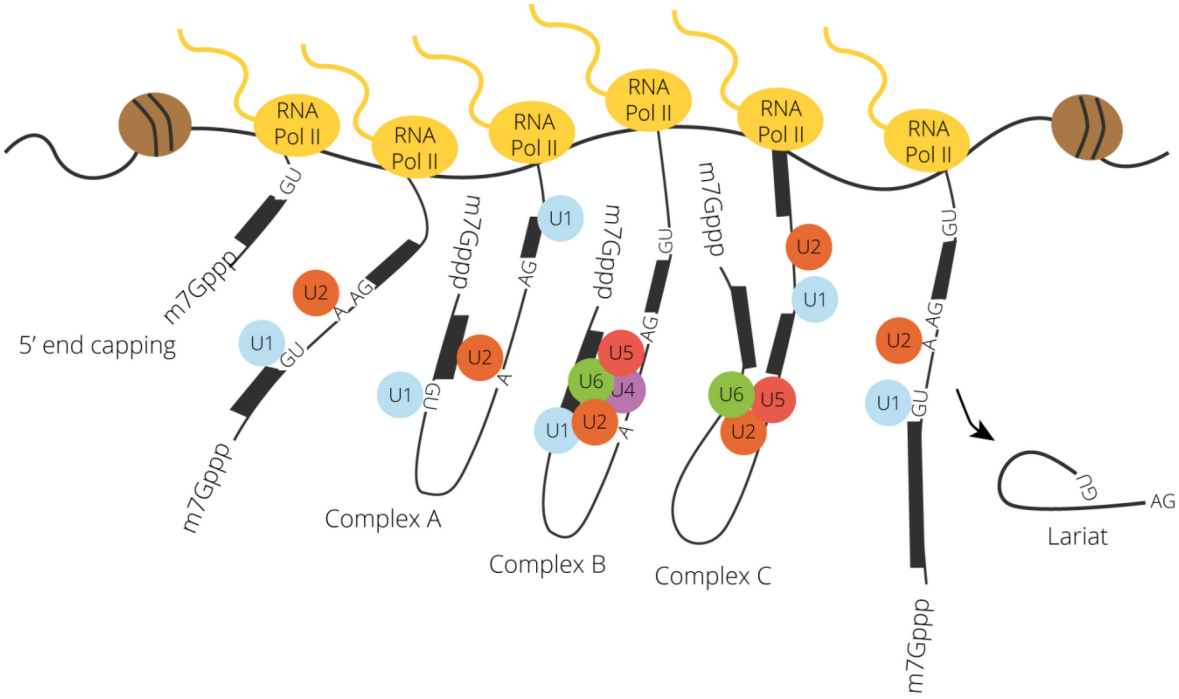

Figure 11. Many pre-mRNAs transcripts are spliced co-transcriptionally. DNA wraps around nucleosomes (light brown) to form chromatin structures. The C-terminal domain (CTD) of RNA polymerase II (RNA Pol II) acts as a docking site for various proteins, facilitates pre-mRNA synthesis, and coordinates co-transcriptional processing events including transcription initiation, 5' end capping, 3' end formation, and RNA splicing. As nascent pre-mRNA is being transcribed, the spliceosome components (U1, U2, U4, U5, and U6 – colored bubbles) are recruited in a step-wise manner onto the pre-mRNA. U1 and U2 assemble at the 5′ splice site (GU) and branch point (A), respectively, to form complex A. Recruitment of U4/U5.U6 tri-snRNP forms complex B, which undergoes several rearrangements to form the catalytically active complex C. The intron is released as a lariat, the exons are ligated together, and the snRNPs disassemble (adapted form Wong et al. 2014).

Co-transcriptional splicing seems to be an evolutionarily conserved mechanism since even in yeast, where splicing occurs in a limited number of genes, the influence of transcription on splicing is visible. It was shown that RNA polymerase II pause downstream of one of the two exons in a group of efficiently spliced genes and it was suggested that this pausing is necessary to allow co-transcriptional splicing (Carrillo Oesterreich et al. 2010).

There is also evidence showing the influence of splicing on transcription. Including an intron sequence into mammalian expression vector very often results in the increase of gene's expression. Moreover, in yeast, the majority of multiexonic genes are also the most highly expressed ones (Ares et al. 1999). In addition, the importance of promoter-proximal 5'ss and U1 snRNP for early steps of transcription and efficient transition to elongation RNA polymerase II complex was shown (Fong and Zhou 2001; Furger et al. 2002; Kwek et al. 2002). U1 snRNP is associating with transcription machinery probably via CTD of RNA polymerase II (Das et al. 2007; Spiluttini et al. 2010; Brody et al. 2011). It seems to be important for producing transcription in the sense direction of bidirectional promoters since there is the enrichment of promoter-proximal U1 recognition sites in sense transcripts and their relative depletion in antisense transcripts (Almada et al. 2013). Moreover, inhibition of splicing or intron deletion leads to the reduction of transcriptional output (Bieberstein et al. 2012).

Mutual Relationship of Chromatin, Transcription, and Splicing

Since the splicing is co-transcriptional, the process of transcription can easily influence splicing dynamics. However, not only the transcription is coupled with splicing but also other RNA processing reactions (Hirose and Manley 2000; Hsin and Manley 2012). The key factor for such coordination is the C-terminal domain (CTD) of RNA polymerase II (David and Manley 2011). This domain in humans typically consists of 52 tandem repeats of the heptapeptide YSPTSPS (Tyrosine-Serine-Proline-Threonine-Serine-Proline-Serine) sequence which is dynamically phosphorylated on serine, tyrosine and threonine residues during the various steps of transcription (**Figure 12**).

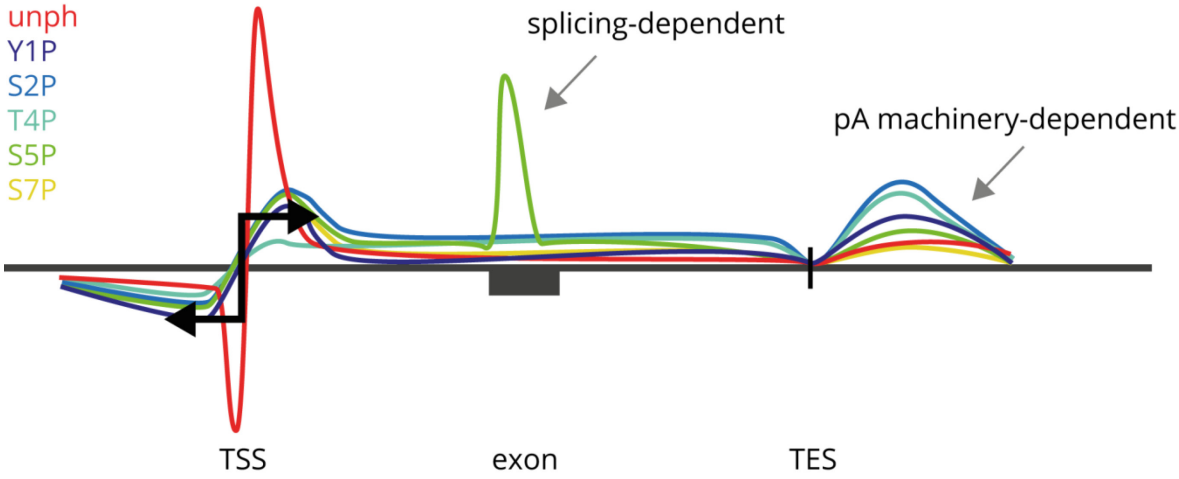

Figure 12. Average ChIP profiles of phosphorylated residues of RNA polymerase II CTD in metazoans. Unphosphorylated (unph) CTD is accumulated over transcription start sites (TSS). Tyrosine 1 phosphorylation (Y1P) level is highest near the promoter and is associated with paused RNA polymerase II. Serine 2 phosphorylation (S2P) level increases across gene bodies and transcription end site (TES), and is important for the recruitment of capping enzymes. Threonine 4 phosphorylation (T4P) accumulates over TES and is important for transcription elongation and termination. Serine 5 phosphorylation (S5P) peaks near the promoter and TSS. Its profiles correlate with active splicing on PCGs and accumulate over 3'ss of intron-containing genes. It is also important for 3' end processing. Serine 7 phosphorylation (S7P) functions in 3' end processing of snRNAs and histone mRNAs (adapted from Naftelberg et al. 2015; Nojima et al. 2015; Harlen and Churchman 2017).

The CTD of RNA polymerase II consists of three parts. One is a flexible linker, second is the region comprising the heptapeptide repeats and third is the tip region located at the end. CTD is in an extended conformation and positioned near the mRNA exit tunnel, where it can interact with many proteins associated with the nascent mRNA (Cramer et al.

2001). Because of its structural flexibility and the variety of binding surfaces generated by extensive post-translational modifications on the repeat residues, the CTD has the ability to interact with and recruit distinct proteins at various stages of transcription (Meinhart et al. 2005; Noble et al. 2005; Lunde et al. 2010; Kubíček et al. 2011; Kubicek et al. 2012; Jasnovidova et al. 2017a; Jasnovidova et al. 2017b).

Role of RNA Polymerase II Transcription Rate on Splicing

Two nonexclusive mechanisms explaining the mutual relationship of splicing and transcription were proposed: recruitment and kinetic coupling models (**Figure 13**). In the recruitment model, CTD of RNA polymerase II functions as a scaffold for various proteins involved in gene expression (e.g. release of RNA polymerase II from promoter-proximal pausing, capping, methylation, splicing, polyadenylation, and cotranscriptional chromatin modification) (reviewed in Srivastava and Ahn 2015). In the kinetic coupling model, the speed of RNA polymerase II influences splicing by affecting the pace at which splice sites and splicing regulatory elements emerge from RNA polymerase II. Indeed, it was shown that the slow mutant of RNA polymerase II changes the splicing pattern (Mata et al. 2003). One mechanism how the speed of RNA polymerase II can influence splicing is that slow elongation (caused by RNA polymerase II pausing (Roberts et al. 1998) or drugs (Nogués et al. 2003; Ip et al. 2011)) gives more time for an upstream splice site and regulatory elements to occur and be recognized by the splicing machinery before downstream signals. This can lead to exon inclusion since it favors the recruitment of splicing factors to the upstream exon before the downstream exon is transcribed (de la Mata et al. 2010). However, slow elongation can also results in greater exon skipping because there is more time for negative regulatory splicing factors to bind to nascent RNA in the upstream intron (Dujardin et al. 2010; Dutertre et al. 2010; Solier et al. 2010; Ip et al. 2011; Dujardin et al. 2014). It was shown that the kinetic coupling model can account for only a fraction of the effects of the elongation rate on splicing (Fong et al. 2014). Moreover, slow and fast elongation rates often disrupt splicing in the same way meaning that slow and fast elongation often do not have opposite effects on exon inclusion or skipping (Fong et al. 2014). So the final splicing outcome (either positive or negative) seems to be dependent on the identity and architecture of the particular gene sequences that surround the alternative splicing event and an optimal elongation rate is required to achieve a normal balance of spliced isoforms (Fong et al. 2014).

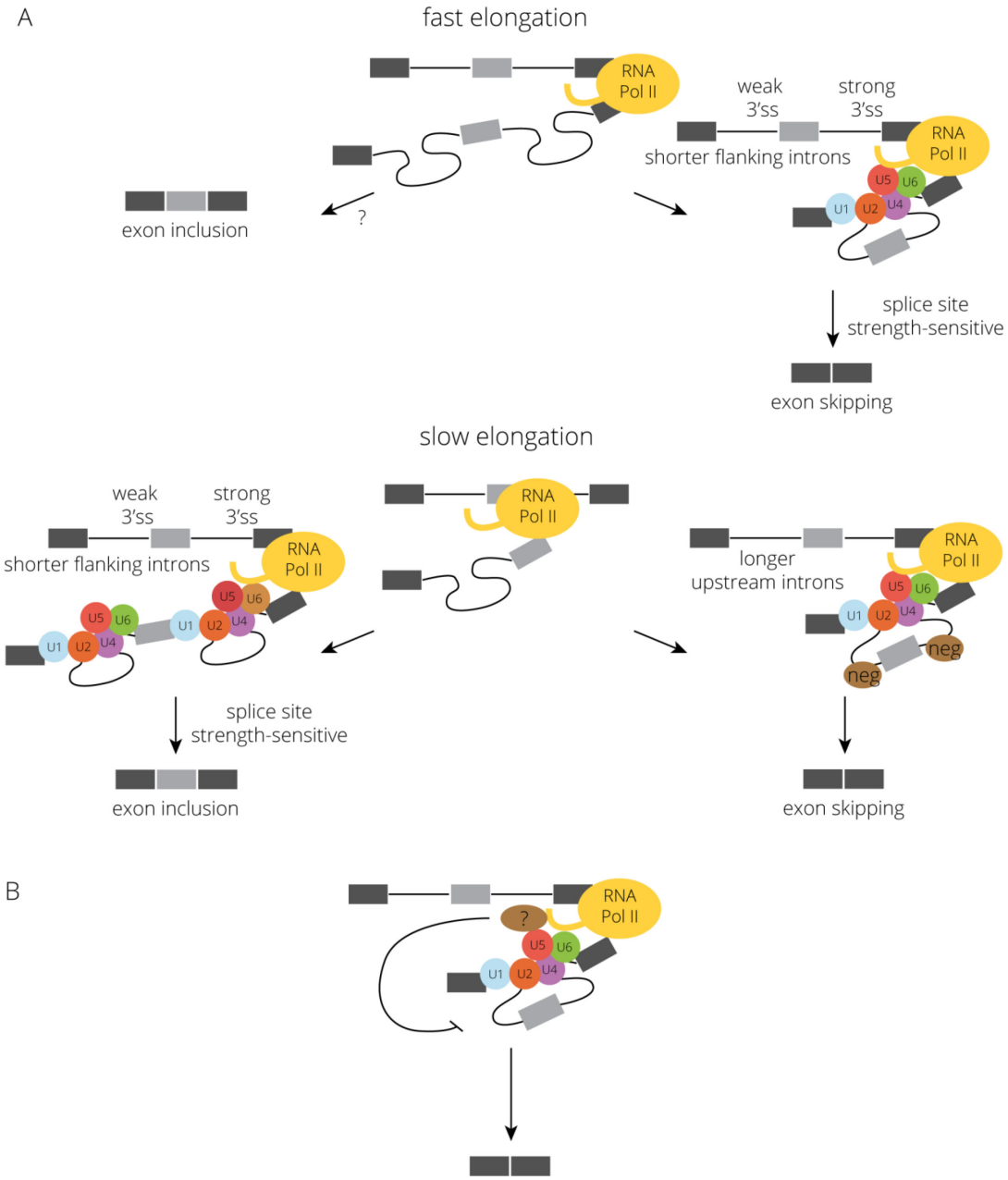

Figure 13. Coupling of transcription and alternative splicing can be explained by two non-mutually exclusive models. (**A**) Two different elongation rates (fast and slow) of RNA polymerase II can result in different splicing outcomes. In the case of a splice site strength-sensitive alternative splicing event, fast elongation leads to exon skipping and slow elongation leads to exon inclusion. This is the result of the competition between weak and strong 3'ss. During fast elongation, there is a short time window between the synthesis of both 3'ss, and the strong 3'ss competes with the weak 3'ss for the commitment of the splicing machinery, promoting alternative exon skipping. On the other hand, when the elongation rate is low, there is more time for the weak 3'ss to be recognized before the strong 3'ss is synthesized which promotes exon inclusion. In other cases, slow elongation can also result in exon skipping because of more time for negative regulatory splicing factors (neg) to bind to the alternative intron. Even though fast elongation can lead to exon inclusion, the exact mechanism is still not known. (**B**) In the recruitment model, splicing factors that are recruited to the transcription machinery can affect alternative splicing decisions (adapted from Fong et al. 2014; Nieto Moreno et al. 2015).

The physiological function of such influence of RNA polymerase II elongation rate on alternative splicing is questionable. In favor of its physiological function are the different elongation rates between genes and within genes (Danko et al. 2013; Jonkers et al. 2014; Veloso et al. 2014) which range from 0.5 to 4 kb/min and their responsiveness to physiological stimuli (Danko et al. 2013). Moreover, several oncogene products are implicated in the control of transcription elongation (Rahl et al. 2010; Smith et al. 2011) which is consistent with widely documented changes in alternative splicing in cancer and other diseases (David and Manley 2010; Germann et al. 2012). Together, this assumes that RNA polymerase II elongation rate has physiological implications on the alternative splicing.

Role of Chromatin Structure and Modifications on Splicing

Not all alternative splicing events are affected by different elongation rates (Ip et al. 2011). Another player of the mutual relationship between splicing and transcription is chromatin serving as a scaffold for co-transcriptional splicing (reviewed in Dujardin et al. 2013; Gómez Acuña et al. 2013). The basic unit of chromatin structure is the nucleosome consisting of two copies of each histones H2A, H2B, H3, and H4 (Luger et al. 1997). Chromatin function can be regulated by substituting canonical histones with their variants and/or through post-translational modifications of histone tail residues (Kouzarides 2007; Talbert and Henikoff 2010). These alterations result in changes of chromatin compaction and modulate the binding of effector proteins. The distribution of such chromatin features is not uniform over the genome, with differences between genes, between promoters and gene bodies, and even between particular exons and introns.

Specific chromatin features can affect splicing by impacting transcription elongation. Global effect of chromatin modification was shown when higher H4 acetylation resulted in altered splicing of hundreds of genes and RNA polymerase II processivity near the altered exons (Hnilicová et al. 2011; Zhou et al. 2011; Dušková et al. 2014). It was also observed that some chromatin remodelling factors can influence the elongation rate by interacting directly with RNA polymerase II (Wilson et al. 1996; Neish et al. 1998) and splicing factors (Batsché et al. 2005). Moreover, it was also shown that inhibition of splicing or intron deletion leads to the reduction of H3K4me3 levels in the promoter-proximal region as well as reduction of transcriptional output (Bieberstein et al.

2012). However, all these studies showing the reciprocal influence of splicing and chromatin modifications have relied on global changes of chromatin, so secondary effects of such chromatin modulation on transcription and/or splicing cannot be excluded. Therefore, the influence of local chromatin changes on alternative splicing remains unclear.

In the case of the recruitment model, the mechanism of chromatin influencing splicing is the association of different histone modifications with specific splicing factors via bridging chromatin-binding proteins (adaptors). Examples of such splicing regulation include the recruitment of U2 snRNP to H3K4me3 via CHD1 protein (Chromodomain-helicase-DNA-binding protein 1) (Sims et al. 2007) or the recruitment of PTBP1 (polypyrimidine tract binding protein 1) protein to H3K36me3 through the MRG15 (MORF-related gene 15) adaptor (Luco et al. 2010). Another example of such regulation is CBX5 (chromobox 5, also called HP1 - heterochromatin protein 1) which was shown to be recruited by repressive histone methylation marks (e.g. H3K9me2/3, H3K27me3) leading to slowing down RNA polymerase II (Allo et al. 2009). The role of methyltransferase EHMT2 (Euchromatic histone-lysine N-methyltransferase 2) describing the adaptor system of CBX3 (chromobox 3, also called HP1γ) recruiting SRSF1 protein upon recognition of H3K9me was proposed (Salton et al. 2014).

Splicing is also influenced by nucleosome occupancy. It has been shown that nucleosomes are preferentially located over exon sequences, mainly when exon has a higher GC content than its flanking introns (Schwartz et al. 2009; Spies et al. 2009; Amit et al. 2012; Tilgner et al. 2012). Exon-enriched nucleosomes may also differ in their histone variant composition. H2A.Bbd, variant of H2A histone, was shown to be associated with active, intron-containing genes and preferentially flanks 5’ and 3’ss (Tolstorukov et al. 2012). H2A.Bbd functions in splicing through the recruitment of splicing components (Tolstorukov et al. 2012) and its depletion results in widespread disruption of constitutive and alternative splicing. Moreover, nucleosomes are more frequently positioned within constitutively spliced rather than alternatively spliced exons (Schwartz et al. 2009; Huang et al. 2012). However, histones are more accumulated on exons with weak splice sites (Tilgner et al. 2009). Additionally, modified 5-methylcytosine at CpG dinucleotides are enriched within exons relative to introns (Feng et al. 2010; Chodavarapu et al. 2010; Laurent et al. 2010) with characteristic patterns at 5’ and 3’ss (Laurent et al. 2010). It was shown that differential methylation pattern correlates with differential alternative splicing

(Lyko et al. 2010). Furthermore, intron sequences were observed to be very rarely occupied by nucleosomes because of the presence of PPT near 3'ss which act as barriers for nucleosomes (Schwartz et al. 2009; Tilgner et al. 2009; Chen et al. 2010).

The communication of the promoter and alternative splicing event downstream was also studied. The first evidence of such coordination was the experiment with promoter swapping demonstrating dependence of splicing on promoter structure but no on promoter strength (Cramer et al. 1997). Moreover, promoter structure was shown to be critical for bindings of some SR proteins (SRSF1, SRSF7) which leads to changes in splicing suggesting that the transcription machinery can modulate splicing through the recruitment of SR proteins (Cramer et al. 1999). In addition, the acetylated histone binding protein Brd2 influencing the splicing of several hundred genes was found to preferentially bound to promoters of its target genes (Hnilicová et al. 2013). The connection of promoter and splicing is also supported by the finding that two point mutations in the promoter that reduce the binding of histone acetyltransferase p300 to the promoter significantly alter the splicing pattern (Dušková et al. 2014). In addition to promoters, the regulation of transcription is also mediated by transcription enhancer elements usually located several kilobases away from the transcription start sites (reviewed in Blackwood and Kadonaga 1998; Levine 2010; Bulger and Groudine 2011; Spitz and Furlong 2012; Calo and Wysocka 2013; Rivera and Ren 2013; Shlyueva et al. 2014). However, the influence of these transcription enhancers on splicing was not studied yet.

Taken together, splicing can be affected through various mechanisms including changes in speed of RNA polymerase II transcription, and/or through differential recruitment of splicing factors to nascent pre-mRNA transcripts or chromatin components. However, it is not clear to what extent are these mechanisms distinct or overlapping since, in some cases, recruitment of splicing factors can influence transcription kinetics, and in other cases, altered elongation affects the recruitment of splicing components to chromatin. The final splicing outcome appears to be dependent on individual characteristics of particular gene locus and mutual interplay between transcription, chromatin and splicing processes.

Long Non-Coding RNAs

LncRNAs are genes with minimal or no protein-coding potential and are longer than 200 nt. The length limit is arbitrarily set to partition well-known short ncRNAs (e.g. tRNAs, miRNAs, snRNAs, snoRNAs) from more mRNA-like transcripts. A majority of lncRNAs were discovered several years ago when it was shown that large parts of the human genome are transcribed (Bertone et al. 2004; Kapranov et al. 2007), and only a fraction of these transcripts account for PCGs (Birney et al. 2007). There are various ways how to categorize lncRNAs, e.g. according to their location within the genome (**Figure 14**) or their regulatory functions. The biggest similarity with PCGs in terms of genomic structure or gene length has been found for long intergenic non-coding RNAs (lincRNAs) (Derrien et al. 2012; Lagarde et al. 2017).

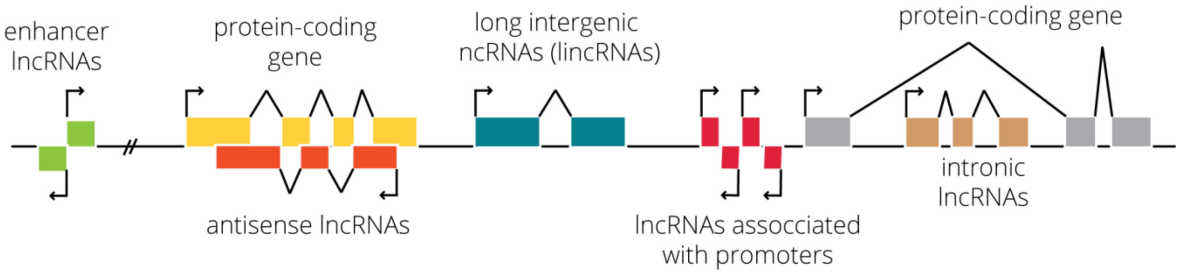

Figure 14. Several types of lncRNAs categorized according to their genomic location. LncRNAs can be categorized according to their proximity to PCGs in the genome: sense, antisense, bidirectional, intronic, and intergenic. Sense lncRNAs overlap with the sense strand of a PCG, while antisense lncRNAs with antisense strand. Bidirectional lncRNAs are located on the opposite strand from a PCG is transcribed with the limit of 1,000 bp away of its TSS. Intronic lncRNAs are completely derived from an intron of another PCG as a true independent transcript or as a product of pre-mRNA processing. Intergenic lncRNAs (lincRNAs) are located between other PCGs, so they do not overlap.

Besides the absence of a long open reading frame (ORF), lncRNAs lack any biochemical distinction from mRNAs. However, there are some differences as well. In general, lncRNAs are expressed at lower levels (~100x less abundant than mRNAs) (Derrien et al. 2012; Djebali et al. 2012; Mukherjee et al. 2016) and display more diverse tissue-specific expression patterns than PCGs (Cabili et al. 2011; Derrien et al. 2012; Lagarde et al. 2017). This tissue-specific transcription could indicate their specific role in these tissues (Mercer et al. 2008; Guttman et al. 2009). Alternatively, tissue-specific chromatin changes and transcription factors might induce lncRNA transcription as a by-

product of their main function in the regulation of PCGs expression. Even though most lncRNAs evolve rapidly in terms of sequence and expression levels, tissue specificities are often conserved (Necsulea et al. 2014). Furthermore, lncRNA expression is often correlated with the expression of mRNA both in *cis* and in *trans*, suggesting that certain lncRNAs may be co‑regulated in expression networks (Guttman et al. 2009).

Previously, it was proposed that lncRNAs tend to have on average two exons (42% of lncRNA transcripts have only two exons compared with 6% of PCGs) (Derrien et al. 2012). However, with new more precise sequencing techniques, the average number of exons per transcript in lncRNAs was estimated for 4.27 (6.69 for PCGs) (Lagarde et al. 2017). Because lncRNAs have less number of exons, overall lncRNA spliced transcripts are shorter than PCGs, even though lncRNAs exons and introns are slightly longer than that of PCGs (2,280 vs 1,602 bp and 149 vs 132 bp exons and introns, respectively) (Derrien et al. 2012). However, the length of spliced transcripts has also come under scrutiny, and it was currently estimated that the median length of lncRNAs is 1,108 nt while for PCGs is 1,240 nt (Lagarde et al. 2017). It is worthy to mention that this new technique selects against shorter transcripts so that the definite answer will be probably provided in the future with other advanced sequencing approaches.

Promoter sequences (~1-10 kb upstream) were shown to be conserved in lncRNAs (Carninci et al. 2005; Guttman et al. 2009; Derrien et al. 2012; Necsulea et al. 2014) and contained binding sites for known transcription factors (TFs) (Carninci et al. 2005; Melé et al. 2017). The number of such conserved TF binding sites is higher in the lncRNAs with annotated functions (Melé et al. 2017). Previous studies showed strong differences between lncRNA and mRNA promoter architectures (Alam et al. 2014; Melé et al. 2017) while the newer study did not find any significant difference in active promoter histone marks (H3K4me, H3K9ac) between lncRNAs and PCGs (Lagarde et al. 2017). On the other hand, repressive chromatin marks (H3K9me3, H3K27me3) show elevated levels in lncRNAs compared to PCGs which can be a consequence of elevated recruitment of the Polycomb repressive complex to lncRNAs (Lagarde et al. 2017; Melé et al. 2017).

Most eukaryotic promoters are bidirectional, so the transcription is happening in both directions (Seila et al. 2008). A divergent transcription includes transcription of mRNA from one strand as well as transcription from the different strand, in the opposite direction. A large proportion of lncRNAs account for such transcription. Usually, this

divergent transcription is coordinated, such that high expression of mRNA also results in higher levels of the corresponding antisense ncRNA (Sigova et al. 2013). However, frequently elongation is productive only in the sense direction (Core et al. 2008; Preker et al. 2008; Seila et al. 2008) which is most likely a result of the asymmetric distribution of polyadenylation and splicing signal sequences (Almada et al. 2013). Polyadenylation signals are enriched in the nearby antisense direction from a TSS, while U1 snRNP splicing signals are enriched in the sense direction ensuring the early termination and polyadenylation of antisense transcripts. A possible function of such antisense ncRNA transcripts is to regulate the promoter and corresponding PCG in *cis* (Guil and Esteller 2012).

Many nascent lncRNAs are transcribed by RNA polymerase II, contain introns and undergo the same RNA processing steps as pre-mRNAs including capping, splicing, polyadenylation, and chemical base modification (reviewed in Quinn and Chang 2016). LncRNAs have standard canonical splice site signals and undergo extensive alternative splicing (Derrien et al. 2012; Deveson et al. 2018). It is believed that introns of lncRNAs are spliced by the same spliceosome machinery as pre-mRNAs (reviewed in Will and Lührmann 2011; Matera and Wang 2014). In PCGs, it was shown that the more expressed gene is, the lower its evolutionary rate is (Pal et al. 2001; Drummond et al. 2006). The same also applies for splice-related constraints (Parmley et al. 2007) including splice site sequences and splicing enhancers which operate close to (within 70 bp) exon-intron junctions (Fairbrother et al. 2004). These motifs evolve at considerably lower rates than non-motif sites (Carlini and Genut 2006; Parmley et al. 2006; Parmley et al. 2007). LncRNAs frequently contain conserved promoter regions and splice sites (Ponjavic et al. 2007; Rose et al. 2011; Nitsche et al. 2014; Washietl et al. 2014). Moreover, it was proposed that the great majority of selection on lncRNAs is splice related, purifying selection being dominantly on exon ends with splicing enhancer motifs especially slow evolving indicating a selection for splicing and transcription in lncRNAs (Schüler et al. 2014). However, several bioinformatic studies reported that lncRNAs/lincRNAs, both steady-state and nascent RNAs, are less efficiently spliced than pre-mRNAs of PCGs (Tilgner et al. 2012; Mukherjee et al. 2016; Lagarde et al. 2017; Melé et al. 2017; Schlackow et al. 2017). One possible mechanism explaining the difference in the splicing efficiency between lincRNAs and PCGs is the absence of proximal RNA Pol II

phosphorylation over 5'ss in lincRNA transcripts (Mukherjee et al. 2016). However, the precise molecular mechanism for this phenomenon has not been elucidated.

It was suggested that lncRNAs transcripts represent important regulators of PCGs expression that control every level of the gene expression program (**Figure 15**) (for reviews see Mattick et al. 2009; Wang and Chang 2011; Patrushev and Kovalenko 2014; Ransohoff et al. 2017). There are several basic mechanisms how lncRNAs can control gene expression programmes (reviewed in Engreitz et al. 2016b). First one includes spatial amplification of regulatory information as lncRNAs can localize and spread across chromatin in proximity to their genomic loci. In such case, they can simultaneously regulate multiple genes that are spatially clustered, so there is no need to regulate each gene independently (e.g. Xist, Firre, Kcnqlot1). This is similar to operon architecture in bacteria when such a gene arrangement can increase efficiency by ensuring the coordinated expression of functionally dependent components (Rocha 2008). Moreover, lncRNAs, unlike DNA regulatory elements, can amplify DNA-encoded regulatory signals to different extents according to their expression levels and mediate contacts even between different chromosomes. Secondly, lncRNAs can partition the nucleus via guiding regulatory complexes to specific locations in the nucleus (e.g. recruiting chromatin regulatory complexes to specific sites on DNA, increasing the local concentration of a particular factor through multimerization). This can improve the kinetic efficiency of some nuclear processes (e.g. Xist, NEAT1). Finally, lncRNAs can also enable rapid assembly and disassembly of nuclear compartments through their transcription/degradation mechanism since they are functional immediately upon transcription and diffused in the nucleus (e.g. NEAT1 is required to maintain their associated nuclear compartments).

The localization of lncRNAs is specific for each particular lncRNA, mostly dependent on its function. Initially, it was assumed that lncRNAs are nuclear-restricted (Khalil et al. 2009). On top of that, many of the most characterized and studied lncRNAs (e.g. Xist, HOTAIR, HOTTIP, MALAT1, NEAT1, Firre) are primarily localized into nucleus (Brown et al. 1992; Rinn et al. 2007; Clemson et al. 2009; Wang et al. 2011b; Yang et al. 2011; Hacisuleyman et al. 2014). However, the new examples of lncRNAs exhibiting a variety of cellular localization pattern have emerged showing that lncRNAs are present in every subcellular compartment from defined subnuclear points to diffuse whole-cell spread (Cabili et al. 2015; Carlevaro-Fita et al. 2016; Mas-Ponte et al. 2017). Nowadays, the consensus is that lncRNAs are modestly enriched in the chromatin and

nuclear fractions relative to cytoplasm (Derrien et al. 2012; Djebali et al. 2012; Cabili et al. 2015) even though more lncRNAs by transcript number are present in the cytoplasm than in the nucleus (Ulitsky and Bartel 2013).

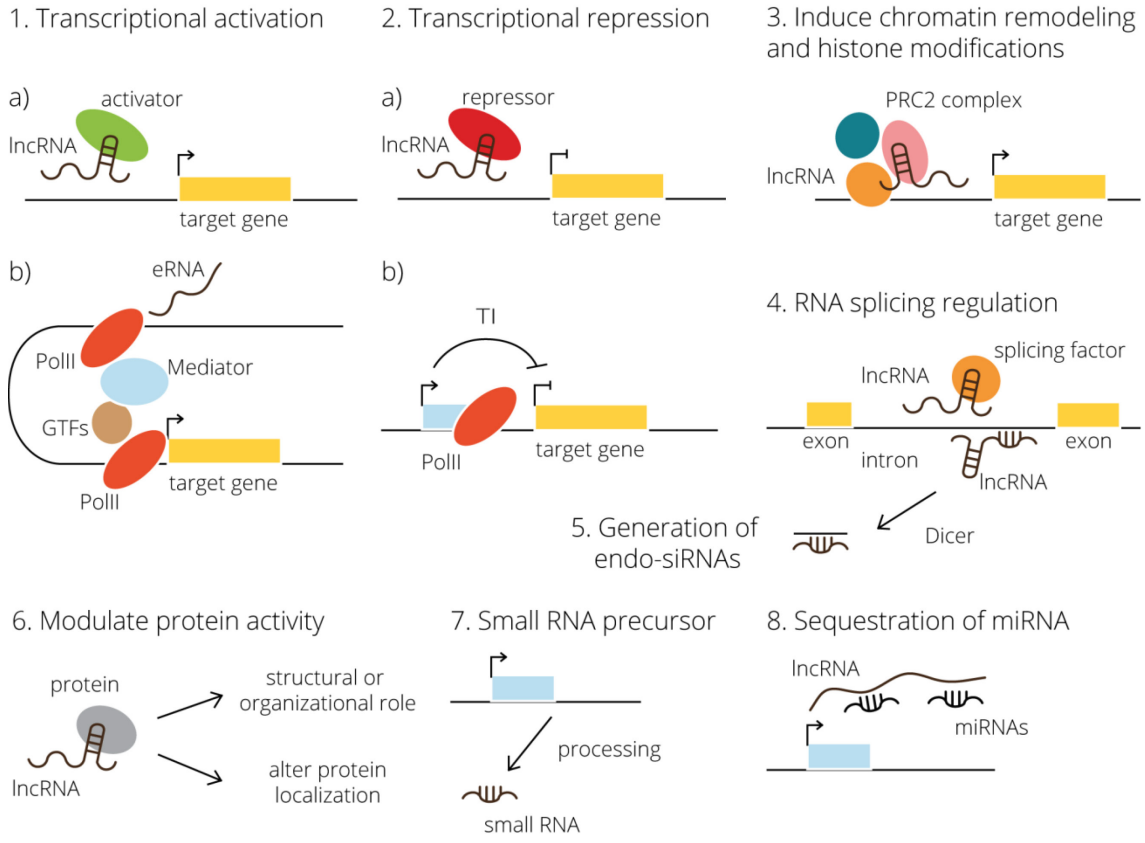

Figure 15. LncRNAs have diverse functions within the cells. LncRNAs can regulate their target gene's expression in various ways. (**1**) LncRNAs can activate the transcription by (a) interaction with transcriptional activators (green) or (b) they can be transcribed from an enhancer region (eRNA). (**2**) LncRNAs may also mediate transcriptional repression either (a) by binding a repressor (red) or (b) by blocking RNA polymerase II (orange) recruitment in the process called transcription interference (TI). (**3**) LncRNAs can also regulate gene expression through inducing chromatin remodeling and histone modifications (e.g. PRC2 complex). (**4**) An antisense lncRNA transcript can hybridize to the overlapping sense transcript and block or recruit splicing factors (light orange) to the splice sites and thus regulate alternative splicing. (**5**) Hybridization of the sense and antisense transcripts can allow Dicer to generate endogenous siRNAs (endo-siRNAs). (**6**) By binding to proteins (gray), lncRNAs can modulate their activity by allowing a larger RNA-protein complex to form or alter the protein localization in the cell. (**7**) LncRNAs can be processed to yield small RNAs (such as miRNAs, piRNAs). (**8**) LncRNAs may also serve as molecular sponges by harboring binding sites for miRNAs and titrating them away from their mRNA targets (modified from Wilusz et al. 2009; Malik and Feng 2016).

In the past, it was assumed that lncRNAs are generally unstable, mainly because of their general lower expression. However, the genome-wide examination of ~800 lncRNAs

(and ~12,000 mRNAs) in mouse revealed that only a minority of lncRNAs are unstable with a wide range of half-lives comparable to mRNAs (Clark et al. 2012). In more details, intergenic and *cis*-antisense RNAs are more stable than those derived from introns, and nuclear-localized lncRNAs are more likely to be unstable. Another study of RNA stability in human cell line showed that a significant proportion of RNAs with a long half-life (>4 h) accounts for ncRNAs. RNAs with a short half-life (<4 h) include regulatory ncRNAs (Tani et al. 2012). The mechanism of lncRNA degradation seems to involve the same RNA quality-control pathways as mRNAs. After the depletion of polyadenylation polymerases, nuclear RNases, and exosome subunits, thousands of ncRNAs accumulated in yeast (Davis and Ares 2006). Moreover, disruption of the NMD pathway in *Arabidopsis thaliana* lead to significantly greater upregulation of lncRNAs relative to PCGs (Kurihara et al. 2009). A hypothesis of lncRNAs as the product of transcriptional noise proposes that the numerous post-transcriptional quality-control mechanisms that have evolved limit their expression (Wyers et al. 2005). It was shown that half-life of a mRNA is closely related to its physiological function. This leads to the possibility that the stability of a particular lncRNA reflects its function. Even though the half-lives were on average shorter than that of mRNAs (Clark et al. 2012), this suggests complex metabolism and widespread functionality of lncRNAs.

With the development of new advanced sequencing techniques, even low-abundance transcripts of lncRNAs could be detected. Even though it is possible that some transcripts are the result of transcriptional noise, the majority of lncRNA transcripts are assumed to be functional. However, their real coding potential has remained controversial. ORFs can exist practically in any stretch of a random RNA sequence, and the probability of ORF presence increases with the length of such sequence. Recently, using genome-wide ribosome profiling followed by sequencing, it was estimated that translation is far more pervasive than previously anticipated and takes place on many transcripts assumed to be non-coding (Brar et al. 2012; Chew et al. 2013; Aspden et al. 2014; Bazzini et al. 2014; Ingolia et al. 2014; Pauli et al. 2014; Ruiz-Orera et al. 2014; Smith et al. 2014; Carlevaro-Fita et al. 2016). Since the majority of peptides produced from such interactions are relatively short (~20-100 aa) and are usually unstable and rapidly degraded, it is technically challenging to identify them. It was suggested that such translation can represent a previously unrecognized source of short proteins in diverse organisms (reviewed in Pauli et al. 2015). However, the association with ribosomes is not always

indicative of real translation of a transcript. Such a transcript can be a part of translation regulatory mechanisms (even the ribosome itself is composed of ncRNAs), or the observed association can be a result of random interactions. Moreover, many transcripts fall into the transition zone of translation with translated ORFs whose peptide products might not be functional (Pauli et al. 2015). Re-analyzing of original ribosome-profiling experiments showed that majority of ribosome-associated lncRNAs resemble the scanning ribosome profile on 5'UTRs without ribosome releasing at stop codons as at bona fide coding ORFs (Guttman et al. 2013; Ingolia et al. 2014). Taken together, it is recently assumed that the vast majority of lncRNAs can be still considered to be non-coding and ~5% of supposedly lncRNAs can be translated similarly to protein-coding transcripts (Guttman et al. 2013).

Enhancer RNAs and Enhancer-Like Long Non-Coding RNAs

LncRNAs that promote the transcription of neighboring PCGs include enhancer RNAs (eRNAs) and enhancer-like RNAs (also called activating ncRNAs - ncRNA-a) (Ørom et al. 2010), which mainly open chromatin or inhibit the binding of repressor proteins (reviewed in Kornienko et al. 2013). Enhancer RNAs are transcribed form enhancer DNA elements (Djebali et al. 2012; Andersson et al. 2014a) which were initially described as short DNA fragments with the ability to positively drive target gene expression independently from a genomic distance and orientation relative to the target gene. Moreover, these sites were shown to be in a decompacted chromatin state (hypersensitivity to DNase treatment) and contain binding motifs for transcription factors together with enriched binding of transcription co-activators (CBP/p300) and histone acetylation (H3K27ac) (reviewed in Blackwood and Kadonaga 1998; Levine 2010; Bulger and Groudine 2011; Spitz and Furlong 2012; Calo and Wysocka 2013; Rivera and Ren 2013; Shlyueva et al. 2014). In the past, the ratio between H3K4me1/H3K4me3 was used for prediction of enhancer elements (Heintzman et al. 2007; Visel et al. 2009; Vučićević et al. 2015). However, it was shown that H3K4me1-enriched regions are generally larger than the associated *cis*-regulatory elements (Barski et al. 2007) so it is difficult to define the exact location of the actual enhancer. On top of that, the presence of H3K4me1 does not strictly correlate with the functional activity of enhancer elements (Cui et al. 2009; Creyghton et al. 2010) so it is required a more accurate definition of the chromatin signature of active enhancers and additional criteria are needed to annotate functional enhancers in the genome. However, it was also observed that H3K4me1/H3K4me3 ratio increases with the enhancer's strength

(Pekowska et al. 2011). Moreover, enhancer association of RNA polymerase II correlated with the presence of H3K4me3 and RNA polymerase II accumulation resulted in a local increase of H3K4me3 suggesting the existence of functional links between Pol II occupancy, H3K4me3 enrichment and enhancer activity (Pekowska et al. 2011). This is also supported by the findings that highly active enhancers display H3K4me3 rather than the H3K4me1 (Henriques et al. 2018). Furthermore, it was observed that in each genomic loci only one out the three methylation marks (DNA methylation, H3K4me1, H3K4me3) is high and determining the fate of that loci: DNA methylated regions are inactive, H3K4me1-enriched are enhancers, and H3K4mr3-enriched are promoters (Sharifi-Zarchi et al. 2017).

eRNAs were discovered when it was observed that multiple hypersensitivity sites produce transcripts which were cell-type and differentiation-stage specific (Lipshitz et al. 1987; Tuan et al. 1992; Ashe et al. 1997; Masternak et al. 2003; Rogan et al. 2004; Feng et al. 2006). This was also supported by transcriptome profiling and RNA polymerase II ChIP-Seq analyses showing RNA polymerase II binds to many extragenic regions (Cheng et al. 2005; Carroll et al. 2006) and can produce largely non-polyadenylated ncRNAs (De Santa et al. 2010; Kim et al. 2010). Interestingly, eRNA induction is a potent, independent indicator of enhancer activity (De Santa et al. 2010; Kim et al. 2010; Melgar et al. 2011; Hah et al. 2013; Lam et al. 2013; Li et al. 2013; Zhu et al. 2013; Andersson et al. 2014a; Wu et al. 2014). eRNA-producing enhancers exhibit higher binding of transcriptional co-activators, greater chromatin accessibility, higher enrichment of H3K27ac, protection from repressive DNA methylation, and correlation with the formation of enhancer-promoter loops (Kim et al. 2010; Melgar et al. 2011; Sanyal et al. 2012; Hah et al. 2013; Pulakanti et al. 2013; Schlesinger et al. 2013; Zhu et al. 2013). When compared with the classic models of promoter activation, enhancers exhibit many similarities to promoters in terms of the assembly of the transcriptional apparatus (recruitment of general transcription factors, Ser5p of RNA polymerase II) together with chromatin state and nucleosome phasing (Lenhard et al. 2012; Core et al. 2014). However, there are still important differences with non-overlapping sets of transcription factors enriched at enhancers and promoters (Encode Project Consortium et al. 2012; Core et al. 2014). Preferentially associated with enhancers are mainly cell-type specific transcription factors (Carroll et al. 2005; Lupien et al. 2008; Spitz and Furlong 2012; Kaikkonen et al. 2013; Ostuni et al. 2013) which is consistent with the concept that enhancers are responsible for driving cell-type-specific gene

expression (Spitz and Furlong 2012). Similarly to mRNA promoters, enhancers are also typically transcribed bi-directionally (De Santa et al. 2010; Kim et al. 2010; Melgar et al. 2011; Hah et al. 2013; Andersson et al. 2014b; Core et al. 2014) but poly(A) cleavage sites are more likely to locate closer to TSS than U1 motifs that counteract cleavage and degradation (Core et al. 2014). The lack of U1 splice sites can also explain rare splicing of eRNAs (~5% of eRNAs are spliced vs to ~30% of lncRNAs and ~80% of mRNAs) (Sigova et al. 2013; Andersson et al. 2014a).

Interestingly, there is a rather complex relationship between the splicing and enhancer activity of enhancer-like lncRNAs. There is an ongoing debate about the importance of splicing-associated processes in the function of enhancer-like lncRNAs. In the study of Yin et al. (2015), splicing-independent regulation of PCG expression mediated by enhancer RNA was proposed. They observed the same phenotype after stimulation of enhancer RNA transcription with or without introns suggesting that splicing process does not play a significant role in the regulatory function by this enhancer RNA. Similar results were obtained from another study (Engreitz et al. 2016a) where the deletion of the first 5'ss of an enhancer-like lncRNA had an effect on the activation of both lncRNA and target gene transcription, but the deletion of downstream splice sites was dispensable. This indicates that this lncRNA is in fact required for expression activation, although this mechanism does not appear to depend on the precise sequence of the RNA beyond the presence of initial 5'ss which can be important only for promoting transcription through direct interaction of spliceosome with transcriptional machinery or via stabilization of enhancer RNA. However, a recent study shows the important role of splicing for the enhancer activity of an enhancer-like lncRNA (Tan et al. 2018). So the controversy of opposing studies (Yin et al. 2015; Engreitz et al. 2016a; Tan et al. 2018) about splicing role in enhancer function still waits for a satisfactory answer.

The significant difference between eRNAs and mRNAs/lncRNAs elongation is the low recruitment of Ser2p RNA polymerase II and minimal levels of H3K36me marks (Koch et al. 2011; Bonn et al. 2012) which can be a consequence of a lack of splicing (Almeida et al. 2011). Generally, eRNAs display low stability and abundance with proposed exosome-based decay mechanism (Andersson et al. 2014a; Andersson et al. 2014b; Lubas et al. 2015; Pefanis et al. 2015) with more stable eRNAs exhibited stronger H3K4me3 marks (Core et al. 2014) and their higher activity (Pekowska et al. 2011).

The fact that eRNAs are transcribed from enhancer elements poses a challenge to elucidate the functionality of DNA element itself from the act of transcription, as well as from the RNA product. An early hypothesis of pervasive transcription at enhancers was that eRNAs represent transcriptional noise (Struhl 2007) which is supported by their relatively low transcriptional levels (Djebali et al. 2012; Andersson et al. 2014a), poor evolutionary conservation (Villar et al. 2015) and high chromatin accessibility of active enhancers (Calo and Wysocka 2013; Rivera and Ren 2013; Shlyueva et al. 2014). Thus, eRNA transcription can be a result of random scanning of the RNA polymerase II or high concentration of transcriptional machinery at active promoters upon the chromosomal interaction between enhancers and promoters (Li et al. 2012). However, many studies have described various functional roles of eRNAs over the years when reducing eRNA transcripts levels concomitantly lead to reduction of RNA levels of adjacent target genes (Ling et al. 2004; Ho et al. 2006; Ørom et al. 2010; Onodera et al. 2012; Kaikkonen et al. 2013; Lai et al. 2013; Lam et al. 2013; Li et al. 2013; Melo et al. 2013; Mousavi et al. 2013; Alvarez-Dominguez et al. 2014; Banerjee et al. 2014; Hsieh et al. 2014; Ilott et al. 2014; Maruyama et al. 2014; Ounzain et al. 2014; Schaukowitch et al. 2014; Pnueli et al. 2015; Sigova et al. 2015). Even though eRNA transcription was reduced, still there were some eRNA transcripts left, so to discriminate the role of RNA product from the act of transcription needs additional evidence. A nice example of such unlinking a lncRNA from its associated *cis* element is the study of Paralkar et al. (2016) which showed that 5' region of a lncRNA gene contains an enhancer for a neighboring gene, whereas lncRNA transcript is dispensable for target gene expression. Interestingly, the study of Engreitz et al. (2016a) supports rather the general processes associated with transcription (and possibly splicing) to be important for promoting expression of neighboring genes in *cis*. Furthermore, they propose that such effects are not limited to lncRNA loci only but can also be applied for protein-coding loci demonstrating cross-talk among neghboring genes to be a prevalent phenomenon.

Taken together, eRNAs were categorized into three classes according to their possible functions (**Figure 16**) (reviewed in Li et al. 2016). For class I eRNAs, there was no discernible function of neither their transcription nor transcripts shown. For class II eRNAs, the act of their transcription was described to be important contributor factor to their function. Class III eRNAs have RNA-dependent functions acting probably through binding proteins to control gene expression. This group is enriched in relatively abundant

and stable eRNAs, especially long non-coding eRNAs (e.g. activating non-coding RNAs (ncRNA-a)).

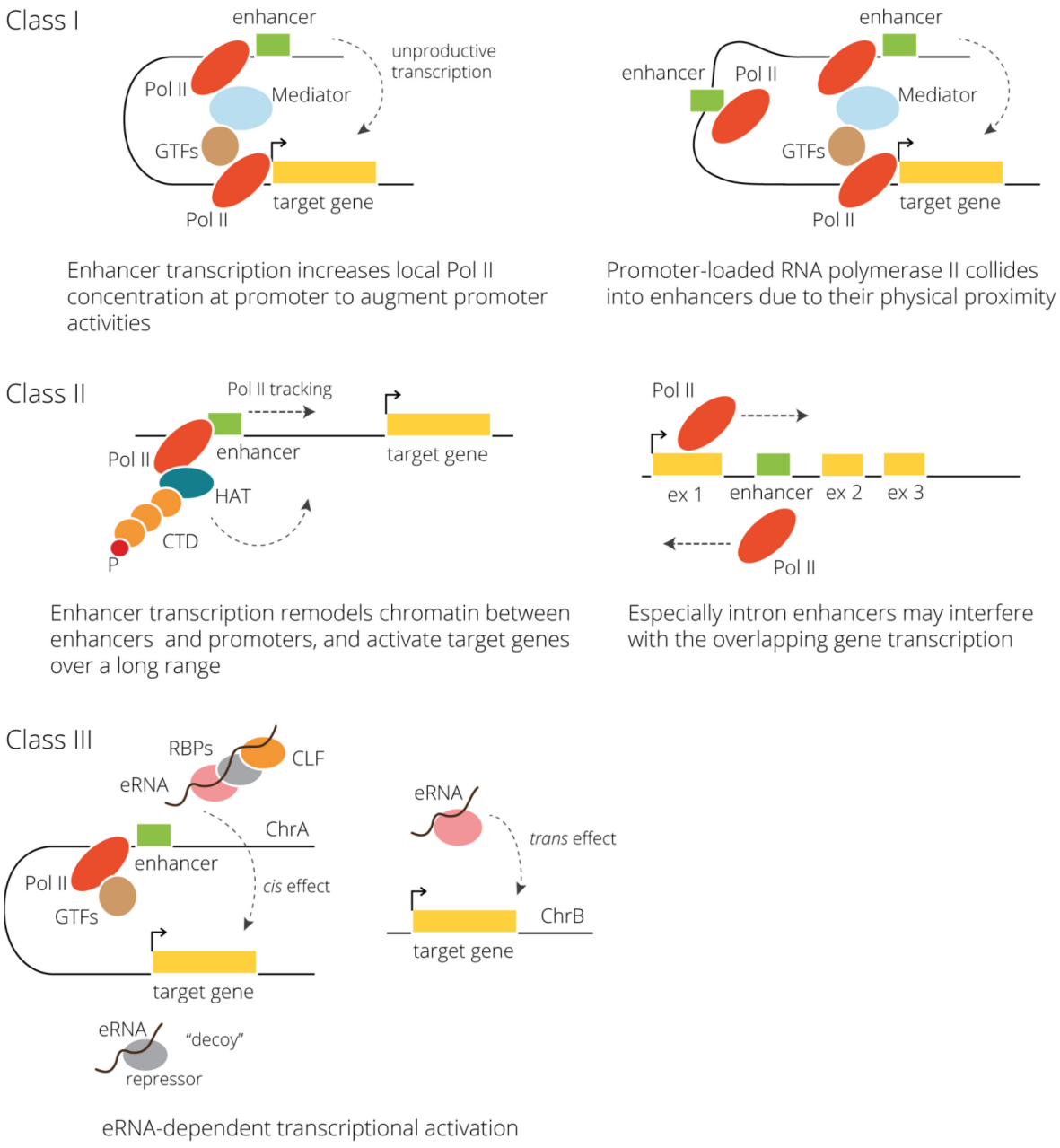

Figure 16. Classes of eRNAs according to their function roles in gene regulation. Class I - the transcription process and eRNAs are non-functional and are merely transcriptional noise. Class II - the act of enhancer transcription mediates function. Class III - genes on the same chromatin fibre (*cis* effect), or on other chromosomes (*trans* effect), are regulated by an eRNA. Pol II – RNA polymerase II, GTFs – general transcription factors, HAT – histone acetyltransferase, CTD – the C-terminal domain of Pol II, P – phosphorylation, ex – exon, eRNA – enhancer RNA, RBPs – RNA binding proteins, CLF - chromosomal looping factor (adapted from Li et al. 2016).

Evolution of Long Non-Coding RNAs

Because lncRNAs do not provide the information for the generation of proteins, there is no such immense pressure to maintain their primary nucleotide sequence. Although lncRNAs are more conserved than neutrally evolving genetic elements (Ponjavic et al. 2007; Cabili et al. 2011; Derrien et al. 2012), they are under weaker evolutionary constraints than PCGs (Wang et al. 2004a; Cabili et al. 2011; Necsulea et al. 2014; Schüler et al. 2014). LncRNAs exons are typically poorly conserved compared to PCGs (Marques and Ponting 2009) and on average evolve slower than their flanking introns (Hurst and Smith 1999; Pang et al. 2006). However, rapid evolution does not need to imply an absence of function (even highly functional Xist lncRNA contains only a few conserved regions) (Pang et al. 2006). Nevertheless, tissue specificity, as well as specific expression patterns of lncRNAs, are generally highly conserved (Chodroff et al. 2010; Washietl et al. 2014; Hezroni et al. 2015) suggesting that some lncRNAs act in similar contexts in different species. Moreover, the secondary structure of lncRNAs is sometimes better conserved than their primary sequence with 14% of the human genome suggested being under purifying selection on RNA structure (Smith et al. 2013). In addition, there is a strong correlation between the amount of non-protein-coding sequences and biological complexity (Taft et al. 2007). The ratio of the total bases on non-protein-coding to total bases of genomic DNA increases gradually with the biological complexity of an organism. Up to 30% of lncRNA transcripts appear to be primate specific (Derrien et al. 2012) and ~20% of human lincRNAs were observed not to be expressed beyond chimpanzee (Washietl et al. 2014). Moreover, the evolution of the brain in humans seems to correlate mainly with changes in non-coding sequences (on the contrary, adaptation via coding changes is dominated in immunity, olfaction and male reproduction) (Haygood et al. 2010). This suggests that particularly cognitive traits evolved in humans partially due to non-coding sequences. Even though this expansion of non-coding regions of genomes in more complex organisms cannot explain C-value paradox (the discrepancy between the amount of cellular DNA in an organism and its biological complexity), it can provide, at least, the partial answer for the G-value paradox (Taft et al. 2007). It expects that increased developmental complexity would be reflected in an increased number of genes which was not met previously. However, originally this concept has been working with the number of PCGs only, so the findings that non-coding genomic parts expand with the complexity of organisms would compensate for such a difference.

Based on their conservation, lncRNAs can be classified into four basic classes with distinct lncRNA features and probably different mechanisms of action (**Figure 17**) (reviewed in Ulitsky 2016). Class I represents lncRNAs whose exon-intron structure and multiple sequences along the length of the lncRNAs are conserved among species. It is expected that many of the *trans*-acting lncRNAs will belong to this group. Some lncRNAs of this class are enriched in the cytoplasm and therefore, probably act independently of their sites of transcription (e.g. Cyrano, NEAT1, MALAT1, NORAD, and GOMAFU) (Sone et al. 2007). In class II lncRNAs, the act of transcription and RNA elements (usually located near 5'end of the RNA) are conserved, while the rest of the locus experienced drastic changes in exon-intron structure and length. Only a few splice sites are conserved, and transposable elements contributed heavily to locus diversification across species. These lncRNAs are more likely to be *cis*-acting and to regulate gene expression in regions surrounding their loci (e.g. lncRNA found downstream of the *ONECUT1* gene). In class III lncRNAs, only promoter sequences and the act of transcription of the specific region is conserved. Usually, there are no regions with recognizable sequence similarity, and there is no conservation of gene structure. LncRNAs can be transcribed from conserved enhancer elements with limited or no function of the RNA product. Typically, the act of transcriptional elongation supported by the productive splicing is important (e.g. Lockd, FENDRR) (Paralkar et al. 2016). Class IV represents lncRNAs that seem to be found and functional in only one species and not the others (e. g. *Haunt*).

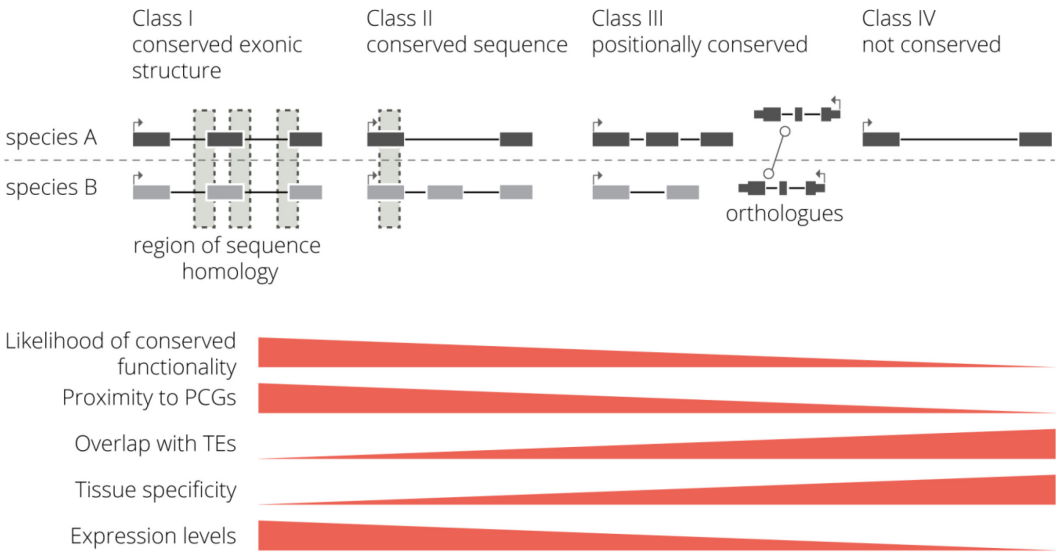

Figure 17. Classes of lncRNAs according to their conservation. Various genomic and functional features are correlated with the degree of conservation. Most conserved lncRNAs are closer to PCGs, less likely to overlap transposable elements (TEs), and are more broadly and highly expressed (adapted from Ulitsky 2016).

The relatively high frequency of new lncRNA origination is supported by the observation that most lncRNAs in vertebrate genomes do not have homologues in species separated by more than 50 million years of evolution (Hezroni et al. 2015). Three most possible processes of creating new lncRNAs include duplication of existing lncRNAs, utilization of already established transcription or exaptation (a shift in the function of a trait during evolution) of the previously non-transcribed locus (reviewed in Ulitsky 2016). However, whole-locus duplication only rarely contributes to the evolution of new lncRNAs, since if this process were common also for lncRNAs, we would expect to see some sequence similarity among lncRNAs within the same species. Although this can occur (e.g. *TUNA* or *MALAT1*/*NEAT1*) (Ulitsky et al. 2011; Derrien et al. 2012; Hezroni et al. 2015), usually it is attributed to unannotated fragments of transposable elements (Hezroni et al. 2015).

Another mechanism of creating new lncRNAs is via losing the coding potential of PCGs through mutations, insertion of transposable elements or genomic rearrangements. If such changes do not lead to a loss of transcription or triggering the nonsense-mediated decay, the new lncRNA can be formed at the same locus (e.g. *XIST*, *JPX* and *FTX*) (Duret et al. 2006; Romito and Rougeulle 2011). It was observed that transposable elements heavily contribute to the evolution of lncRNAs: ~40% of lncRNAs sequences are recognizable as derived from transposable elements, ~80% of lncRNAs overlap at least one transposable element, and ~25% of promoters and polyadenylation sites of human lncRNAs are derived from transposable elements (Kelley and Rinn 2012; Kapusta et al. 2013; Karlic et al. 2017). This is mainly caused by the fact that when a transposable element is inserted into lncRNA, in most of the cases, insertion does not affect the function, which is in contrast to the insertion into PCG, where it might disrupt functional OR. A transposable element with a functional promoter (such as ERVs – endogenous retroviruses) was shown to be sufficient to drive transcription initiation at a previously non-transcribed locus (Kapusta et al. 2013). As it was mentioned above, U1 snRNP splicing signals are enriched in the sense direction in the case of a divergent transcription. However, if antisense transcript gains splicing signals through mutations or the insertion of transposable elements, the splicing is favoured over polyadenylation and transcription is even further promoted. Additionally, this leads to the suppression of early polyadenylation, and thus stabilization of the transcription product. This can easily transform cryptic

transcripts into stable RNAs which can then acquire new functions (Gotea et al. 2013; Wu and Sharp 2013).

The last way how to create a new lncRNA is to generate the transcription from a previously non-transcribed locus through a series of mutations creating a favourable combination of promoters, splice sites and polyadenylation signals. Some promoters of newly created lncRNAs can correspond to conserved DNA sequences at syntenic loci that have an enhancer chromatin signature in other species (Engreitz et al. 2016a). These sequences may have conserved functional roles as *cis*-regulatory elements, rather than as lncRNA promoters. Additionally, particularly functional splice sites can further promote transcription since splicing has been shown to be able to stimulate RNA polymerase II initiation and elongation (Fong and Zhou 2001; Furger et al. 2002; Kwek et al. 2002; Kotovic et al. 2003; Lin et al. 2008). These lncRNAs are usually expressed under the control of enhancer elements located in their vicinity and are highly tissue-specific (Ruf et al. 2011). The existence of such lncRNA creation mechanism is supported by the observation that lncRNAs that are found away from PCGs are typically more tissue-specific that those expressed from divergent promoters with PCGs. Moreover, transposable elements with a functional promoter (such as ERVs – endogenous retroviruses) were shown to be sufficient to drive transcription initiation at a previously non-transcribed locus (Peaston et al. 2004; Faulkner et al. 2009; Jacques et al. 2013; Kapusta et al. 2013; Fort et al. 2014; Lu et al. 2014; Consortium 2015; Göke et al. 2015; Melé et al. 2015). It was suggested that more than 30% of transcription star sites overlap with retrotransposons (Faulkner et al. 2009). On top of that many ERV-derived promoters are even highly cell-type-specific (Faulkner et al. 2009; Kelley and Rinn 2012; Göke et al. 2015), and thereby increase transcriptome complexity. For many genes, the TSS in the retrotransposon is the only one and is essential as it provides the only promoter. The vast majority (97%) of such genes are lncRNAs (Göke and Ng 2016). Strikingly, ~80% of lncRNAs contain transposable elements, a considerably higher percentage compared with protein-coding genes (Kelley and Rinn 2012; Kapusta et al. 2013) with 19% of lncRNAs consist of more than 50% transposable element sequence (Kapusta et al. 2013), suggesting that exaptation of transposable elements and evolution of lncRNAs are closely related.

Material and Methods

Cell Culture, Plasmids and Transfections

HeLa and U2OS-TO cells were cultured in high glucose (4.5 g/l) DMEM (Sigma) supplemented with 100 U/ml penicillin, 100 µ g/ml streptomycin (Penicillin/Streptomycin, Gibco) and 10% (v/v) fetal bovine serum (FBS, Gibco) at 37°C and 5% CO_2.

The *ncRNA-a2* gene (PCAT6 - ENSG00000228288) was placed under the control of the CMV promoter in the pEGFP-C1 backbone using NheI and HindIII restriction sites, replacing the sequence of the GFP gene with the *ncRNA-a2* gene (pEGFP-C1_ncRNA-a2). We introduced Multiplex Identifier barcode sequences (10 nt MID3 and 10 nt MID4, Roche) at the 3'end of the *ncRNA-a2* gene to specifically detect transiently expressed ncRNA-a2 transcripts.

The human hemoglobin subunit beta (*HBB*) gene (ENSG00000244734) was amplified from genomic DNA and cloned between the KpnI and HindIII restriction sites of the pcDNA3 plasmid (pCDNA3_HBB). The *ncRNA-a2* gene containing *HBB* intron 2 was prepared from pEGFP-C1_ncRNA-a2, whereby the *HBB* intron 2 sequence was amplified from genomic DNA and cloned into the pEGFP-C1_ncRNA-a2 construct by site-directed mutagenesis PCR. The *HBB* gene containing the ncRNA-a2 intron was prepared from pCDNA3_HBB whereby the ncRNA-a2 intron sequence was PCR amplified and cloned into pcDNA3_HBB without *HBB* intron 2 by site-directed mutagenesis PCR. The plasmid of the *ncRNA-a2* gene containing the *HBB* PPT was cloned in the same way (nucleotides 823-847 of *HBB* intron 2 were used as the *HBB* PPT).

NcRNA-a2 deletion mutants (FΔ1-8, RΔ1-7, Δ1-7, ΔPPT, Δ60), ncRNA-a2 mutants with ISE motifs, ncRNA-a2 mutants with modified PPT and 5'ss mutants were prepared by PCR with specific primers using pEGFP-C1_ncRNA-a2 as a template. The FΔ1-8 mutants were prepared by deletion of regions gradually increasing by 20 bp starting 6 bp downstream of the 5'ss. Similarly, RΔ1-7 mutants were prepared by deletion of regions gradually increasing by 20 bp starting 40 bp upstream of the 3'ss. The Δ1-7 mutants were prepared by sequential deletion of 20 bp starting 6 bp downstream of the 5'ss. Mutants with ISE motifs and a negative control with a degenerated motif (Wang et al. 2012) were

introduced 25 bp downstream of the 5'ss (individual ISE motif sequences are listed in the Supplementary Material). In ΔPPT, 4 bp were deleted in the ncRNA-a2 intron sequence (intron positions: 198-201). In Δ60 mutants, 60 bp regions were deleted in the middle of the ncRNA-a2 intron (intron positions: 67-125). In ncRNA-a2 mutants with a modified PPT, nucleotides 181-201 of the ncRNA-a2 intron were modified (T21 – all nucleotides to Ts, CtoT – all Cs to Ts, GAtoT – all Gs and As to Ts). Branch points of all lncRNAs used in the study and the HBB intron 2 were predicted by SVM-BPfinder online tool (Corvelo et al. 2010), and sequences between the predicted branch point and the 3' YAG motif were mutated. For further information on modified sequences, see Supplementary Material.

SNHG8 (ENSG00000269893), BX088651.4 (ENSG00000237357), BX005266.2 (ENSG00000226007), AC005840.2 (ENSG00000256433), and AC116021.1 (ENSG00000254639) genes were amplified from genomic DNA and cloned into the pEGFP-C1 backbone using NheI/AgeI and HindIII/KpnI restriction sites, replacing the sequence of the GFP gene. We introduced Multiplex Identifier barcode sequences (MID5 and MID6, Roche) immediately downstream of the genes to specifically detect ectopically expressed transcripts.

All constructs have been verified by DNA sequencing. TALE plasmids (Bieberstein et al. 2016) were transiently transfected into cells using Lipofectamine® LTX Transfection Reagent (Thermo Fisher Scientific) according to the manufacturer's instructions and incubated for 48 h. *FOSL1* (ENSG00000175592) was induced by addition of 2 μ g/ml ionomycin calcium salt (LifeTechnologies) for 30 min 48 f after transfection, just before further experiments. All other plasmids were transiently transfected into cells using Lipofectamine® 3000 Transfection Reagent (Thermo Fisher Scientific) according to the manufacturer's instructions and incubated for 24 h with a medium change 6 h after the transfection.

RNA Isolation, Reverse Transcription and Quantitative PCR

Cells were grown to 90% confluency and RNA was isolated using either the TRIzol reagent (Thermo Fisher Scientific), which allows for simultaneous isolation of RNA and proteins, or the RNAzol reagent (Molecular Research Center). RNA was further precipitated with isopropanol, resuspended in Nuclease-Free Water (Ambion) and treated with Turbo DNase (Ambion) according to the manufacturer's protocol. Reverse

transcription was performed with SuperScript III (Thermo Fisher Scientific) using 5 µg of total RNA per 20 µl reaction and either random hexamer primers or primers complementary to barcode sequences downstream of ectopically expressed lincRNAs, respectively. cDNA was analyzed by quantitative PCR using LightCycler 480 (Roche) and LightCycler® 480 SYBR Green I Master (Roche) using the $2^{-\Delta\Delta Ct}$ method [(Ct gene of interest – Ct internal control) sample A – (Ct gene of interest – Ct internal control) sample B]. A list of used primers is provided in the Supplementary Material.

Western Blot and Antibodies

Proteins were isolated from TRIzol fractions, precipitated with isopropanol and resuspended in NEST-2 buffer (50 mM Tris-HCl pH 6.8, 20 mM EDTA, 5% (w/v) SDS). Proteins were resolved on a 12% (cellular fractions), 10% (TALEs, hnRNP H RNAi, RIP) or 8% (CRISPRa/i) SDS-PAGE, blotted onto a nitrocellulose membrane and detected using the indicated antibodies and SuperSignal Femto/Pico West (Thermo Fisher Scientific).

For Western blot, the following antibodies were used: rabbit α-H3 (Abcam ab1719), rabbit U2B‘‘ (PROGEN 57036), mouse α-tubulin kindly provided by Pavel Draber (Institute of Molecular Genetics of the Czech Academy of Sciences, Prague, Czech Republic), mouse hnRNP F/H (Santa Cruz sc-32310), mouse U2AF2 (Santa Cruz sc-53942), rat HA (Merck 11867423001), mouse GFP (Santa Cruz sc-9996), and mouse GAPDH (Abcam ab9484).

For ChIP experiments, the following antibodies were used: rabbit H3K4me2 (Abcam ab7766), rabbit H3K9me3 (Abcam ab8898), rabbit H3K36me3 (Abcam ab9050), mouse H4ac (Santa Cruz sc-377520), α-H3 (Abcam ab1719), mouse IgG (Sigma I5381).

Native Chromatin Immunoprecipitation

Firstly, beads were blocked before immunoprecipitation. 30 µl beads per IP (Protein A/G PLUS-Agarose, Santa Cruz sc-2003) was washed twice in nChIP buffer (50 mM NaCl, 5 mM EDTA, 50 mM Tris-HCl, pH 7.5) and resuspended in 30 µl nChIP buffer per IP. Then, 150 µg BSA and 60 µg sheared salmon sperm DNA was added, and beads were incubated at 4˚C with rotation until needed.

Cells were grown to 90% confluency, washed twice with ice-cold PBS + PIC (Protease Inhibitor Cocktail Set III, EDTA-Free, Merck) (2 µl /10 ml) and scraped into ice-cold PBS + PIC (2 µl/1 ml). Cells were centrifuges 600 rcf 5 min 4˚C and pellets resuspended in 0.3 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 0,2% NP-40, 15 mM Tris-HCl, pH 7.7, 0.5 mM DTT, PIC (2 µl/ml). Nuclei were released by passage through a 22 G needle and loaded on a sucrose gradient (1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 15 mM Tris-HCl, pH 7.7, 0.5 mM DTT, PIC (2 µl/ml)) and centrifuged for 20 min at 2000 g, 4˚C. Pellets were resuspended in MNase digestion buffer (0.32 M sucrose, 1 mM $CaCl_2$, 4 mM $MgCl_2$, 15 mM Tris-HCl pH 7.7 and PIC (2 µl/ml)), 1X Micrococcal buffer, 1X Purified BSA and 5.10^3 gel units of Micrococcal nuclease (New England BioLabs). Digestion was performed for 6 min at 37˚C. Reactions were stopped by EGTA (final concentration 10 mM) and centrifuged (14,000 rpm 10 min 4˚C). The supernatant was taken and the pellet resuspended in 0.2 mM EDTA, 1 mM Tris-HCl, pH 7.5, incubated for 1 h at 4˚C with rotation, centrifuged again (14,000 rpm 10 min 4˚C) and both supernatants mixed. Chromatin was diluted in nChIP buffer (50 mM NaCl, 5 mM EDTA, 50 mM Tris-HCl, pH 7.5), 10% inputs were saved for later and IPs incubated overnight at 4˚C with appropriate antibody (4 µl anti-H3K4me2, 4 µl anti-H3K9me3, 4 µl anti-H3K36me3, 4 µl nonspecific IgG, 4 µg anti-H4ac) and blocked beads. The beads were centrifuged (4,000 rcf 1 min 4˚C), washed once with ice-cold nChIP buffer, and transferred into new tubes. Then, beads were washed for 5 min at 4˚C with rotation and centrifuged (4,000 rcf 1 min 4˚C). Once again washed with ice-cold nChIP buffer and then twice in the same ice-cold buffer with increasing salt concentration (75 mM NaCl, 125 mM NaCl, 175 mM NaCl). Complexes were eluted twice with elution buffer (1% w/v SDS, 50 mM NaCl, 5 mM EDTA, 50 mM Tris-HCl, pH 7.5) for 15 min at 25˚C shaking at 1,200 rpm. Then, both 10% inputs and IPs were treated with 20 mg proteinase K for 1 h at 45˚C. DNA was precipitated with 0.5 vol. of 7.5M NH_4Ac and 2 vol. of 100% ethanol at -80˚C over night, washed with 70% ethanol, and resuspended in nuclease-free water. Signals were quantified by qPCR and signal IP over Input ($2^{(Ct(input)-Ct(IP))}$) and normalized to a non-transcribed intergenic control region.

Cross-Linked Chromatin Immunoprecipitation.

Firstly, the beads were blocked before immunoprecipitation. Twice 30 µl beads per IP (Protein A/G PLUS-Agarose, Santa Cruz sc-2003) were washed twice in RIPA buffer (50

mM NaCl, 1% v/v NP-40, 0.5% w/v Na deoxycholate, 0.1% w/v SDS, 50 mM Tris-HCl, pH 8, 5mM EDTA) and resuspended in RIPA buffer. For pre-clearing, beads were resuspended in 900 µl RIPA buffer + PIC (Protease Inhibitor Cocktail Set III, EDTA-Free, Merck) + PMSF (5 µl/ml each) and kept on ice until needed. For IPs, beads were resuspended in 450 µl RIPA buffer with BSA (final conc. 1 mg/ml) and sheared salmon sperm DNA (final conc. 0.3 mg/ml), and incubated at 4˚C with rotation until needed. Then centrifuged 4,000 rcf 1 min 4˚C.

Cells were grown to 90% confluency and crosslinked with 1% formaldehyde/PBS for 10 min at room temperature shaking under the fume hood. The reaction was stopped by the addition of glycine (final conc. 125 mM). Cells were washed twice with ice-cold PBS + PIC (2 µl /10 ml) and scraped into ice-cold PBS + PIC + PMSF (5 µl/ml each). Cells were centrifuges 600 rcf 5 min 4˚C, pellets resuspended in RIPA buffer (50 mM NaCl, 1% v/v NP-40, 0.5% w/v Na deoxycholate, 0.1% w/v SDS, 50 mM Tris-HCl, pH 8, 5mM EDTA) + PIC + PMSF (5 µl/ml), sonicated 3 times 20 impulses, and centrifuged 14,000 rpm 10 min 4˚C. Into supernatant (lysate) was added RIPA buffer (10% inputs saved for later). IPs were blocked with beads for pre-clearing at 4˚C for 3 h with rotation. The pre-cleared lysate was centrifuged 4,000 rcf 1 min 4˚C, transferred to blocked beads, and immunoprecipitated with the appropriate antibodies (5 µl anti-H3 per reaction, 5 µl nonspecific IgG) at 4˚C overnight with rotation. The beads were centrifuged (4,000 rcf 1 min 4˚C), washed once with ice-cold RIPA buffer, and transferred into new tubes. Then, beads were washed for 5 min at 4˚C with rotation and centrifuged (4,000 rcf 1 min 4˚C). Once again washed with ice-cold RIPA buffer, then four times with Szack's IP buffer (100 mM Tris-HCl, pH 8.5, 500 mM LiCl, 1% v/v NP-40, 1% NA deoxycholate), twice again with RIPA and twice with TE buffer (10 mM Tris-HCl, pH8, 1 mM EDTA). Protein-DNA complexes were eluted in freshly prepared Taliandis elution buffer (70 mM Tris-HCl, pH 6.8, 1 mM EDTA, 1.5% w/v SDS) for 10 minutes at 65˚C shaking 1,200 rpm. Then, both 10% inputs and IPs were decrosslinked in the presence of 200 mM NaCl and treated with 20 mg proteinase K for 5 h at 65˚C. DNA was precipitated with 0.5 vol. of 7.5M NH_4Ac and 2 vol. of 100% ethanol at -80˚C over night, washed with 70% ethanol, and resuspended in nuclease-free water. Signals were quantified by qPCR and signal IP over Input ($2^{(Ct(input)-Ct(IP))}$) and normalized to a non-transcribed intergenic control region.

SRSF-Binding and Splicing Silencer Motif Prediction

SRSF protein and splicing silencer consensus motif distributions around splice sites were visualized using seqPattern Bioconductor R package (Gentleman et al. 2004). The references for SR binding consensus motifs with IUPAC nucleotide ambiguity codes are listed in the Supplementary Material. Windows of 200 bp (-100 to +100) were extracted around the 5'ss and 3'ss for both, PCGs and lincRNAs. The import.bed function was used from the Rtracklayer Bioconductor R package (Lawrence et al. 2009), and the corresponding sequences were extracted using the getseq function. The getPatternOccurrenceList function of seqpattern R package was used to obtain the position of each consensus pattern occurrence for each sequence. A total number of motifs was counted for each position for the given region. Binding site densities were calculated for each position separately as the number of motifs divided by the number of analyzed sequences. The visualization was done in R (version 3.5.1) with the ggplot2 (Wickham 2016) package for the final graphical output.

Cellular Fractionation

Cellular fractionation assays were performed as previously described (Pandya-Jones and Black 2009). Cells were grown to 90% confluency, washed with PBS and scraped into PBS/1mM EDTA. The pellet was resuspended in ice-cold NP-40 lysis buffer (10 mM Tris-HCl pH 7.5, 0.15% NP-40, 150 mM NaCl) for 5 min. Then, the lysate was placed on 2.5 volumes of an ice-cold sucrose cushion (24% sucrose in NP-40 lysis buffer) and centrifuged for 10 min at 4°C. The supernatant (cytoplasmic fraction) was stored at 4°C for subsequent RNA isolation. The pellet was washed with ice-cold PBS/1mM EDTA and resuspended in glycerol buffer (50% glycerol, 20 mM Tris-HCl pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 0.85 mM DTT, and 0.125 mM PMSF). An equal volume of nuclei lysis buffer (10 mM HEPES pH 7.6, 1 mM DTT, 7.5 mM $MgCl_2$, 0.2 mM EDTA, 0.3 M NaCl, 1M urea, 1% NP-40) was added, samples were incubated for 2 min on ice and centrifuged for 2 min at 4°C. The supernatant (soluble nuclear fraction) was stored at 4°C for subsequent RNA isolation. The pellet (chromatin fraction) was washed with ice-cold PBS/1mM EDTA. RNA from all fractions was isolated using the TRIzol reagent as described above.

SiRNA Treatment

Pre-annealed siRNA duplexes were obtained from Ambion - hRNP H1 (s6728): 5' GAAGCAUACUGGUCCAAAUtt 3', ncRNA-a2: 5' CCTCCTTACTCTTGGACAAtt 3', ncRNA-a5: 5' CCTTGGAGAATAAAGCTTAtt 3'. The negative control # 5 siRNA from Ambion was used as a negative control. SiRNAs were transfected with Oligofectamine (Thermo Fisher Scientific) at the final concentration of 50 nM according to the manufacture's protocol. Cells were incubated for 72 h (hnRNP H) and 48 h (ncRNA-as) and then harvested and analyzed. After siRNA treatment of ncRNAs, the expression of PCGs in in their vicinity were evaluated – KDM5B (ENSG00000117139), RABIF (ENSG00000183155), KLHL12 (ENSG00000117153), ADIPOR1 (ENSG00000159346), PQLC3 (ENSG00000162976), ROCK2 (ENSG00000134318), E2F6 (ENSG00000169016).

RNA Immunoprecipitation

Cells were grown to 80-90% confluency and 24 hours after transfection with various ncRNA-a2 constructs. Cells were washed with PBS and scraped into 2 ml PBS. Two ml of nuclear isolation buffer (1.28 M sucrose, 40 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$, 4% Triton X-100) and 6 ml of water were added, and cells were incubated 20 min on ice with frequent mixing. Nuclei were pelleted by centrifugation at 2,500g for 15 min and resuspended in 1 ml RIP buffer (150 mM KCl, 25 mM Tris-HCl pH 7.4, 5 mM EDTA, 0.5 mM DTT, 0.5% NP40) with freshly added 100 U/ml RNasin (Promega) and 5 µl Protease Inhibitor Cocktail Set III, EDTA-Free (Calbiochem). Then, nuclei were split into two 500 µl fractions (IP, mock) and mechanically sheared by a dounce homogenizer with 3 times 20 strokes (0.5 s, 4% amplitude). Nuclear membranes and debris were removed by centrifugation at 13,000 rpm for 10 min. Supernatants were transferred into siliconized tubes, and 10% was frozen and stored at -80˚C for RNA/protein isolation (10% inputs). Antibodies were added (IP:2 µg of U2AF2 – Santa Cruz sc-53942, mock: 4 µg of IgG from mouse serum – Sigma I5381) to the remaining supernatants and samples were incubated at 4˚C overnight with gentle rotation. Then, 40 µl of Protein G PLUS agarose beads (Santa-Cruz Biotechnology, sc-2002) were added to the lysates and further incubated for 1h at 4˚C with gentle rotation. Beads were pelleted at 2,500 rpm for 30 s and washed 3 times with 500 µl RIP buffer, followed by one additional wash with PBS. Co-precipitated and input

RNA and proteins were isolated by resuspending the beads in 1 ml (10% inputs) or 500 µl (IPs) TRIzol reagent (LifeTechnologies). RNA and proteins were isolated and analyzed as described above.

CRISPRa and CRISPRi

For designing short-guide RNAs (sgRNAs), online CRISPR Design Tool (Hsu et al. 2013) (http://crispr.mit.edu/) was used. SgRNA targeting the 5'end of the *ncRNA-a2* gene was cloned into pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene plasmid #42230) (Cong et al. 2013) using BbsI restriction sites for testing guide efficiency or into edited lentiGuide-Puro (Addgene plasmid #52963) (Sanjana et al. 2014) (the sequence coding puromycin N-acetyltransferase was swapped with sequence coding mCherry protein) using BsmbI restriction sites for transfection. Guide target sequence for testing guide efficiency was cloned into pARv-RFP (Red Fluorescent Protein, a gift from Radislav Sedláček, Addgene plasmid # 60021) (Kasparek et al. 2014) using EcoRV and PvuI restriction sites with the introduction of the BamHI restriction site at 5'end of a guide target sequence to allow restriction digest analysis of positive clones. For delivering dCas9 protein (catalytically dead Cas9) tagged with 10 copies of the GCN4 peptide v4 and BFP, pHRdSV40-dCas9-10xGCN4_v4-P2A-BFP (Addgene plasmid #60903) (Tanenbaum et al. 2014) was used and for delivering transcriptional activation domain VP64, plasmid encoding an antibody that binds to the GCN4 peptide from the SunTag system, pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS (Addgene plasmid #60904) (Tanenbaum et al. 2014) was used. For delivering transcriptional repression domain KRAB (Krüppel associated box) fused with dCas9, pHR-SFFV-dCas9-BFP-KRAB (Addgene plasmid #46911) (Gilbert et al. 2013) was used. As an empty vector control, pHRdSV40-dCas9-10xGCN4_v4-P2A-BFP (Addgene plasmid #60903) (Tanenbaum et al. 2014) was used, but the sequence for the VP64 domain was cut out. All sequences were confirmed by DNA sequencing. The sequences of all sgRNAs and guide target sequences used in this study are listed in the Supplementary Material.

Cells were grown on PD6 to 70% confluency, and an equimolar mixture of plasmids (in the case of activation domain: 2 µg of pHR-dCas9-GCN4-BFP, pHR-scFv-GCN4-sfGFP-VP64, lentiGuide_mCherry_Guide, in the case of repressor domain: 3 µg pHR-SFFV-dCas9-BFP-KRAB, lentiGuide_mCherry_Guide) was transiently transfected into HeLa cells using the Lipofectamine 3000 Transfection Reagent (Thermo Fisher

Scientific) according to the manufacturer's instructions and incubated for 72 h. The culture medium was changed 6 h after transfection. 72 h after transfection, cells were FACS sorted for BFP, GFP and RFP triple positivity (in the case of activation domain) or BFP and RFP double positivity (in the case of repressor domain). Total RNA and proteins were isolated as described above. RNA was quantified by qRT-PCR, and the presence of proteins was checked by Western Blot.

CRISPR/Cas9-Mediated poly(A) Knock-In

For designing short-guide RNA (sgRNA), online CRISPR Design Tool (Hsu et al. 2013) (http://crispr.mit.edu/) was used. SgRNA targeting 58-80 nt downstream of ncRNA-a2 TSS was cloned into pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene plasmid #42230) (Cong et al. 2013) using BbsI restriction sites. Guide target sequences for testing guide efficiency was cloned into pARv-RFP (Red Fluorescent Protein, a gift from Radislav Sedláček, Addgene plasmid # 60021) (Kasparek et al. 2014) using EcoRV and PvuI restriction sites with the introduction of the BamHI restriction site at 5'end of a guide target sequence to allow restriction digest analysis of positive clones. The homology-directed repair (HDR) template including two pole(A) sites and 800bp homology arms flanking the targeted *ncRNA-a2* sequence was amplified from genomic DNA and cloned into the pBluescript II vector. Before transfections, HDR template was amplified by PCR with specific primers. All sequences were confirmed by DNA sequencing. The sequences of sgRNA, guide target sequences, and inserted sequences with pA sites used in this study are listed in the Supplementary Material.

Cells were grown to 90% confluency, and an equimolar mixture of plasmids (8 µg of pX330_Cas9_Guide, pARV_GuideTargetSequence, HDR template) was transiently transfected into HeLa cells in Opti-MEM™ I Reduced Serum Medium (Thermo Fisher) using the Lipofectamine® 3000 Transfection Reagent (Thermo Fisher Scientific) according to the manufacturer's instructions and incubated for 72 h. The culture medium was changed 6 h after transfection and then every 24 h. The inhibitor of non-homologous end-joining (NHEJ) (SCR7, final concentration 1 µM, Xcess Biosciences) was added 16 h prior to transfection and then every 24 h with changing the medium. 72 h after transfection, cells were FACS sorted for GFP and RFP positivity in single-cell mode. Single cells were grown for 2 weeks until full confluency in a 1:1 fresh/conditioned medium. Positive clones were selected by PCR (sequences are provided in Supplementary Material).

Isolation of Biotin-Labeled Nascent Transcripts

Cells were grown to 80-90% confluency and provided with fresh media containing 500 µM 4-Thiouridine (4-sU; Sigma). Cells were pulsed labeled for 60 min, and RNA was extracted by TRIzol (Thermo Fisher Scientific) as described before. To biotinylate 4-sU-labeled RNAs, 120 µg total RNA was mixed with 240 µl of 4 mM EZ-Link® HPDP-Biotin (Thermo Fisher Scientific), 120 µl biotinylation buffer (10 mM HEPES pH 7.5, 1 mM EDTA) and 840 µl Nuclease-Free Water (Ambion). The samples were incubated in the dark for 90 min at room temperature. Total RNA including 4sU-Biotin-labeled RNA was extracted by adding 250 µl phenol:chloroform and precipitated overnight with 2.5 vol. 100% ethanol, washed with 70% ethanol and resuspended in 100 µl Nuclease-Free Water (Ambion). 4sU-Biotin-labeled RNA was captured on 100 µl BcMag^{TM} Streptavidin Magnetic Beads (Bioclone Inc), washed twice with washing buffer (0.5 M NaCl, 20 mM Tris-HCl pH 7.5, 1 mM EDTA), eluted with washing buffer containing 0,1 mM DTT (Invitrogen) and precipitated with 2.5 vol. of 100% ethanol. Total RNA that did not bind to magnetic beads served as input. Reverse transcription and quantitative PCR was done as described before.

Results

This section is composed of my published and unpublished results about long non-coding RNAs, and the relationship between chromatin modifications and splicing.

BIEBERSTEIN, N. I., KOZÁKOVÁ, E., HURANOVÁ, M., THAKUR, P. K., KRCHŇÁKOVÁ, Z., KRAUSOVÁ, M., CARRILLO OESTERREICH, F., STANĚK, D. 2016. **TALE-directed local modulation of H3K9 methylation shapes exon recognition**. Sci Rep. 6: 29961.

This project was published in Scientific Reports journal (impact factor: 4.122 for year 2017), and I participated in this project by showing that JMJD2D N202M (catalytically not active) mutant is not affecting the splicing of *FOSL1* gene (ionomycin treatment, RNA and protein isolation, RT-qPCR, WB – for more detail see Material and Methods).

KRCHŇÁKOVÁ, Z., THAKUR, P. K., KRAUSOVÁ, M., BIEBERSTEIN, N. I., HABERMAN, N., MÜLLER-MCNICOLL, M., STANĚK, D. 2018. **Splicing of Long Non-Coding RNAs Primarily Depends on Polypyrimidine Tract and 5' Splice-Site Sequences Due to Weak Interactions with SR Proteins.** NAR. doi: 10.1093/nar/gky1147. [Epub ahead of print]

This project was published in Nucleic Acids Research journal (impact factor: 11.561 for year 2017), and I performed all experiments presented here if not stated otherwise.

Mutual Regulation of Chromatin and Splicing

Role of Histone Modifications on Alternative Splicing

Currently, it is widely accepted that transcription, chromatin and pre-mRNA splicing are functionally coupled. Since the splicing of introns takes place predominantly co-transcriptionally (Carrillo Oesterreich et al. 2010; Tilgner et al. 2012; Brugiolo et al. 2013), it is not so surprising that chromatin can influence splicing as well. Furthermore, it has been previously shown that histone post-translational modifications and promoter sequences can significantly influence the splicing outcome (Hnilicová et al. 2011; Dušková et al. 2014; Salton et al. 2014; Curado et al. 2015; Bieberstein et al. 2016). First studies implying that chromatin can affect splicing show the correlation of the presence of certain histone modification with exon inclusion rates (Schwartz et al. 2009; Tilgner et al. 2009). However, determining whether these histone marks are causative for splicing changes or simply a consequence of splicing cannot be addressed by this correlative approach. Previously, most studies have used a global approach to perturb histone modifications genome-wide, using small molecule inhibitors or knockdown/overexpression of histone modifying enzymes (HMEs) (Sims et al. 2007; Luco et al. 2010; Hnilicová et al. 2013; Guo et al. 2014; Salton et al. 2014). As this method globally affects the transcriptional program of the cell, secondary effects cannot be fully excluded. Because of that, we set out to determine the role of chromatin by direct, local manipulation of the chromatin environment at a selected target exon. We used TALE (Transcription-Activator-Like-Effector) domains to tether HMEs to a target exon locus, analogous to TALENs for genome engineering or TALE associated with transcription factors for transcriptional regulation (Miller et al. 2010; Cermak et al. 2011; Cermak et al. 2015). Importantly, this strategy has been successfully utilized to modify chromatin modifications at enhancers (Mendenhall et al. 2013).

To test whether locally restricted changes in the chromatin context can directly affect splicing of an alternative exon, we altered H3K36 and H3K9 methylation at the EDB exon in human fibronectin (*FN1*) (Bieberstein et al. 2016) since this gene is one of the first described and till nowadays a widely used model gene for studying alternative splicing (Kornblihtt et al. 1984). It has three regions undergoing alternative splicing. Two are cassette exons EDB and EDA (sometimes also called EIIIB and EIIIA) which are either

included or excluded and the region located near the 3'end of the gene called IIICS. As mentioned before, histone modifications (e.g. methylation of H3K9 and H3K36) were previously proposed to regulate alternative splicing (Luco et al. 2010; Saint-Andre et al. 2011; Pradeepa et al. 2012; Guo et al. 2014; Salton et al. 2014), and splicing of the EDB exon was described to be sensitive to the chromatin environment (Pagani et al. 2003; Hnilicová et al. 2011; Dušková et al. 2014). To specifically and locally target chromatin around the EDB exon, a TALE domain recognizing a region close to 3'ss of EDB exon was designed by the TAL Effector Targeter25 (https://tale-nt.cac.cornell.edu/node/add/single-tale) with no predicted off targets in the human genome. To modulate H3K9 methylation, the TALE domain was fused to the H3K9 tri-methyltransferase Suv39H1, the catalytic domain of the di-methyltransferase G9a (EHMT2) or the H3K9me2/3 demethylase JMJD2D (KDM4D) (Rea et al. 2000; Tachibana et al. 2001; Whetstine et al. 2006; Wu et al. 2010). For H3K36me3 we selected the methyltransferase ASH1L and the catalytically active domain of SETD2 (Sun et al. 2005; An et al. 2011). The HA-tag was included as a linker between the TALE domain and HME to allow detection of fusion proteins by Western blot. As a control, the TALE domain was fused to GFP, and all results were normalized or compared to the TALE-GFP construct in order to exclude unspecific effects due to the binding of the TALE construct to DNA. The levels of H3K36 and H3K9 methylation were monitored by chromatin immunoprecipitation (ChIP). All histone modifying domains (G9a, Suv39H1, JMJD2D, and SETD2) were able to influence the methylation signals over EDB exon. More importantly, such changes were specific mainly for EDB exon itself (only in the case of Suv39H1 exon 38 downstream of EDB and in the case of SETD2 also exon upstream of EDB was affected) (Bieberstein et al. 2016). The effect of chromatin alteration on alternative splicing was assessed using RT-qPCR. Majority of changed chromatin signals had a significant influence on alternative splicing of EDB exon demonstrating that the local level of H3K9 methylation can directly impact splicing and that higher methylation promotes exon inclusion. Furthermore, this proves the principle that a local change in the chromatin environment impacts alternative splicing.

Given the positive effect of local H3K9me3 on alternative exon inclusion (Saint-Andre et al. 2011; Schor et al. 2013; Bieberstein et al. 2016), we speculated this histone modification might help to recognize exons in general. Moreover, we have shown using publicly available ChIP-Seq data from three different human cell lines that H3K9me3 was

depleted around transcriptional start sites and poly(A) sites and enriched around internal exons in all tested cells (Bieberstein et al. 2016). To experimentally test the effect of H3K9 methylation on splicing efficiency of constitutive exons, we analyzed co-transcriptional splicing of the exon 3 in *FOSL1* gene after targeting H3K9 methyltransferases (G9a, Suv39H1) and demethylases (JMJD2D) by TALE-HMEs binding exon 3. This exon is surrounded by weak splice sites, and H3K9me3 might promote its inclusion similarly to the alternative EDB exon. To monitor the splicing efficiency of the constitutive exon 3 in *FOSL1*, we analyzed co-transcriptional splicing of nascent RNA to avoid potential artefacts due to the high amount of spliced mRNA present in total RNA samples. A primer downstream of the poly(A) cleavage site for reverse transcription was used, thereby only nascent transcripts still attached to chromatin through the transcription machinery were selected (Pandya-Jones and Black 2009; Carrillo Oesterreich et al. 2010; Bieberstein et al. 2016). The ratio of pre-mRNA to mRNA revealed an accumulation of unspliced transcripts after JMJD2D overexpression and depletion of H3K9me3 (**Figure 18**). In contrast, tethering of methyltransferases G9a or Suv39H1 did not further improve co-transcriptional splicing suggesting that naturally occurring H3K9 methylation is sufficient to promote FOSL1 splicing, and increasing H3K9 methylation does not further enhance splicing efficiency. Together, these data demonstrate that H3K9me3 has a functional role in the splicing of constitutive exons.

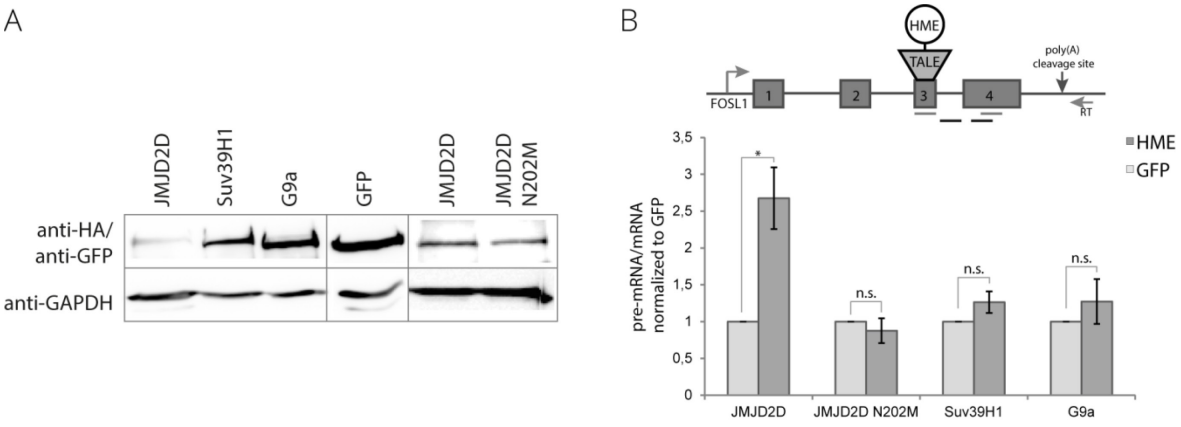

Figure 18. Histone modification can influence splicing. (**A**) Transient expression of TALE-HME constructs targeting FOSL1 exon was confirmed by Western blot using an anti-HA antibody against the HA-linker or GFP in the case of TALE-GFP. (**B**) Ratios of unspliced pre-mRNA to spliced mRNA are normalized to TALE-GFP. The mean of three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by the two-tailed Student's T-test, *p<0.05, **p<0.01, ***p<0.001 (published in Bieberstein et al. 2016).

Role of DNA Enhancer Element on Alternative Splicing

Because of the observed mutual relationship between transcription, chromatin and RNA splicing, it was not surprising to see studies showing that promoter identity and sequence can significantly affect alternative splicing (Cramer et al. 1997; Cramer et al. 1999; Auboeuf et al. 2002; Pagani et al. 2003). Promoters were shown to contain elements that can control alternative splicing independently of transcription regulation and that this regulation involves chromatin acetylation (Dušková et al. 2014). Inside cells, the promoter activity is regulated by enhancer elements. Therefore, we decided to test whether endogenous transcription enhancers modulate alternative splicing.

As a model gene, we again selected the *FN1* gene. Putative transcription enhancer element of *FN1* gene was predicted by two publicly available programs: FANTOM5 (Lizio et al. 2015) and The Ensemble Regulatory Build (Zerbino et al. 2015) which combine transcription factor-binding information, ChIP-Seq data of various histone modifications, binding of various regulatory proteins (e.g. CTCF, p300), and DNaseI hypersensitivity sites. Both programs identified a region of ~2400 bp located ~23 kilobases upstream of *FN1* TSS (Volek 2018). To investigate the role of such transcription enhancer on splicing, we deleted whole ~2400 bp region by CRISPR/Cas9 (Volek 2018) and selected three different cell lines by two-step genotyping PCR (**Figure 19**). All of the three selected cell lines have a whole region deleted (**Figure 19A**), but they also contain a WT allele, so all of them are heterozygotes (**Figure 19B**).

Alternative splicing of EDB exon was assayed in enhancer-deleted cell lines by semiquantitative RT-PCR. Transcripts with included EDB exon were detected primarily in enhancer-deleted lines demonstrating that alternative splicing of EDB exon is altered after the deletion of the enhancer element (**Figure 20A**). In addition, we observed a reduced amount of both splicing isoforms after deletion of the enhancer, which is consistent with the transcription enhancing the function of the deleted genomic element (different signal in 25. cycle between WT and mutants) (**Figure 20A**).

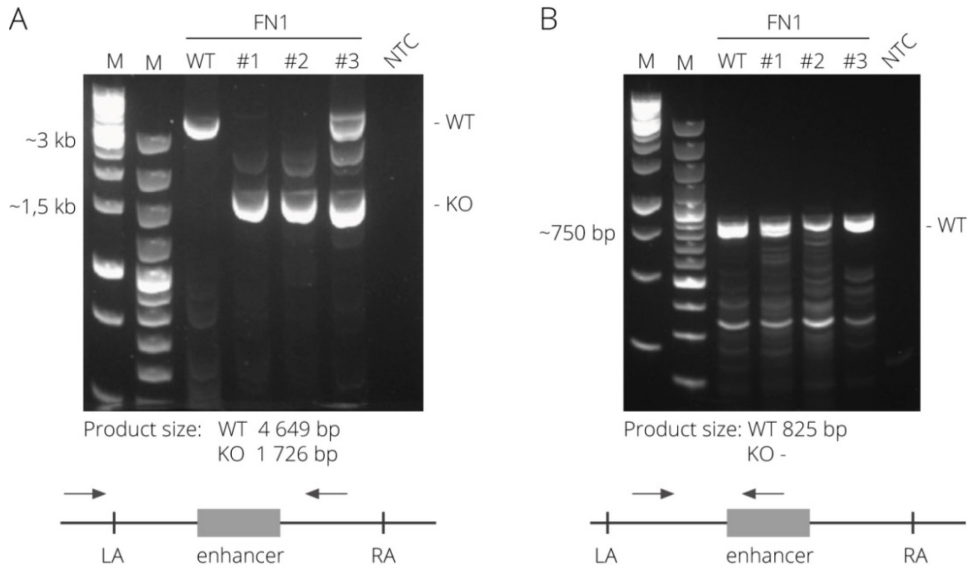

Figure 19. Three cell lines with deleted FN1 transcription enhancer. (**A**) The first round of selection PCR from genomic DNA using a set of primers (bottom scheme) that one of them recognizes sequence outside the homology template. By this we confirmed the deletion is within endogenous FN1 locus; M – marker, NTC – non-template control. (**B**) The second round of selection PCR from genomic DNA using a set of primers (bottom scheme) that one of them recognizes a sequence within the transcription enhancer region. By this we determined that all cell clones also contain WT allele; M – marker, NTC – non-template control (unpublished data).

To confirm the previous results, we have measured the ratios of included over excluded EDB and EDA exons by RT-qPCR. In the case of EDB exon, we have observed the significant increase in EDB inclusion in deletion mutants compared to WT while splicing of EDA exon seems to be unaffected by enhancer deletion (**Figure 20B**). In addition, the relative transcription of the *FN1* gene was significantly decreased in all mutants when compared to WT confirming the previous data from semiquantitative RT-PCR (**Figure 20C**). The lower transcription of the *FN1* gene in mutants is also consistent with the deleted region to be indeed a functional transcription enhancer element.

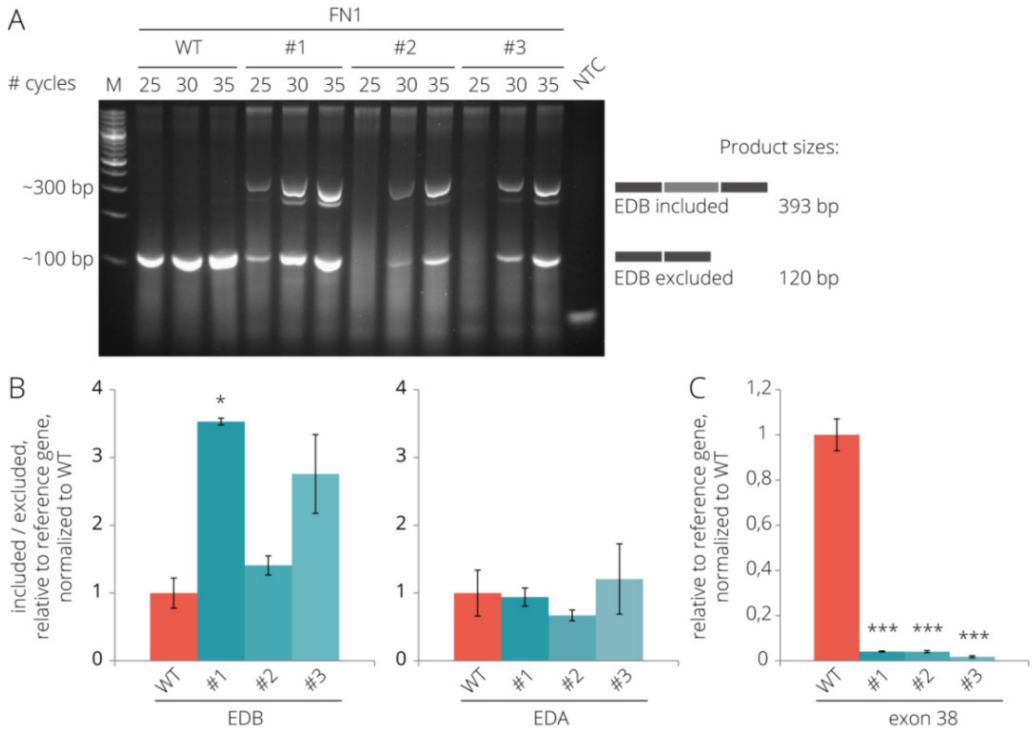

Figure 20. EDB exon of *FN1* gene is more efficiently spliced in FN1 enhancer-deleted cell lines than in WT. (**A**) Exon inclusion or skipping of EDB exon of *FN1* gene in WT and three FN1 enhancer-deleted cell lines measured by semiquantitative RT-PCR.; M – marker, NTC – non-template control. (**B**) The ratio of included over excluded of exon EDB (exon 25) and EDA (exon 33) in WT and three FN1 enhancer-deleted cell lines. (**C**) The relative expression of the *FN1* gene (exon 38) in WT and three FN1 enhancer-deleted cell lines. (**B-C**) Bar-plots show RNA levels determined by RT-qPCR. The mean of at least two independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by the two-tailed Student's T-test comparing the individual mutant with WT, *p<0.05, **p<0.01, ***p<0.001 (unpublished data).

It was previously shown that alternative splicing of the *FN1* gene can be altered by modulating histone modifications (Hnilicová et al. 2011; Dušková et al. 2014; Bieberstein et al. 2016), we decided to measure histone methylation levels along the *FN1* gene. We tested two histone marks previously associated with splicing changes (H3K36me3 and H3K9me3) and a histone modification H3K4me2 which overlaps with transcription factor binding regions (Wang et al. 2014). Interestingly, in deletion mutants, H3K4me2 levels are partially decreased over EDB exon, while the partial increase over promoter was observed in comparison to WT (**Figure 21A**). We also observed ~2-fold increase of H3K9me3 immediately downstream of deleted enhancer element when compared to WT (**Figure 21B**) indicating that this region could become heterochromatized and thus no longer functioning as transcription enhancer. Concomitantly, levels of H3K9me3 were significantly depleted over EDB exon in deletion mutants in comparison to WT.

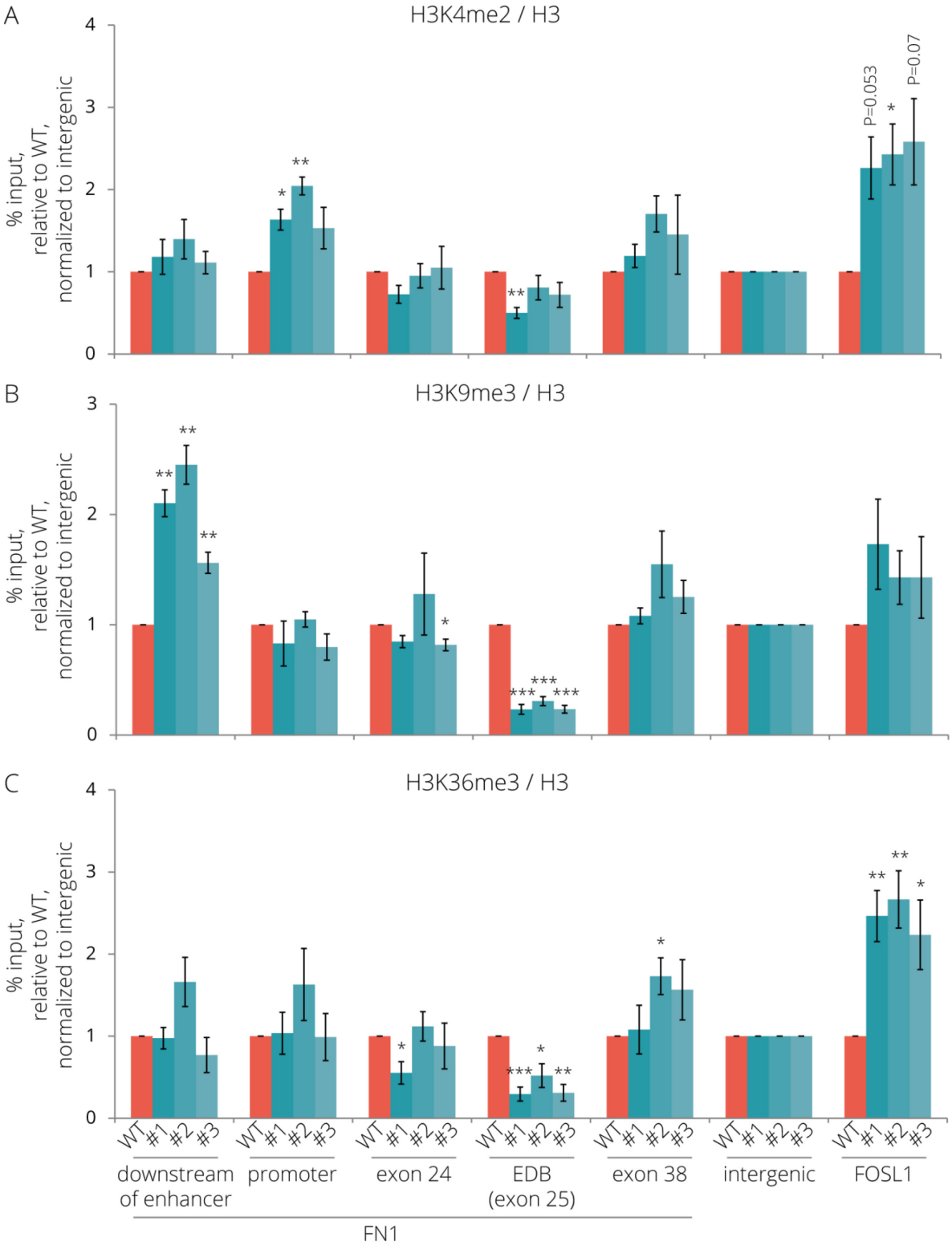

Figure 21. Deletion of transcription enhancer affects methylation levels over EDB exon of the *FN1* gene. Bar-plots show RNA levels determined by RT-qPCR. The mean of at least two independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by the two-tailed Student's T-test comparing the individual mutant with WT, *p<0.05, **p<0.01, ***p<0.001 (unpublished data).

Finally, we have looked on H3K36me3 which is a general mark of the active transcription. Surprisingly, H3K36me3 levels were not affected along the *FN1* gene upon enhancer deletion (**Figure 21C**), even though the overall FN1 transcription was substantially reduced (**Figure 20A, C**). The only exception was EDB exon where we observed a significant drop in H3K36me3 levels in mutants compared to WT (**Figure 21C**). This lower H3K36me3 levels can be a result of altered splicing upon enhancer deletion since it was shown that even though H3K36me3 marks are enriched over exons indicating its relation to splicing (Kolasinska-Zwierz et al. 2009), altered splicing can affect H3K36me3 levels as well (Kim et al. 2011).

Additionally to methylation, we also checked H4 acetylation and observed lower H4 acetylation over EDB exon (**Figure 22**) followed by the increase in EDB inclusion (**Figure 20**) which is consistent with previously shown action of acetylation in alternative splicing (Hnilicová et al. 2011). Interestingly, H4 acetylation was slightly elevated immediately downstream of deleted enhancer element and over the promoter region (**Figure 22**).

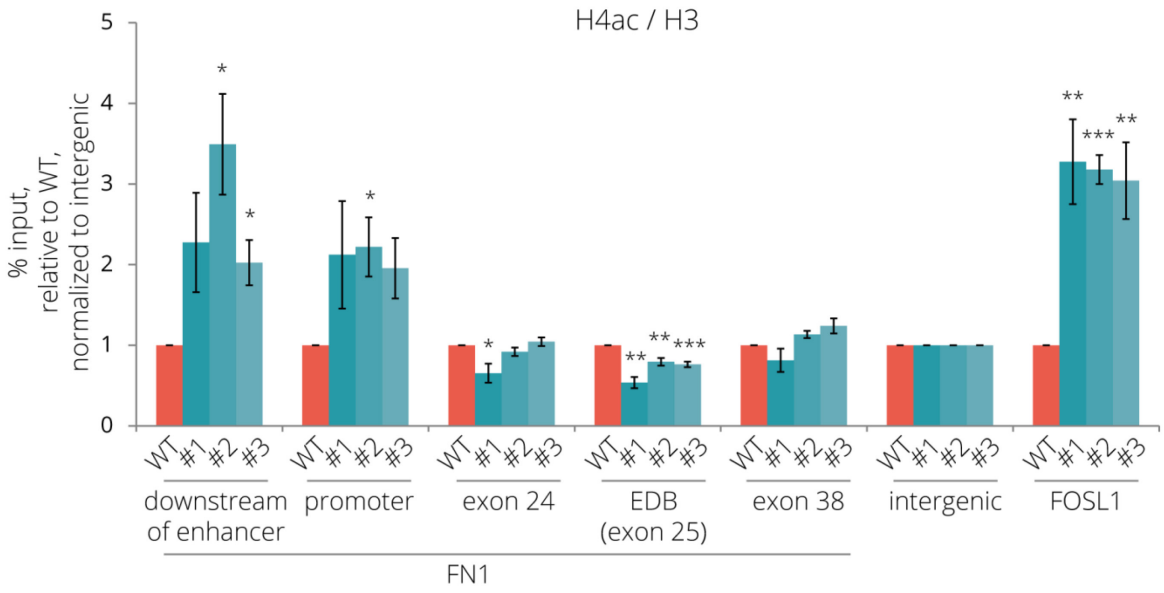

Figure 22. Deletion of transcription enhancer affects acetylation levels over EDB exon of the *FN1* gene. H4ac antibody detects Ser 1, Lys 5, Lys 8 and Lys 12 acetylated H4. Bar-plots show RNA levels determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by the two-tailed Student's T-test comparing the individual mutant with WT, *p<0.05, **p<0.01, ***p<0.001 (unpublished data).

Together, RT-qPCR and ChIP results suggest that our predicted DNA enhancer element located ~23 kbp upstream of *FN1* gene is truly a transcription enhancer since, upon its deletion, the overall FN1 transcription was significantly reduced. Additionally, the deletion of this region lead to altered alternative splicing of FN1 EDB exon, and at the same time, several histone modifications levels over this exon were affected. Because of that, we propose that alternative splicing can also be modified by enhancer element located several kilobases away of the particular exon, likely through the modulating histone modification marks.

Splicing of Long Intergenic Non-Coding RNAs

Several bioinformatic studies using various human or mouse cell lines have reported that, on general, lncRNAs/lincRNAs are less efficiently spliced than pre-mRNAs (Tilgner et al. 2012; Mukherjee et al. 2016; Lagarde et al. 2017; Melé et al. 2017; Schlackow et al. 2017). This was observed for both steady-state as well as nascent RNAs. One possible mechanism explaining the apparent difference in the splicing efficiency between lincRNAs and PCGs is the absence of proximal RNA Pol II phosphorylation over 5'ss in lincRNA transcripts (Mukherjee et al. 2016). However, the precise molecular mechanism for this phenomenon has not been elucidated yet.

In this study, we combined bioinformatic and experimental approaches to determine *cis*- and *trans*-acting factors that are responsible for the poor splicing of intergenic lncRNAs (lincRNAs). We analyzed available RNA-Seq data from five different human cell lines (embryonic stem cells H1-hESC, lung carcinoma A549, cervix carcinoma HeLa, liver carcinoma HepG2 and breast cancer cell line MCF7). To avoid a potential overlap with PCGs, we focused on lincRNAs only. Splicing indices (a proxy for splicing efficiencies) were then determined for each individual intron in each cell line separately. Although the lincRNA expression is highly cell-specific, we found that lincRNAs were, similarly to previous analyses, less efficiently spliced (have lower splicing indices) compared to PCGs in all tested cell lines (data not shown).

Because of the inefficient splicing of lincRNAs, we asked if lincRNAs possess some inhibitory factors that would affect their splicing. Therefore, we analyzed cumulative lengths of exons and introns and found that lincRNAs contain only slightly longer introns and exons than PCGs (Krchňáková et al. 2018). Because the length of introns has been associated with splicing efficiency (Klinz and Gallwitz 1985; Sterner et al. 1996; Bell et al. 1998; Fox-Walsh et al. 2005; Dewey et al. 2006; Louloupi et al. 2018), longer introns can partially explain lower splicing efficiency of lincRNAs. We also looked at the probability of lincRNAs to form secondary structures and did not find any relevant differences between lincRNAs and PCGs suggesting that RNA secondary structure is not the major factor that would determine the splicing difference between PCGs and lincRNAs (Krchňáková et al. 2018). Finally, we analyzed the presence of known splicing inhibitory sequences 100bp upstream and downstream of 3'ss. We found only one (out of 12) inhibitory motif (GTAGGT) enriched in lincRNAs over PCGs (**Figure 23** and **Figure 24**)

(Krchňáková et al. 2018). Together these results indicate that except longer introns, lincRNAs do not contain any particular feature that would specifically inhibit their splicing.

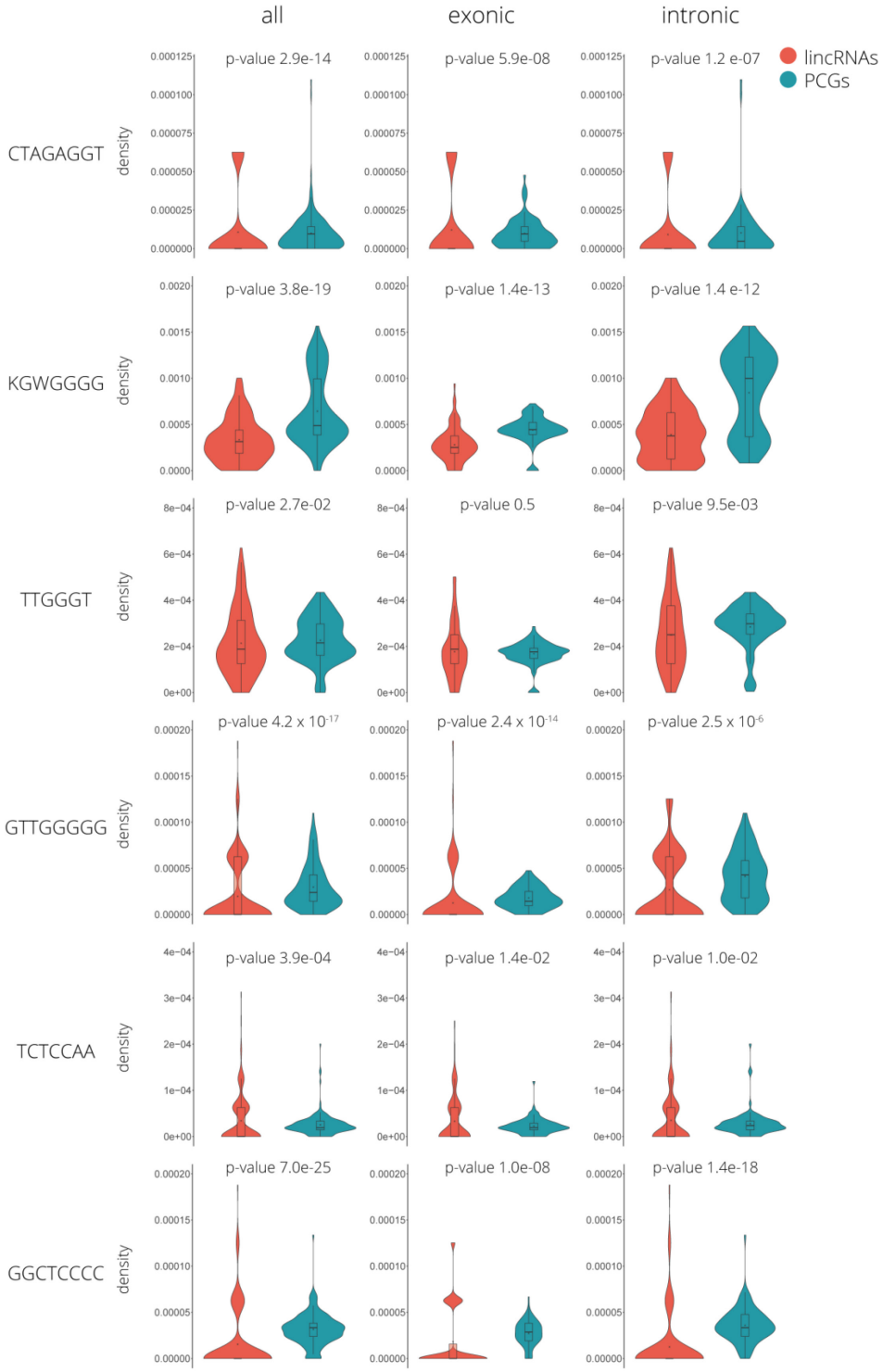

Figure 23. LincRNAs do not have more splicing silencer motifs. Splicing silencer motif distributions 100 nt upstream (intronic) and 100 nt downstream (exonic) of 3'ss in lincRNAs and PCGs. Motifs are taken from Sironi et al. (2004). P-values are calculated by Wilcoxon rank sum test (published in Krchňáková et al. 2018).

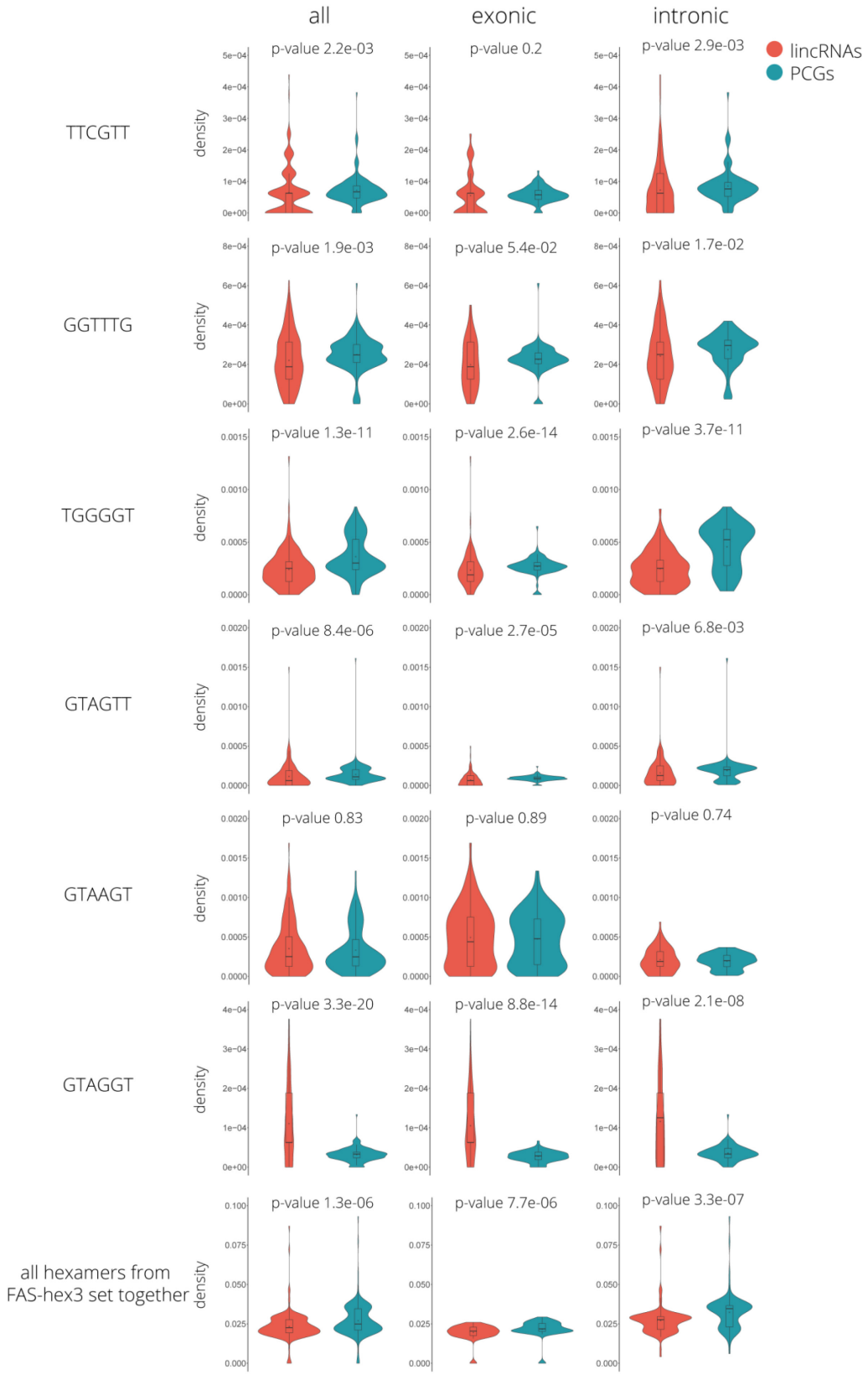

Figure 24. LincRNAs do not have more splicing silencer motifs. Splicing silencer motif distributions 100 nt upstream (intronic) and 100 nt downstream (exonic) of 3'ss in lincRNAs and PCGs. Motifs are taken from Wang et al. (2004b) and FAS-ESS web server (http://genes.mit.edu/fas-ess/). P-values are calculated by Wilcoxon rank sum test (published in Krchňáková et al. 2018).

To study the splicing efficiency of lincRNAs, we selected two activating lncRNAs, ncRNA-a2 (PCAT6) and ncRNA-a5 (LINC00570), which stimulate the expression of PCGs located in their genomic vicinity (Ørom et al. 2010). It should be noted that three different transcripts of the ncRNA-a2 (PCAT6) gene and four different transcripts of the ncRNA-a5 (LINC00570) gene are annotated in the Ensembl database (http://www.ensembl.org/index.html). However, in HeLa cells, we detected only two variants produced by alternative usage of 3'ss separated by 114 nt producing two ncRNA-a2 transcripts (PCAT6-201, PCAT6-202) (**Figure 25**, **Figure 26**). Only one ncRNA-a5 transcript (LINC00570-201) is supported by multiple ESTs in the Ensembl database and annotated in the NCBI Reference Sequence database (www.ncbi.nlm.nih.gov/refseq/). Therefore we focused our analysis on two ncRNA-a2 transcripts (PCAT6-201, PCAT6-202) and the second intron of the LINC00570 transcript (LINC00570-201).

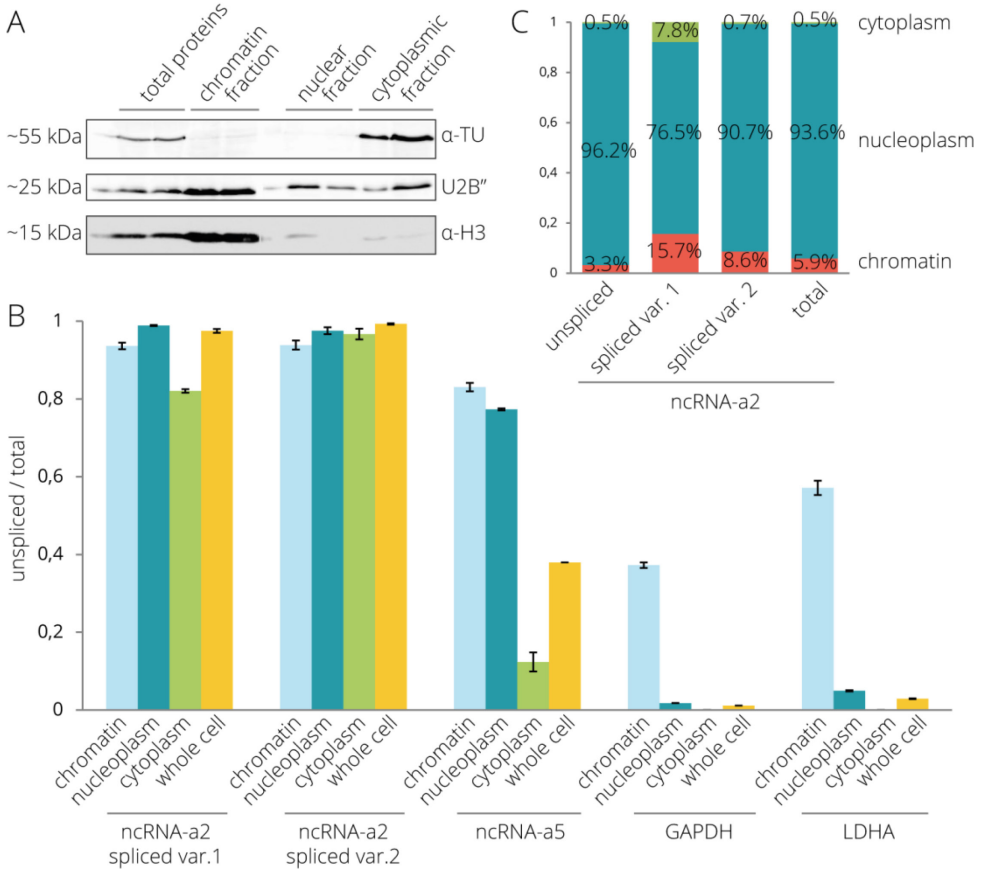

Figure 25. NcRNA-as are less efficiently spliced than PCGs. (**A**) Western blot of total proteins and proteins from cellular fractionations (two biological replicates). Anti-α-tubulin, anti-U2B", and anti-histone3 antibodies detect the cytoplasmic, nucleoplasmic, and chromatin fractions, respectively. (**B**) Splicing efficiencies in different cellular fractions. Bar-plots show RNA levels determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM. (**C**) The cellular distribution of ncRNA-a2 transcripts (published in Krchňáková et al. 2018).

To determine their splicing status in different cellular compartments, we fractionated HeLa cells into chromatin, nucleoplasmic and cytoplasmic fractions (**Figure 25A**). Using reverse transcription coupled with quantitative PCR (RT-qPCR), we found that nuclear fractions contained predominantly unspliced forms of both lncRNAs (**Figure 25B**). Strikingly, ~80% of cytoplasmic ncRNA-a2 retained the intron, compared to only ~10% for ncRNA-a5 transcripts, revealing large differences in splicing efficiencies between lncRNAs. In contrast, unspliced pre-mRNAs of two PCGs, *GAPDH* and *LDHA*, were only detected in the chromatin fraction (**Figure 25B**). A more detailed analysis of ncRNA-a2 transcripts revealed that the upstream 3'ss was preferentially used, but, in general, splicing at both 3'ss was inefficient (**Figure 26**). In addition, ncRNA-a2 seems to reside primarily in the nucleus since ~76-96% of its transcripts are localized in the nucleoplasm and chromatin fractions (**Figure 25C**).

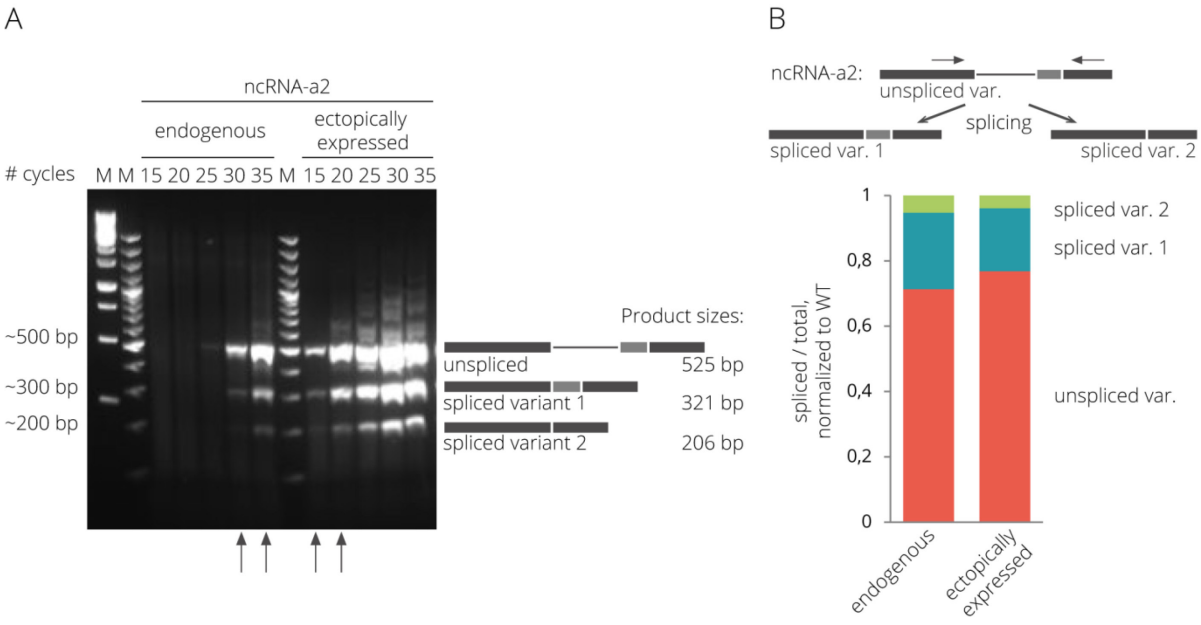

Figure 26. Neither chromatin nor promoter is extensively affecting the splicing inefficiency of ncRNA-a2. (**A**) Splicing efficiency of endogenous and ectopically expressed ncRNA-a2 measured by semiquantitative RT-PCR. Results are quantified from experiments indicated by arrows; M - marker. (**B**) Fraction calculation of ncRNA-a2 unspliced and spliced variants measured by semi-quantitative RT-PCR in (**A**). Primers are depicted as arrows above the transcript (published in Krchňáková et al. 2018).

We selected ncRNA-a2 for further analysis as an example of an inefficiently spliced lincRNA. It has been previously shown that chromatin modifications and promoter sequences can significantly influence the splicing outcome of a PCG (Hnilicová et al.

2011; Dušková et al. 2014; Salton et al. 2014; Curado et al. 2015; Nieto Moreno et al. 2015; Bieberstein et al. 2016). In order to determine whether promoter or chromatin elements influence *ncRNA-a2* splicing, we cloned the whole transcribed ncRNA-a2 sequence into a plasmid containing the CMV promoter, expressed ncRNA-a2 transiently in HeLa cells, and analyzed its splicing using semi-quantitative RT-PCR (**Figure 26**). We did not observe any significant differences in the splicing pattern between endogenous and transiently expressed ncRNA-a2. This suggests that the ncRNA-a2 sequence is the dominant factor affecting the efficiency of ncRNA-a2 splicing.

Intronic Sequences Determine the Inefficient Splicing of NcRNA-a2

To determine the contribution of exonic or intronic sequences to the observed splicing inefficiency, we swapped introns of *ncRNA-a2* and a PCG, human hemoglobin beta subunit (*HBB*) and transiently expressed chimeric transcripts (**Figure 27A**). We chose *HBB* intron 2 because our previous experiment showed its efficient splicing (unpublished data). We also observed efficient splicing of the *HBB* intron when inserted between *ncRNA-a2* exons (**Figure 27A**), and less than 1% of ncRNA-a2 transcripts with the HBB intron remained unspliced (**Figure 27B**). In contrast, the ncRNA-a2 intron remained largely unspliced when placed between HBB exons (**Figure 27A**). Since 5'ss sequences extend into the upstream exon, substituting just the intron changes the 5'ss strength of both hybrids. Therefore, we kept the original 5'ss of ncRNA-2 and HBB (8 bp downstream of the exon-intron boundary) and replaced only intronic sequences downstream (constructs "ncRNA-a2 with HBB intron-ncRNA-a2 5'ss" and "HBB with ncRNA-a2 intron HBB 5'ss"; **Figure 27A**). In both cases, keeping the original 5'ss sequence increased splicing efficiency of both constructs (1.2x for "ncRNA-a2 with HBB intron ncRNA a2 5'ss" and 2.1x for "HBB with ncRNA-a2 intron HBB 5'ss"). In the case of the ncRNA-a2 construct, the result is surprising because the original ncRNA-a2 5'ss has a much weaker MaxEnt score (MES) (5.28) than the artificial HBB/ncRNA-a2 5'ss (MES 10.90). This suggests that the 5'ss identity does not have a dominant impact on splicing of hybrid RNAs. While some contribution of exonic sequences cannot be ruled out, these results suggest that the ncRNA-a2 intron is largely responsible for the inefficient splicing of the ncRNA-a2 transcript.

To search for possible splicing regulatory elements, we prepared several deletion mutants of the 204 bp long ncRNA-a2 intron. We gradually deleted nucleotides either from the 5' end (F) or the 3' end (R) of the intron starting 5 nt downstream of the 5'ss and leaving 39 nt upstream of the 3'ss intact (**Figure 28**). The deletion of nucleotides 6-25 (mutant FΔ1) partially increased the splicing efficiency, but the majority of deletions reduced splicing efficiencies. We observed a particularly large drop in splicing efficiency when the central intronic region spanning nucleotides 66-125 was deleted (FΔ5 2.5x, FΔ6 15.5x, RΔ5 2.5x, RΔ6 4.9x reduction in splicing efficiency compared to WT). This could be explained either by the fact that the truncated intron is too short to be efficiently recognized by the splicing machinery or that the central intronic sequence contains elements that enhance splicing.

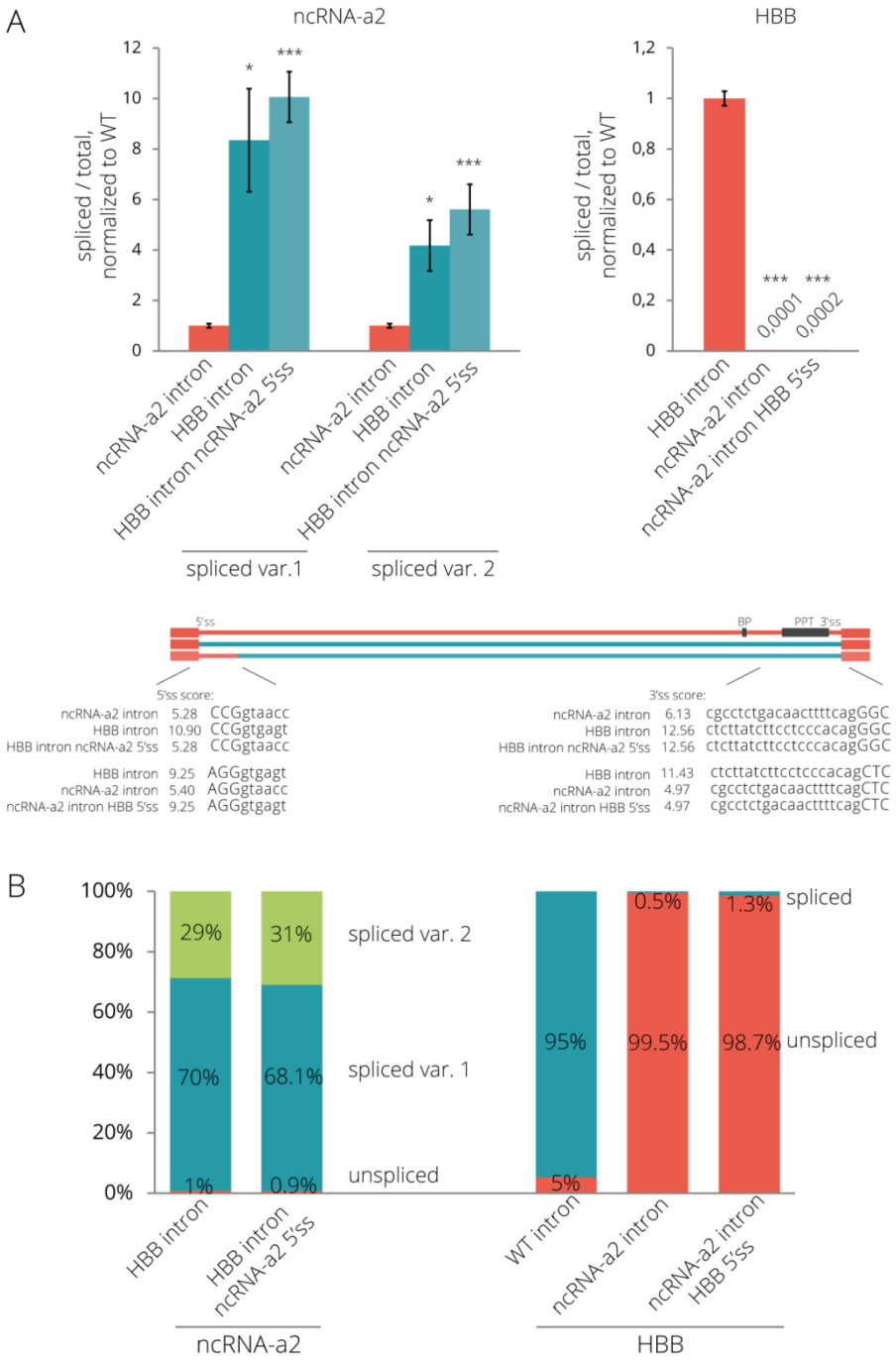

Figure 27. Intron sequence of ncRNA-a2 is a major determinant of its inefficient splicing. (**A**) The ncRNA-a2 intron is inefficiently spliced out when inserted into the human β-globin (HBB) pre-mRNA. Bar-plots show RNA levels determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by the two-tailed Student's T-test comparing the individual mutant with WT, *p<0.05, **p<0.01, ***p<0.001. (**B**) Fractions of spliced and unspliced transcripts of ncRNA-a2 and HBB after intron swapping. HBB intron 2 is efficiently spliced when placed between ncRNA-a2 exons. While ncRNA-a2 intron remains unspliced when placed between HBB exon (published in Krchňáková et al. 2018).

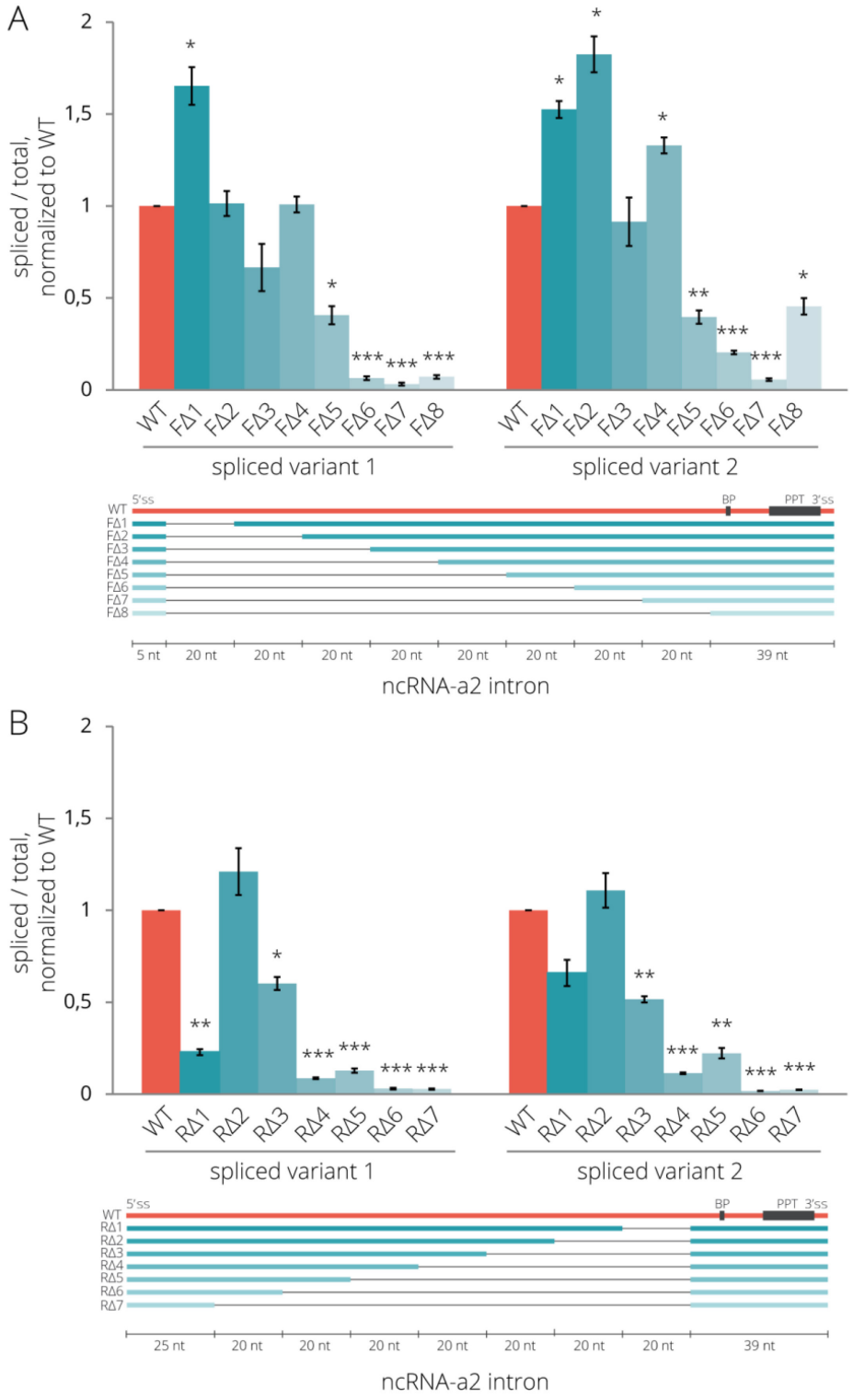

Figure 28. Splicing efficiencies of ncRNA-a2 forward (A) and reverse (B) intron deletion mutants. Splicing efficiencies are measured as a fraction of spliced transcripts relative to the total amount of transcripts. Schemes under the charts represent the intron sequence of ncRNA-a2 gene with predicted a branch point (BP) and the PPT. Bar-plots show RNA levels as determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test comparing the individual mutant with WT, *p<0.05, *p<0.01, ***p<0.001 (published in Krchňáková et al. 2018).

To distinguish this, we prepared additional deletion mutants (Δ1-7) and gradually removed a 20 nt sequence window along the intron (**Figure 29**). We observed a partial enhancement of splicing efficiency in mutants Δ1 and Δ4 (nucleotides 6-25 and 66-85) suggesting that these sequences could act as weak splicing silencers. The splicing efficiency significantly decreased in mutants Δ5 (5x) and Δ6 (11x for spliced variant 1 and 5x for spliced variant 2 compared to WT) indicating that this intronic sequence harbors splicing enhancer(s). Indeed, this sequence contains several G-rich sequences (**Figure 29**) that were previously characterized as splicing enhancers that recruit U1 snRNP and hnRNP F/H proteins when located downstream of 5'ss (McCullough and Berget 1997; Chou et al. 1999; McCullough and Berget 2000; Wang et al. 2007; Xiao et al. 2009; Wang et al. 2011a).

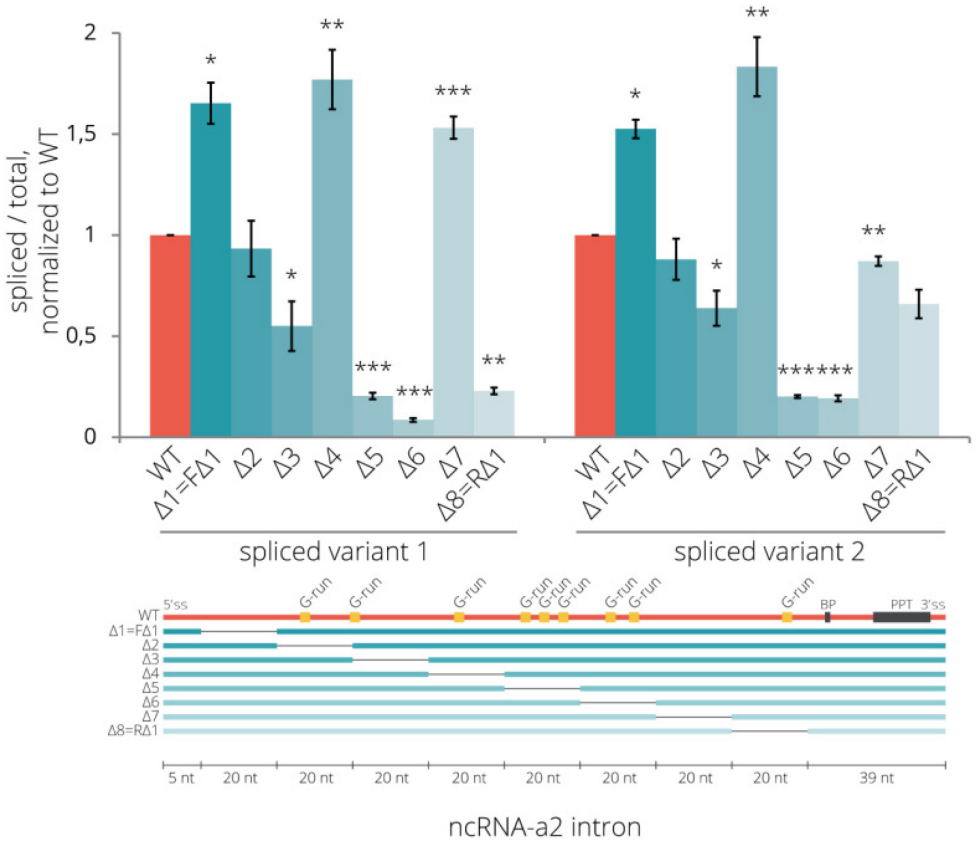

Figure 29. Surprisingly middle region of ncRNA-a2 harbors G-rich intronic splicing enhancers. Splicing efficiencies are measured as a fraction of spliced transcripts relative to the total amount of transcripts. Schemes under the charts represent the intron sequence of ncRNA-a2 gene with predicted a branch point (BP), the PPT and G-run motifs (yellow). Bar-plots show RNA levels as determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test comparing the individual mutant with WT, *p<0.05, *p<0.01, ***p<0.001 (published in Krchňáková et al. 2018).

To test whether hnRNP H protein enhances ncRNA-a2 splicing, we knocked it down by RNA interference and observed increased levels of unspliced lncRNAa-2 variants (Figure 30). Altogether, this suggests that the ncRNA-a2 contains splicing enhancer(s) in the middle of the intron that is regulated by hnRNP H.

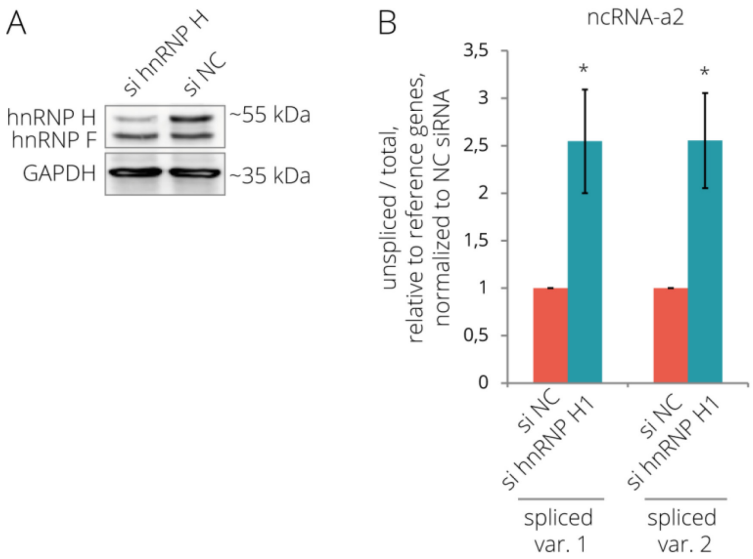

Figure 30. HnRNP H protein plays a role in the splicing of ncRNA-a2. (**A**) Western blot confirmation of hnRNP H protein down-regulation. (**B**) The fraction of unspliced transcripts of the ncRNA-a2 gene after hnRNP H siRNA knock-down. Barplots show relative RNA levels as determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test comparing the hnRNP H knock-down with the negative control, *p<0.05 (published in Krchňáková et al. 2018).

Next, we tested whether removal of G-runs in the middle of the intron can be rescued by insertion of intronic splicing enhancers that support recognition of 5'ss. We introduced one, two and three copies of a known intronic splicing enhancer (ISE) motif containing G-runs (Wang et al. 2012) downstream of the ncRNA-a2 5'ss (**Figure 31**). As a control, we introduced a mutated ISE element (ISEctrl). The splicing efficiency significantly increased (1.5-2.5x) in all cases except in the control. Insertion of the ISE element into the ncRNA-a2 lacking the middle G-rich sequence (3xISE-Δ60) rescued splicing to WT level, which is consistent with a model wherein the middle intron G-run sequences promote recognition of the 5'ss.

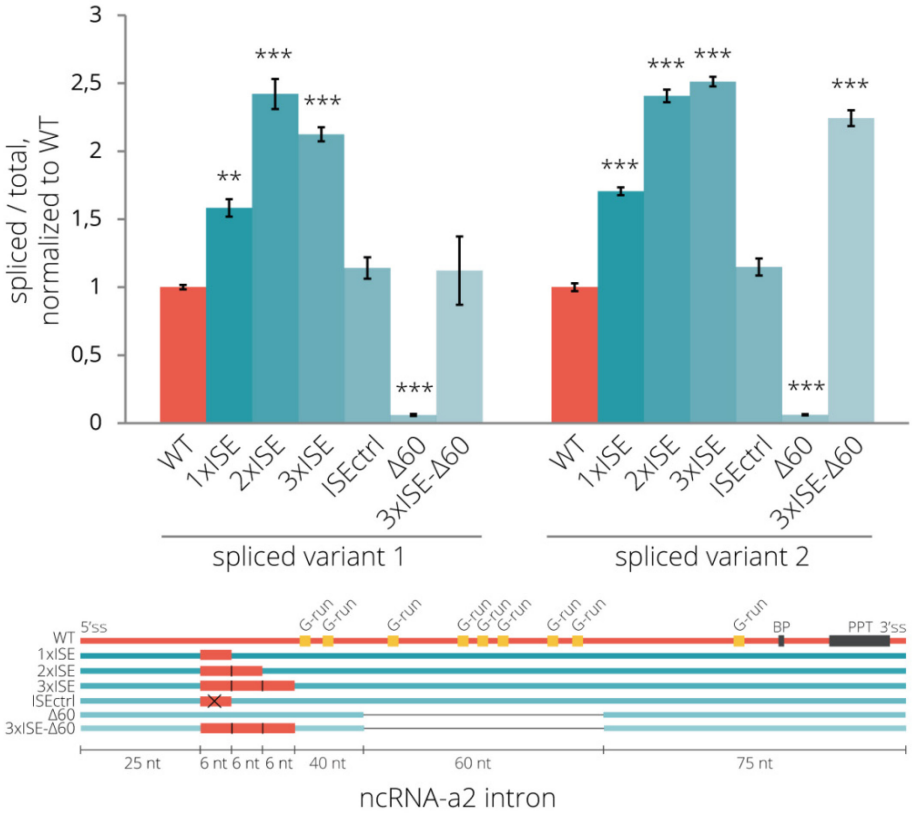

Figure 31. Introduction of intronic splicing enhancer motifs can elevate the slicing efficiency of ncRNA-a2. Splicing efficiencies are measured as a fraction of spliced transcripts relative to the total amount of transcripts. Schemes under the charts represent the intron sequence of ncRNA-a2 gene with predicted a branch point (BP), the PPT, G-run motifs (yellow) and inserted ISE motif(s) (red). Bar-plots show RNA levels as determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test comparing the individual mutant with WT, *p<0.05, *p<0.01, ***p<0.001 (published in Krchňáková et al. 2018).

5’ ss and Polypyrimidine Tract are Important for LincRNA Splicing

Previous results indicate that enhancer sequences that promote 5’ss recognition enhance ncRNA-a2 splicing. To test the role of the 5’ss in ncRNA-a2 splicing more rigorously, we prepared several mutants with increased strengths of the 5’ss. We utilized the WT ncRNA-a2 and the Δ1 mutant lacking the putative inhibitory sequences. Mutations increased the 5’ss MES (MaxEnt score) to 7.53 (WT) and 7.66 (Δ1 mutant), which is approximately one point below the average 5’ss strength of lincRNAs (8.56), 9.60 (WT) and 9.35 (Δ1 mutant), which is similar to the threshold of top 25% 5’ss (9.79 for lincRNAs and 9.80 for PCGs) and to 11.08, which falls into the top 10% of the strongest 5’ss (**Figure 32**, for MES distribution see Supplementary Material). While the average 5’ss strength did not improve splicing efficiency, substitutions leading to strong 5’ss significantly enhanced ncRNA-a2 splicing efficiency (>3.5x), primarily the spliced variant 1. The effect was stronger for WT ncRNA-a2 compared to the Δ1 mutant.

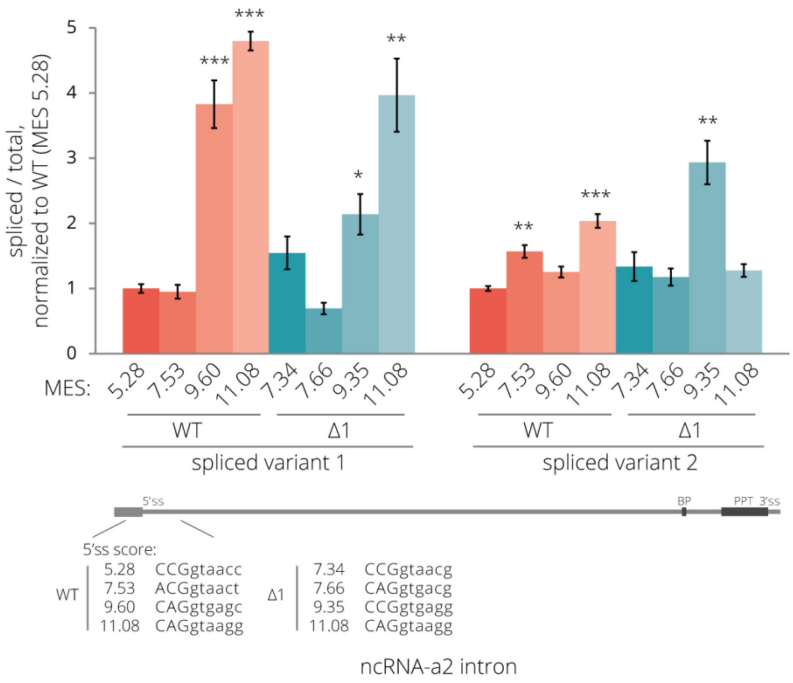

Figure 32. 5’ss strength plays an important role in splicing of ncRNA-a2. Mutation strengthening of the 5’ss improved splicing efficiency of WT as well as Δ1 mutant lacking a putative splicing inhibitory sequence. Splicing efficiencies were measured as a fraction of spliced transcripts relative to the total amount of transcripts. A scheme under the charts indicate modifications of the intron sequence of ncRNA-a2 gene. Predicted branch point (BP), the PPT (black) are indicated. Bar plots show relative RNA levels as determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student’s T-test comparing the individual mutant with either WT (MES 5.28) or Δ1 (MES 7.34), *p<0.05, **p<0.01, ***p<0.001 (published in Krchňáková et al. 2018).

Therefore, we wanted to investigate whether 5'ss strength is a general determinant of lncRNA splicing. We analyzed data used for calculating splicing indices to evaluate 5'ss strength in differently spliced genes. We categorized lincRNA and PCG introns into four groups based on their splicing efficiencies (increasing splicing index), and we calculated the mean 5'ss MES for each group (Krchňáková et al. 2018). We found a positive correlation between 5'ss strength and splicing efficiency of lincRNAs in four tested cell lines (Pearson's correlation coefficients 0.67-0.94) while no such correlation was found for PCGs (Pearson's correlation coefficients -0.75-0.57). These results suggest that lincRNA splicing is more dependent on 5'ss strength than splicing of PCGs.

To further identify sequences that could be responsible for less efficient splicing of lincRNAs, we compared the consensus sequences at 5' and 3'ss of lincRNAs and PCGs, which represent a group of effectively spliced genes (Krchňáková et al. 2018). Similarly to previous studies (Derrien et al. 2012; Melé et al. 2017), we did not detect any differences in 5'ss composition (nucleotides -3 to +6 with respect to the 5'ss). However, 3'ss sequences of lincRNAs (nucleotides -20 to +3 including PPT and the YAG motif) showed lower homology to the consensus 3'ss sequence than PCGs. A detailed analysis of PPT sequences (nucleotides -40 to -1) revealed slightly better conservation of C/T nucleotides at position -3 of the YAG sequence in PCGs in all five tested cell lines (Krchňáková et al. 2018). Interestingly, we found that the stretch of thymidines (Ts) within the PPT of lincRNAs is longer than in PCGs. In line with this finding, a higher number of Ts in lincRNA genes versus PCGs was observed in a recent study analyzing lincRNA splice-site strengths (Melé et al. 2017).

To better understand the role of PPT length and T content, we utilized the model ncRNA-a2 and increased the number of Ts in its PPT. We either replaced all cytidines (CtoT), all purines (GAtoT), or all nucleotides (T21) with Ts, or deleted a stretch of four Ts upstream of the CAG 3'ss (ΔPPT; **Figure 33**). All PPT modifications that increased the T content had a positive effect on the splicing efficiency (4.6-5.3x increase compared to WT), the deletion of Ts inhibited splicing of splicing variant 1 (8.3x reduction with respect to WT) but not splicing variant 2 (1.7x increase with respect to WT). The strong PPT was able to compensate for splicing reduction induced by deletion of the G-run enhancer since the Δ60-T21 construct was spliced 6.4x better than WT ncRNA-a2.

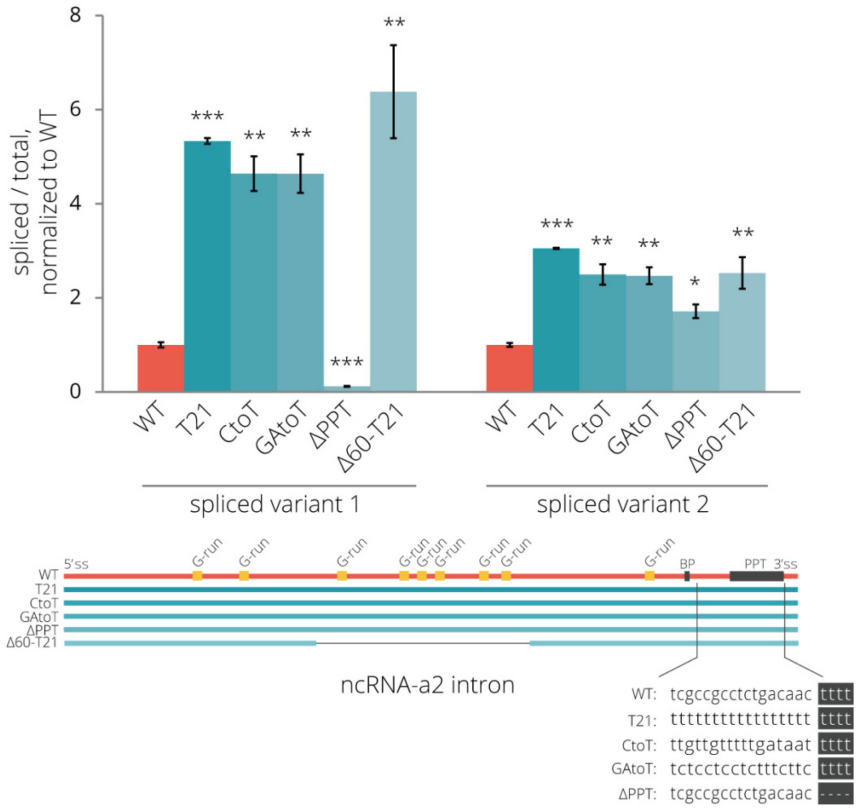

Figure 33. The higher number of Ts in PPT can promote lincRNA splicing efficiency. Splicing efficiencies of ncRNA-a2 after the substitution of Cs (CtoT), Gs (GtoT), all nucleotides (T21) for Ts and deletion of four Ts (ΔPPT). The scheme under the chart represents the intron sequence of the ncRNA-a2 gene with predicted a branch point (BP), the PPT sequences of individual mutants, and G-run motifs (yellow). Splicing efficiencies are measured as a fraction of spliced transcripts relative to the total amount of transcripts. Bar-plots show RNA levels as determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test comparing the individual mutant with WT, *p<0.05, *p<0.01, ***p<0.001 (published in Krchňáková et al. 2018).

To test the contribution of individual intronic elements to ncRNA-a2 splicing, we prepared combinations of our deletion mutants and measured their splicing efficiencies (**Figure 34**). In all cases, increasing the number of Ts within the PPT significantly enhanced splicing by 5.3x (3xISE-Δ60-T21), 6.4x (Δ60-T21), and 9.3x (3xISE-T21), underlining the importance of the PPT sequence for ncRNA-a2 splicing. Additional insertion of ISE sequences downstream of the 5'ss (3xISE-T21) further promoted splicing by up to 9.3x (spliced variant 1) and 2.5x (spliced variant 2). Interestingly, the artificial ISE inserted downstream of the 5'ss rescued inefficient splicing of the mutant lacking the

endogenous enhancer sequence in the middle of the intron (3xISE-Δ60). These data suggest that the endogenous enhancer promotes recognition of the 5'ss.

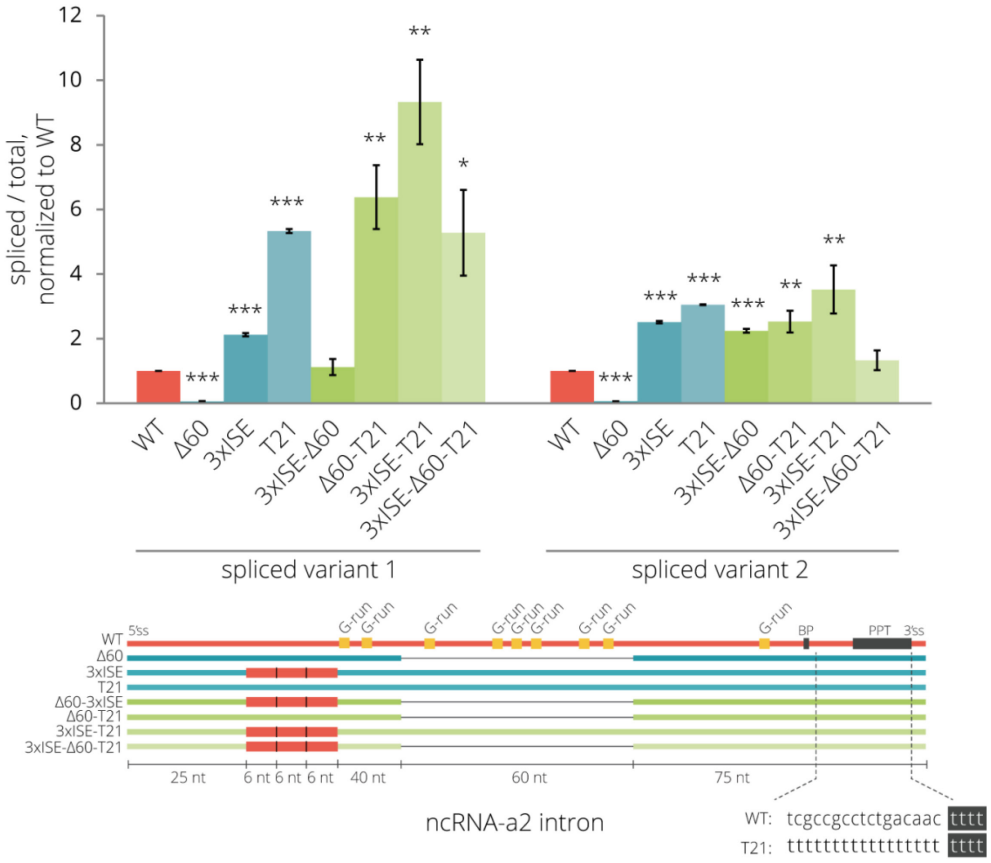

Figure 34. An artificial intronic splicing enhancer rescues the depletion of the endogenous G-run sequence. The scheme under the chart represents the intron sequence of the ncRNA-a2 gene with individual changes made. Splicing efficiencies are measured as a fraction of spliced transcripts relative to the total amount of transcripts. Bar-plots show RNA levels as determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test comparing the individual mutant with WT, *p<0.05, *p<0.01, ***p<0.001 (unpublished data).

Finally, we combined enhancement of 5'ss with the T21 mutation (**Figure 35**) and found that the improvement of 5'ss further stimulated splicing of ncRNA-a2 spliced variant 1, but reduced recognition of the downstream 3'ss and the production of spliced variant 2. The only exception was the strongest 5'ss with MES 11.08, which did not enhance splicing when compared with WT PPT containing ncRNA-a2. These results confirm that the 5'ss and the T content in the PPT have a strong and cumulative effect on ncRNA-a2 splicing.

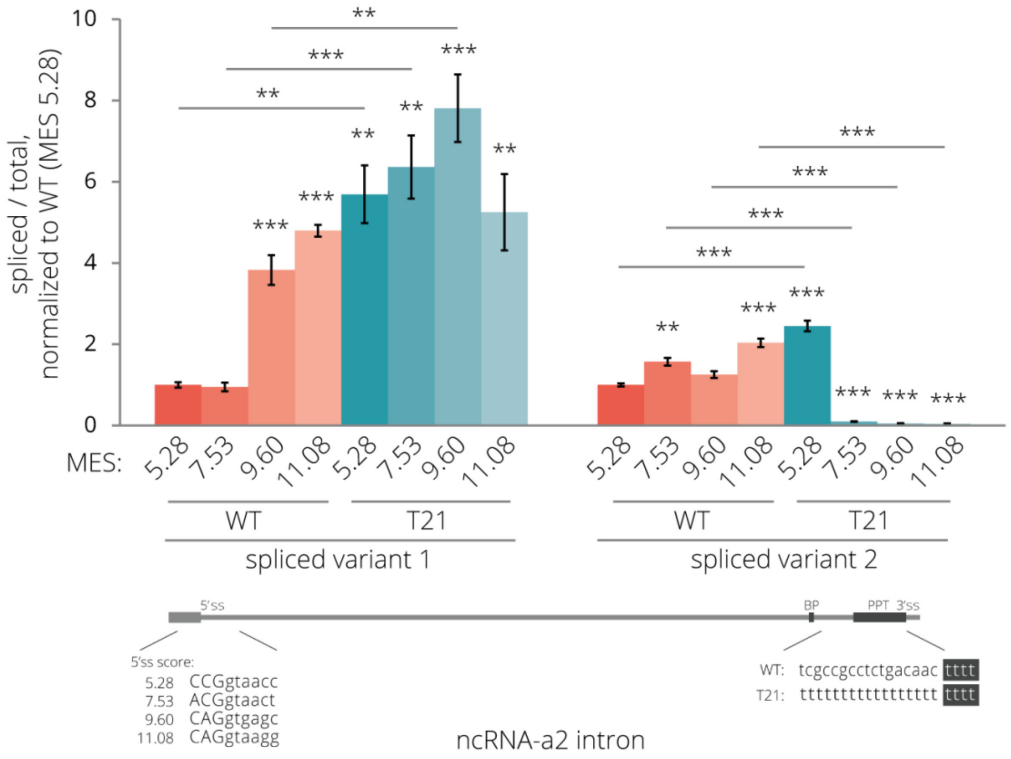

Figure 35. The PPT is a key determinant of lincRNA splicing efficiency. Strengthening 5'ss and PPT has a cumulative effect on ncRNA-a2 splicing. Mutations that improve MES of 5'ss (see **Figure 32**) were introduced into the T21 mutant, and splicing efficiency was analyzed and compared with WT (the data for WT are identical as in **Figure 32**). Asterisks above bars indicate the statistical significance of the individual mutant with respect to WT ncRNA-a2 and asterisks above lines compares WT and T21 constructs with identical 5'ss. Splicing efficiencies are measured as a fraction of spliced transcripts relative to the total amount of transcripts. Bar plots show relative RNA levels as determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test comparing the individual mutant with either WT (MES 5.28) or T21 (MES 5.28), *p<0.05, **p<0.01, ***p<0.001 (published in Krchňáková et al. 2018).

To further test the importance of the PPT for lincRNA splicing, we selected five lincRNAs with low splicing indices and mutated their PPTs. First, we compared splicing efficiencies of endogenous lincRNAs with lincRNAs transiently expressed from CMV-driven plasmid vectors and showed that splicing efficiency is not affected by ectopic expression (**Figure 36A**). Then, we converted all nucleotides between the putative branch point and the YAG motif into Ts, which significantly enhanced the splicing efficiencies of four out of five tested lincRNAs (**Figure 36B**). These results are consistent with the model that a higher number of Ts in the PPT correlates with more efficient lincRNA splicing.

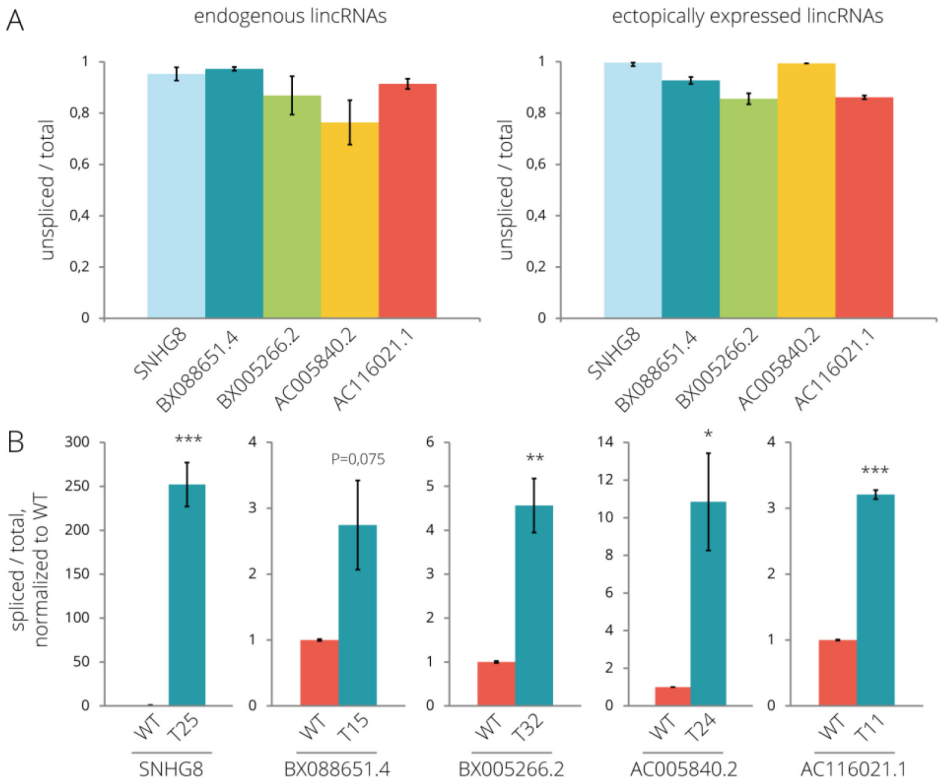

Figure 36. Increasing the number of Ts in PPT enhances splicing efficiencies of inefficiently spliced lincRNAs. (**A**) Inefficient splicing of selected lincRNAs. Splicing efficiencies of endogenous lincRNAs (left) and splicing efficiencies of transiently expressed lincRNAs (right). Splicing efficiencies of transiently expressed lincRNAs increases after the substitution of nucleotides in their PPT by Ts. (**B**) The PPT of *HBB* enhances splicing of ncRNA-a2. (**A-B**) Splicing efficiencies are measured as a fraction of spliced transcripts relative to the total amount of transcripts. Bar-plots show RNA levels as determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test comparing the individual mutant with WT, *p<0.05, *p<0.01, ***p<0.001 (published in Krchňáková et al. 2018).

Finally, to test whether a PPT sequence from PCG that is optimized for splicing can enhance splicing of lincRNA, we replaced the ncRNA-a2 PPT with the PPT sequence from HBB that had higher T/G and T/C ratios than the ncRNA-a2 PPT (**Figure 37**). The insertion of the HBB PPT into ncRNA-a2 significantly increased its splicing efficiency, confirming that the ncRNA-a2 PPT is weaker than the HBB PPT. Replacing natural 5'ss ncRNA-a2 sequence with stronger 5'ss from HBB (3 nt upstream and 6 nt downstream of 5'ss) showed partial, but not statistically significant enhancement of spliced variant 1 splicing, which suggests that the PPT sequence is more important than the 5'ss for ncRNA-a2 splicing.

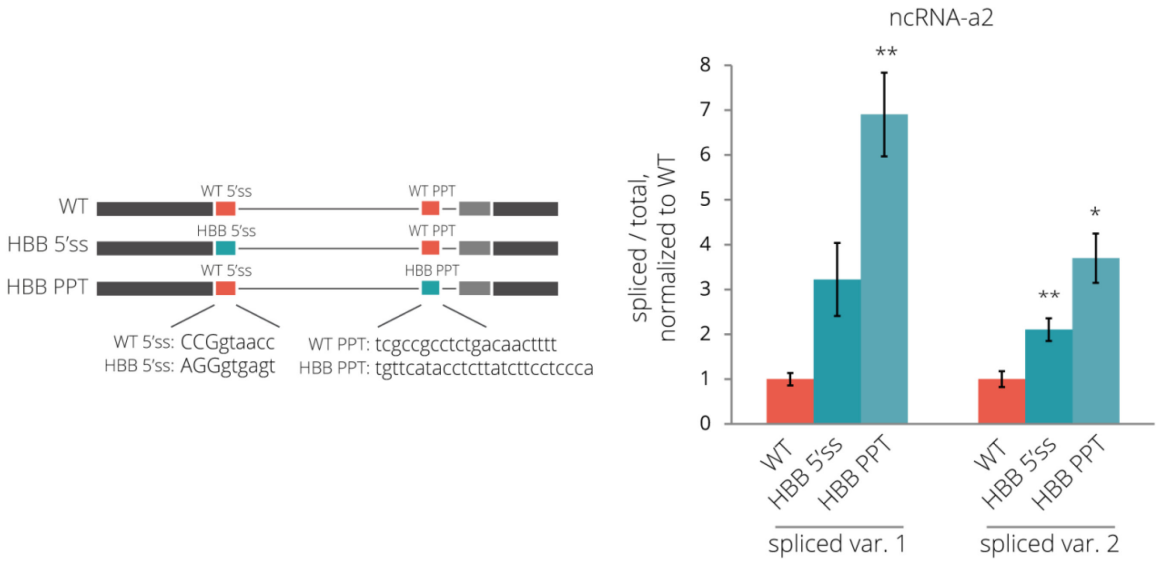

Figure 37. The PPT of *HBB* enhances the splicing of ncRNA-a2. Splicing efficiencies are measured as a fraction of spliced transcripts relative to the total amount of transcripts. Bar plots show relative RNA levels as determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test comparing the individual mutant with WT, *p<0.05, **p<0.01, ***p<0.001 (published in Krchňáková et al. 2018).

Our analyses showed the importance of the PPT sequence for lincRNA splicing. PPT serves as a binding platform for several key splicing proteins. Therefore, we analyzed how splicing factors, which preferentially bind U-rich sequences in the PPT, interact with lincRNAs and how their binding affect lincRNA splicing. We focused on U2AF2 (U2AF65), hnRNP C and PTBP1 (polypyrimidine tract binding protein 1), which were all shown to bind to the U-rich sequences in the PPT (Mulligan et al. 1992; Wagner and Garcia-Blanco 2001; König et al. 2010; Zarnack et al. 2013). We utilized publicly available eCLIP (enhanced crosslinking and immunoprecipitation) data from HepG2 cells (Van Nostrand et al. 2017) and iCLIP (individual nucleotide–resolution crosslinking and immunoprecipitation) data from Hela cells (Xue et al. 2009; Zarnack et al. 2013) and compared splicing efficiencies of lincRNAs associated/not associated with these proteins (Krchňáková et al. 2018). In the U2AF2 data set, lincRNAs bound by U2AF2 have significantly higher splicing efficiencies than lincRNAs that are not bound by U2AF2, in agreement with a recently published analysis (Melé et al. 2017). In contrast, PTBP1-bound and hnRNP C-bound lincRNAs were spliced as efficiently as unbound lincRNAs. These

data suggest that U2AF2 binding improves lincRNA splicing efficiency, while PTBP1 and hnRNP C binding do not.

To test this prediction experimentally, we transiently expressed ncRNA-a2 WT and its T21 mutant and analyzed their interactions with U2AF2 by RNA immunoprecipitations (RIP) followed by RT-qPCR. We observed that the T21 mutant more efficiently co-precipitated with the U2AF2 protein than the WT transcript (**Figure 38**). Altogether these results are consistent with the model that inefficient U2AF2 binding is one of the key factors that reduces splicing efficiencies of lincRNAs.

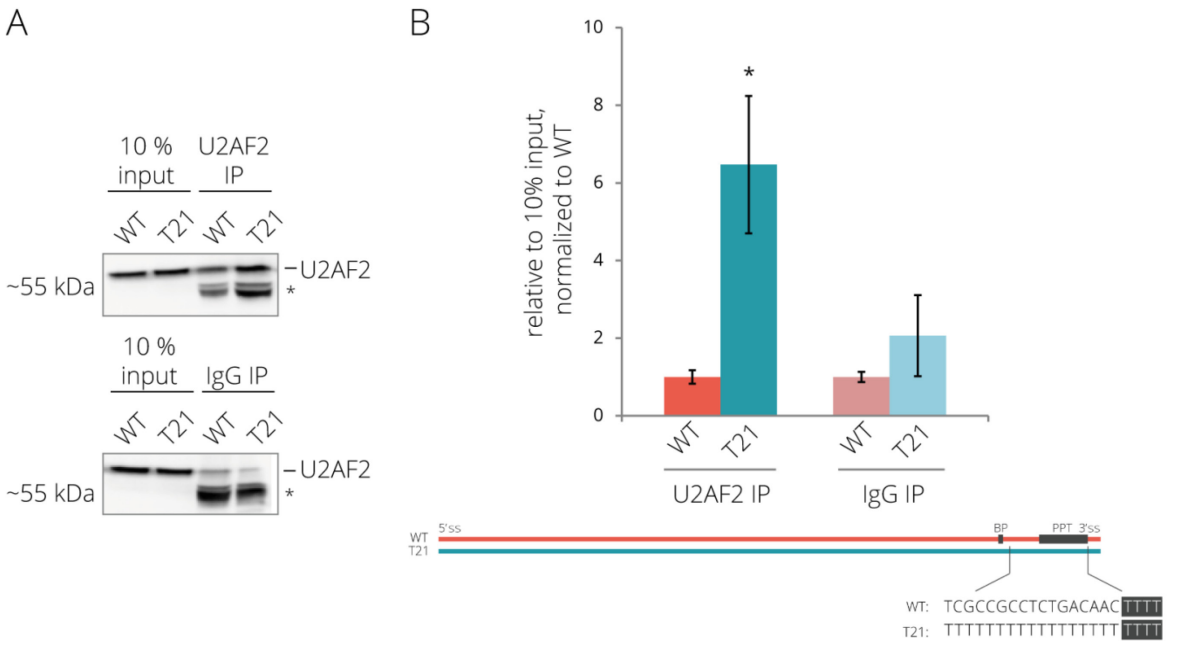

Figure 38. A higher number of Ts in PPT enhances the U2AF2 binding to ncRNA-a2. (**A**) Results of RNA immunoprecipitation using the anti-U2AF2 antibody. The position of U2AF2 is shown on WB, the asterisk marks unspecific proteins pulled down in both U2AF2 and control IgG IPs. (**B**) Results of RT-qPCR measurement of splicing efficiencies determined as a fraction of spliced transcripts relative to the total amount of transcripts. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test comparing the mutant with WT, *p<0.05, **p<0.01, ***p<0.001 (published in Krchňáková et al. 2018).

SR Proteins Bind Less Efficiently to LincRNAs

Our data suggest that a strong 5'ss and T-rich PPT together with productive U2AF2 binding are required for efficient lincRNA splicing. Such high dependence on strong splice sites could signal that lincRNAs lack additional splicing enhancer sequences that navigate the basic splicing machinery to splice sites. However, previous bioinformatic analyses showed that the global density of exonic splicing enhancers (ESEs) is even slightly higher in lincRNAs than in PCGs (Melé et al. 2017). In addition, ESEs are conserved in lincRNAs, and no difference in the number of ESEs has been observed between efficiently and inefficiently spliced lincRNAs (Schüler et al. 2014; Haerty and Ponting 2015; Melé et al. 2017).

To perform a more focused analysis, we searched for the occurrence of motifs that are known to be recognized by SR proteins, general splicing enhancers (Paz et al. 2010; Mueller and Hertel 2011; Müller-McNicoll et al. 2016). We determined the occurrence of 29 consensus motifs in exons (100 nt upstream of 5'ss or 100 nt downstream of 3'ss) and observed a striking difference in motif densities between lincRNAs and PCGs (**Figure 39**, **Figure 40**, **Figure 41**). Only one motif (SRSF3 - WCWWC) was significantly enriched in lincRNA exons while the majority of analyzed SR binding motifs were more prevalent in PCGs.

Figure 39. LincRNAs have less SRSF binding site densities than PCGs. Combined distribution of 29 SR protein binding motifs within 100 nt regions upstream of 5'ss and downstream of 3'ss between lincRNAs and PCGs. Motifs enriched in PCG exons are indicated in the right. Two SRSF9 motifs were not found in lincRNAs (published in Krchňáková et al. 2018).

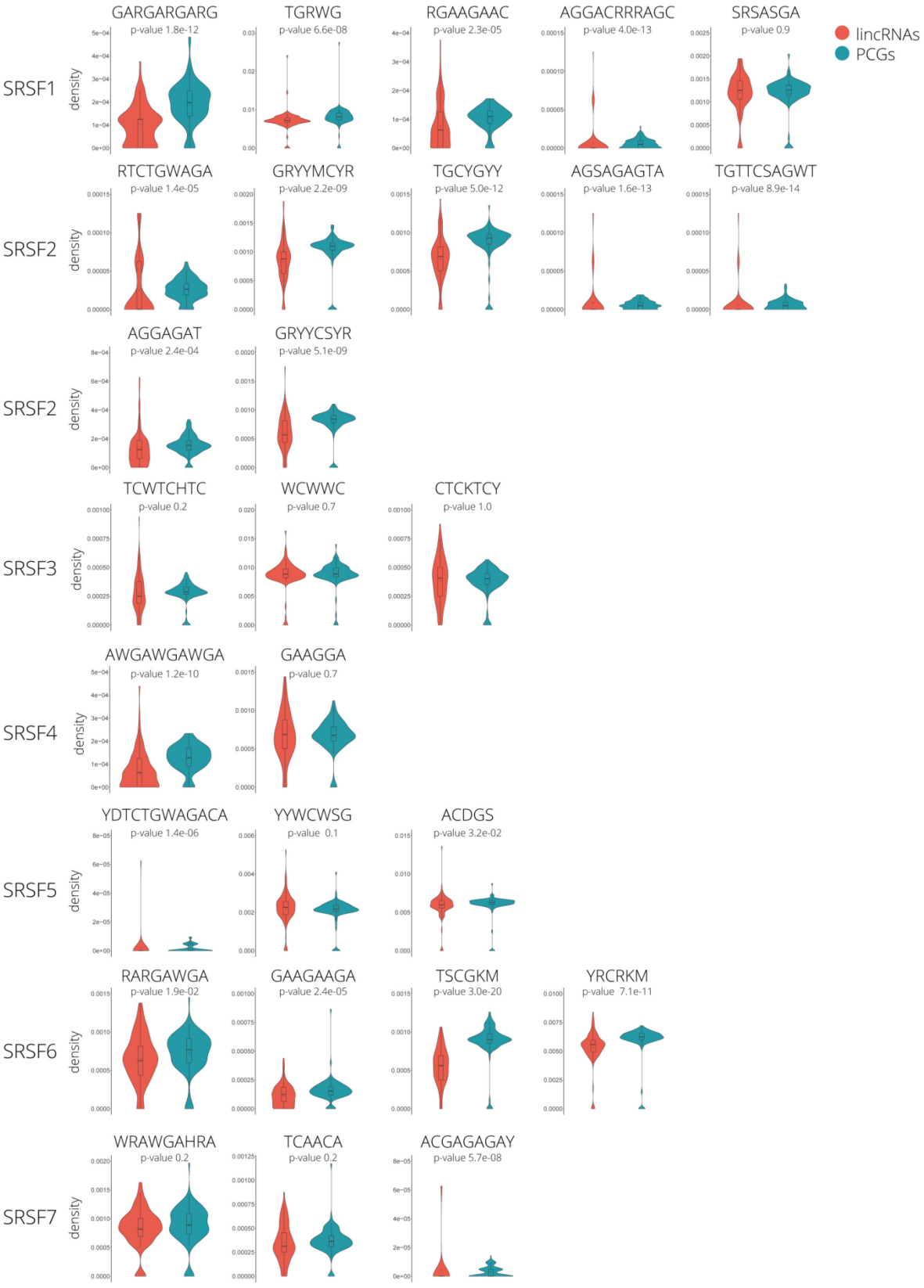

Figure 40. SRSF binding motif distributions in lincRNAs and PCGs 100 nt upstream of 5'ss. References for motifs are provided in Supplementary Material. P-values are calculated by Wilcoxon rank sum test. Two binding motifs of SRSF9 are not shown because no motifs were found in lincRNAs (published in Krchňáková et al. 2018).

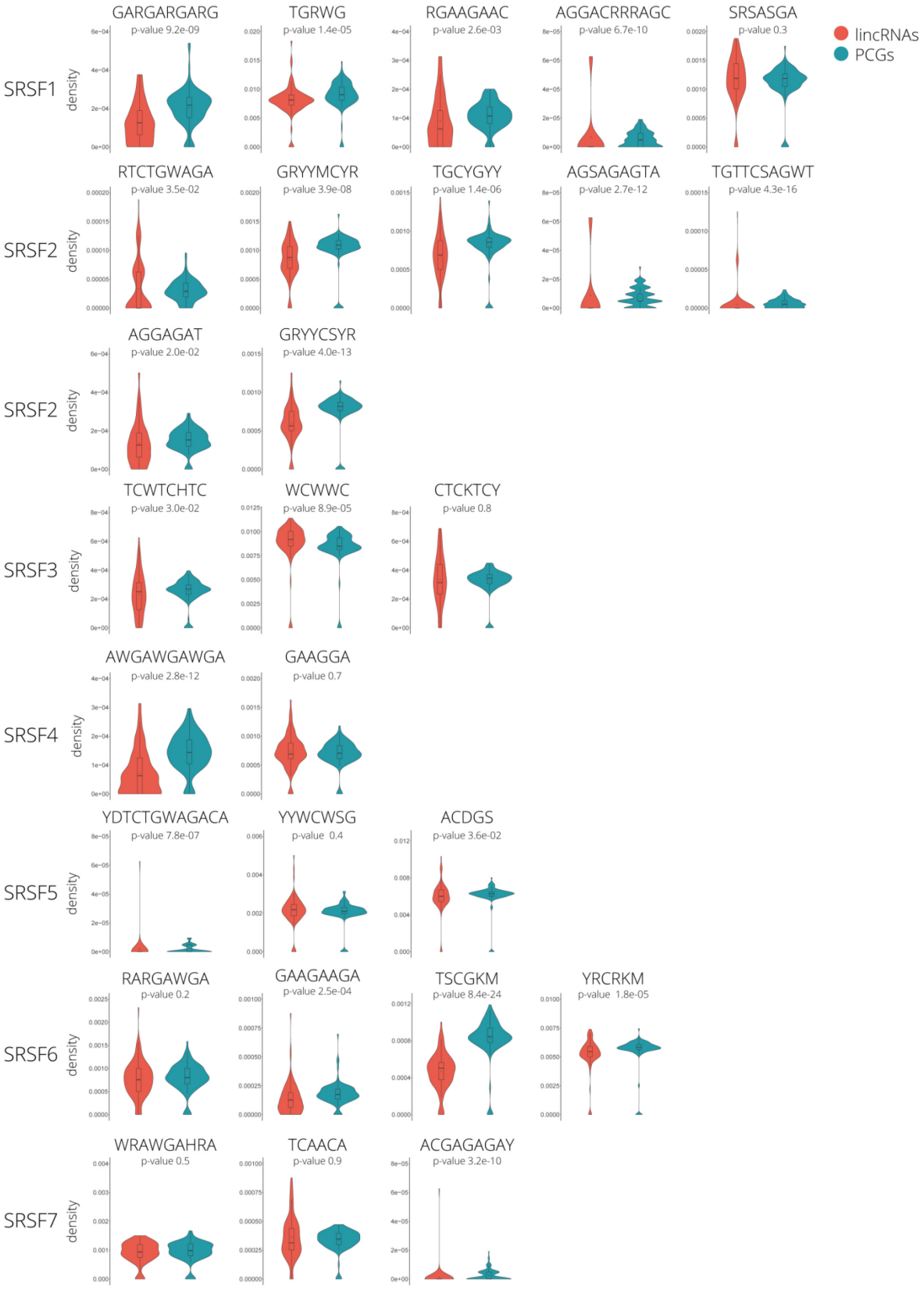

Figure 41. SRSF binding motif distributions in lincRNAs and PCGs 100 nt downstream of 3'ss. References for motifs are provided in Supplementary Material. P-values are calculated by Wilcoxon rank sum test. Two binding motifs of SRSF9 are not shown because no motifs were found in lincRNAs (published in Krchňáková et al. 2018).

To test whether a smaller number of SR binding motifs in lincRNAs results in a lower interaction with SR proteins, in collaboration with Nejc Haberman from MRC London Institute of Medical Sciences, we analyzed available eCLIP data performed with SRSF1, 7 and 9 in HepG2 cells (Van Nostrand et al. 2017). All three SR proteins bound efficiently within PCG exons, while their association with lincRNAs was much weaker and we did not detect any significant enrichment over exons (Krchňáková et al. 2018). SR protein binding to lincRNAs was lower compared to the total expressed PCGs (18-26% of binding to PCGs). To normalize for the expression level of PCGs and lincRNAs, we created a subset of PCGs that match number and expression level of lincRNAs and repeated the analysis. Similarly, the binding of SR proteins to lincRNAs was reduced to 20-30% of expression-matched PCGs (Krchňáková et al. 2018).

To investigate the binding of additional SR proteins not covered by eCLIP, we established a collaboration with Michaela Müller-McNicoll from Goethe University in Frankfurt and performed iCLIP in HeLa cell lines stably expressing GFP-tagged SRSF2, SRSF5 or SRSF6 from bacterial artificial chromosomes at near endogenous levels using anti-GFP antibodies as described before (Botti et al. 2017). Similarly to previous studies (Fairbrother et al. 2004; Xiao et al. 2007), we show that SR proteins bound preferentially to exonic regions of PCGs (Krchňáková et al. 2018). In agreement with the eCLIP data, binding of all three analyzed SR proteins to lincRNAs was much lower compared to all expressed PCGs (13-30% of binding to PCGs) or expression-matched PCGs (56-68% of binding to PCGs). Altogether, this confirmed that SR proteins interact poorly with lincRNAs, which is independent of lincRNA expression level. To our knowledge, this was for the first time that such an inefficient binding of SR proteins to lincRNAs compared to PCGs was observed.

Role of Intron in the Function of Long Non-Coding RNAs

The complex relationship of the importance of splicing-associated processes in the function of enhancer-like lncRNAs is still not resolved, so the controversy of opposing studies (Yin et al. 2015; Engreitz et al. 2016a; Tan et al. 2018) about splicing role in enhancer function is not definitely answered. Because of that, we wanted to know if splicing plays a role in the function of our model ncRNAs-a2.

NcRNA-a2 together with other RNAs including ncRNA-a5 has been suggested to act as transcription enhancers because their depletion by RNAi decreased the expression of some adjacent PCGs (**Figure 42**) (Ørom et al. 2010). Specifically, after ncRNA-a2 transcript depletion, the amount of KLHL12 transcripts was reduced. Similarly, the depletion of ncRNA-a5 resulted in ROCK2 down-regulation. Surprisingly, this decrease was specific for one, target PCG located in lncRNA vicinity but not the closest one which is *KDM5B* for ncRNA-a2 and *E2F6* for ncRNA-a5.

Such stimulatory effects on neighboring genes were shown previously but limited to an immediate neighboring gene (Engreitz et al. 2016a) suggesting that specific RNA transcript is not required for such expression regulation but instead involves general processes associated with their production, including the activity of gene promoters, the process of transcription, and maybe even the splicing of the transcript. However, the down-regulation of not the most adjacent genes upon RNAi depletion of ncRNA-a2 and ncRNA-a5 indicates the different modes of their activation.

Therefore, to better understand the activating function of ncRNA-a2, we tested whether ectopic expression of this RNA can induce expression of the target gene. Our result indicates that ncRNA-a2 acts in *cis* because overexpression of ncRNA-a2 from a CMV-driven plasmid did not increase expression of its target PCG *KLHL12*, so the ncRNA-a2 transcripts not transcribed from its endogenous locus cannot stimulate expression of the target gene (**Figure 43**).

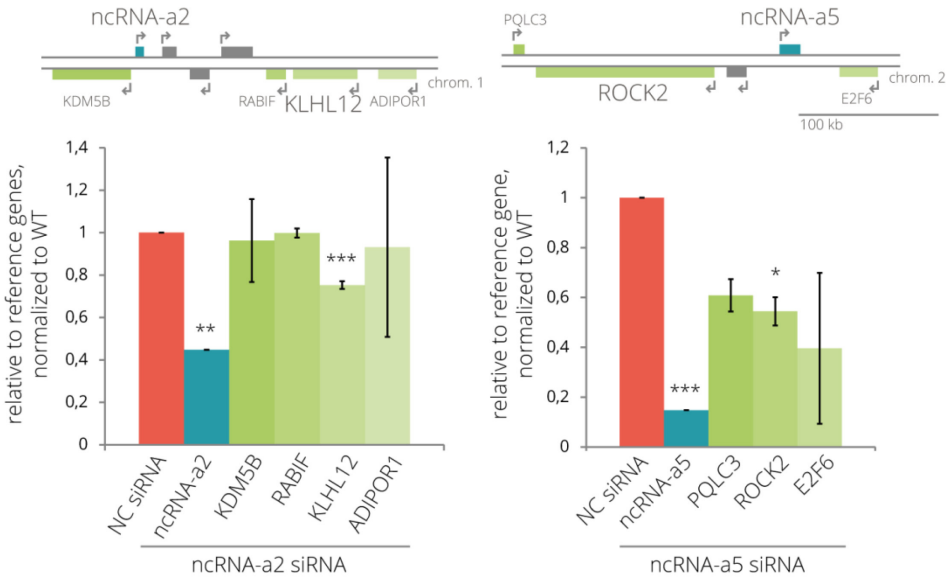

Figure 42. NcRNA-a2 and ncRNA-a5 can activate gene expression of a PCG located in their genome vicinity. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test comparing the individual transcript with negative control (NC) siRNA, *p<0.05, *p<0.01, ***p<0.001 (published in Krchňáková et al. 2018).

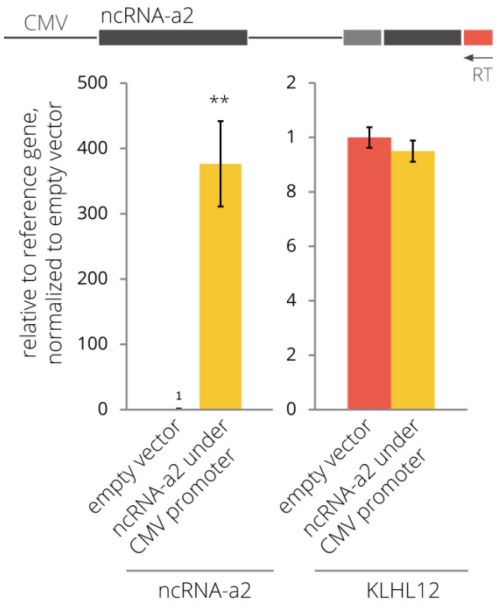

Figure 43. NcRNA-a2 act in *cis*. The expression of ncRNA-a2 and its target PCGs after the transient expression of the ncRNA-a2 gene. The arrow represents a reverse primer used for RT that is specific for transiently expressed ncRNA-a2. Bar plots show RNA levels as determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test, *p<0.05, *p<0.01, ***p<0.001 (published in Krchňáková et al. 2018).

Since ncRNA-a2 seems to act in *cis*, we decided to either stimulate or repress the expression of ncRNA-a2 from its endogenous locus using CRISPRa (CRISPR activation) and CRISPRi (CRISPR interference) mechanism. Both methods use a catalytically deactivated version of widely used *Streptococcus pyogenes* Cas9 protein with the mutations in its nuclease domains (dCas9). In the case of CRISPRa, activation domain of VP64 is used. This is a tetrameric repeat of the minimal activation domain of VP16 (Herpes Simplex Viral Protein 16) which is a strong transcriptional activator of viral immediate early promoters. To further enhance the activation, in this work we used an additional amplification step. We utilized SunTag system which includes an array of small peptide epitopes fused to the C terminus of dCas9 to recruit multiple copies of scFV (single-chain variable fragment - fusion protein of the variable regions of the heavy and light chains of immunoglobulins) fused to VP64 (Tanenbaum et al. 2014). We used this system to localize activation domain immediately upstream of *ncRNA-a2* gene. Surprisingly, an increased amount of ncRNA-a2 transcripts in its locus did not promote the expression of its target gene (**Figure 44**).

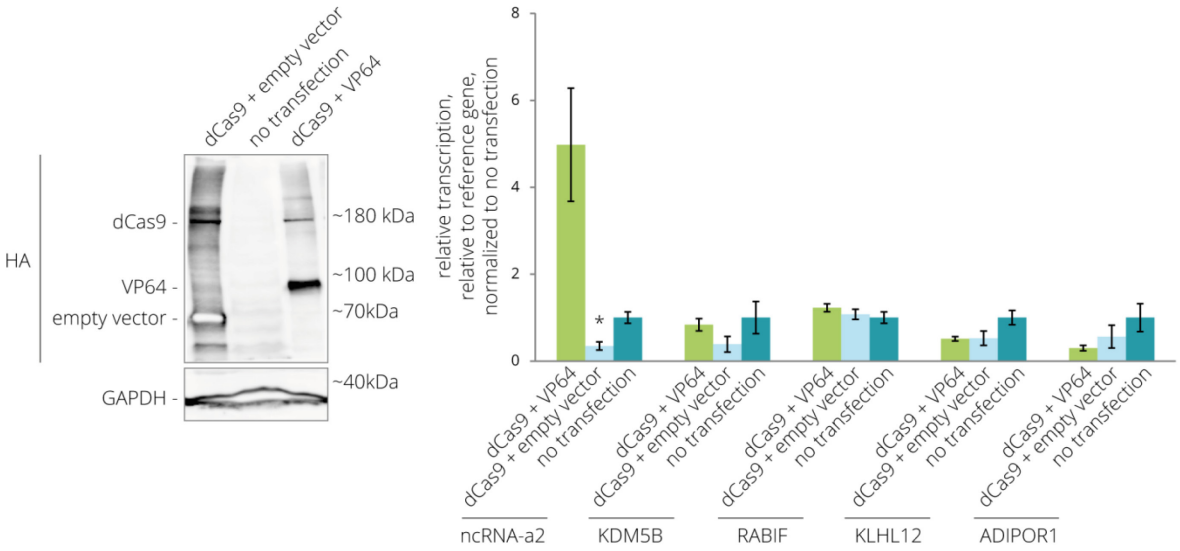

Figure 44. The endogenous activation of ncRNA-a2 does not increase the expression of its target *KLHL12* gene. The expression of ncRNA-a2 and PCGs located in its genomic vicinity was assayed by RT-qPCR. The mean of three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test compared with no transfection, *p<0.05, *p<0.01, ***p<0.001 (unpublished data).

Contrary to the proposed activating function of *ncRNA-a2*, we observed a partial, albeit non-statistically significant, transcription decrease of target *KLHL12* and

neighbouring *ADIPOR1* gene. Interestingly, the difference in transcription between empty control (dCas9 + empty vector) and no transfected control suggest the repression effect of the dCas9 complex binding to the ncRNA-a2 promoter, likely via sterical hindrance with RNA polymerase II or transcription initiation factors.

To repress the endogenous transcription of *ncRNA-a2* gene, we used CRISPRi system which utilizes dCas9 protein fused with a transcriptional repressor KRAB (Krüppel-associated box) which recruits the KAP1/TIF1β corepressor complexes. KAP1 acts as a scaffold to further recruit factors associated with DNA methylation and formation of repressive chromatin, such as heterochromatin protein 1 (HP1), histone deacetylases and histone methyltransferases. Analogously to CRISPRa, we localized dCas9 fused with repressor immediately upstream of *ncRNA-a2* gene and observed a decrease in the amount of ncRNA-a2 transcripts (**Figure 45**). This depletion also resulted in the decrease of expression of its target gene (*KLHL12*) together with *KDM5B* which is the most adjacent to *ncRNA-a2*, and *ADIPOR1* located farther away next to *KLHL12*.

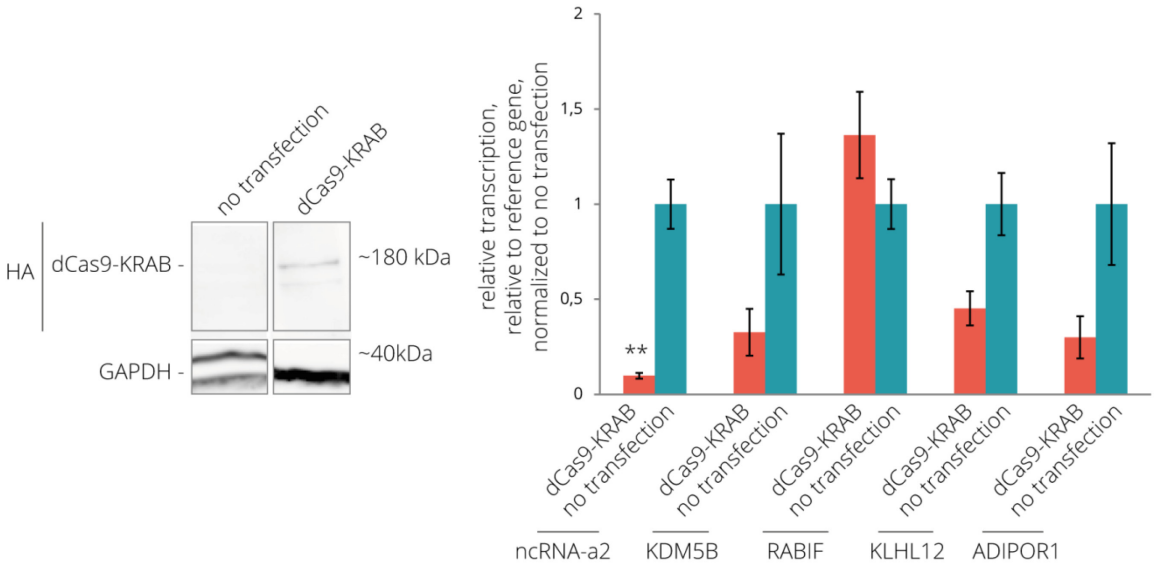

Figure 45. The endogenous depletion of ncRNA-a2 leads to decreased expression of its target *KLHL12* gene. The expression of ncRNA-a2 and PCGs located in its genomic vicinity was assayed by RT-qPCR. Bar plots show RNA levels as determined by RT-qPCR. The mean of three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test compared with no transfection, *p<0.05, *p<0.01, ***p<0.001 (unpublished data).

Taken together, experiments from overexpression of *ncRNA-a2* (**Figure 43**) and CRISPRa/i (**Figure 44** and **Figure 45**) suggest that *ncRNA-a2* expression is not essential

for promoting neighboring genes expression. On the contrary, elevated expression of *ncRNA-a2* by CRISPRa lead to a decrease of two neighboring genes (*KLHL12*, *ADIPOR1*; **Figure 44**). Similar transcription decrease of these two genes was observed after expression of CRISPRi and even the empty dCas9. The most plausible explanation is that the DNA sequence bound by dCas9 (7-30 nt upstream of *ncRNA-a2* TSS) is important for the expression of these PCGs. In addition to these two PCGs, *KDM5B* was down-regulated upon expression of inhibitory dCas9 (**Figure 45**). The down-regulation of this gene can be explained by the closed chromatin at *ncRNA-a2* locus and its vicinity caused by the presence of dCas9 fused with the repression domain.

In order to distinguish if the DNA sequence or RNA product itself has a role in the activating function of *ncRNA-a2*, two poly(A) sites were inserted 40 nt downstream of endogenous *ncRNA-a2* TSS using CRISPR/Cas9 by providing a template for homologous recombination. After single-cell sorting, we isolated a cell line with poly(A) sites insertion which we confirmed by two-step genotyping PCR (**Figure 46A**). However, the transcription of ncRNA-a2 was not inhibited, and thus this experiment did not allow us to answer the question of the importance of ncRNA-a2 RNA for the activating function (**Figure 46B**). The reason for inefficient inhibition of *ncRNA-a2* transcription could be due to a low number of inserted poly(A) sites and/or the poly(A) sites are too close to TSS and do not induce ncRNA-a2 degradation.

Finally, we tested whether the intron of ncRNA-a2 contributes to its enhancer function. We removed the ncRNA-a2 intron from the endogenous *ncRNA-a2* gene locus using CRISPR/Cas9 by providing a template for homologous recombination. After single-cell sorting, we isolated three different intron-deleted cell lines which we confirmed by extensive genotyping and sequencing (intron deletion knock-outs preparation and genotyping was done by Michaela Krausová, Laboratory of RNA Biology, Institute of Molecular Genetics, Czech Academy of Sciences). For the simplicity and similar results from all three cell lines, in further experiments, we show only results for clone #1. In the beginning, we looked at the general expression of ncRNA-a2 together with the adjacent PCGs and observed a decrease in ncRNA-a2 transcripts to approximately in half (**Figure 47**, bottom left). This can be explained by the absence of the splicing process which can lead to a drop in transcription since splicing has been shown to be able to stimulate RNA polymerase II initiation and elongation (Fong and Zhou 2001; Furger et al. 2002; Kwek et

al. 2002; Kotovic et al. 2003; Lin et al. 2008). Interestingly, no significant decrease was observed for any neighboring PCGs.

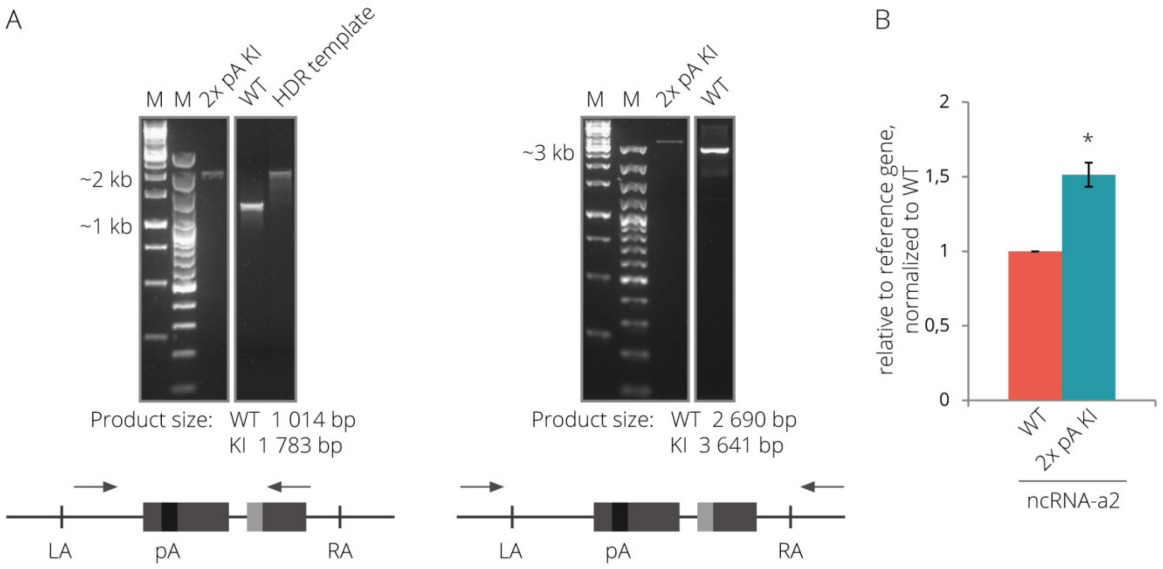

Figure 46. Inhibition of ncRNA-a2 by insertion of poly(A) sites into the ncRNA-a2 gene. (**A**) The first and second round of selection PCRs from genomic DNA using a set of primers (bottom schemes). The second round of selection PCR used primers that one of them recognizes a sequence outside the HDR template. By this we confirmed the insertion is within endogenous ncRNA-a2 locus; M – marker, WT – wild-type, HDR – homologous-directed repair template. (**B**) The expression of the *ncRNA-a2* gene was analyzed in 2x pA and WT cell line using primers recognizing region downstream of pA insertion. Bar plots show RNA levels as determined by RT-qPCR. The mean of three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test comparing the individual mutant with WT, *p<0.05, *p<0.01, ***p<0.001 (unpublished data).

However, since ncRNA-a2 seems to act in *cis* and close spatial proximity to its endogenous locus (**Figure 43**), we wanted to test the importance of ncRNA-a2 splicing on newly transcribed mRNAs. Because of that we either isolated RNAs associated with the chromatin fraction or metabolically labelled and isolated nascent RNAs using 4sU-biotin labeling (**Figure 47**). Again, as in the case of total RNA, we did not detect any significant difference in the transcription of neighboring genes after the deletion of the *ncRNA-a2* intron. These results suggest that the intron itself and/or its splicing do not play a significant role in the activating function of ncRNA-a2.

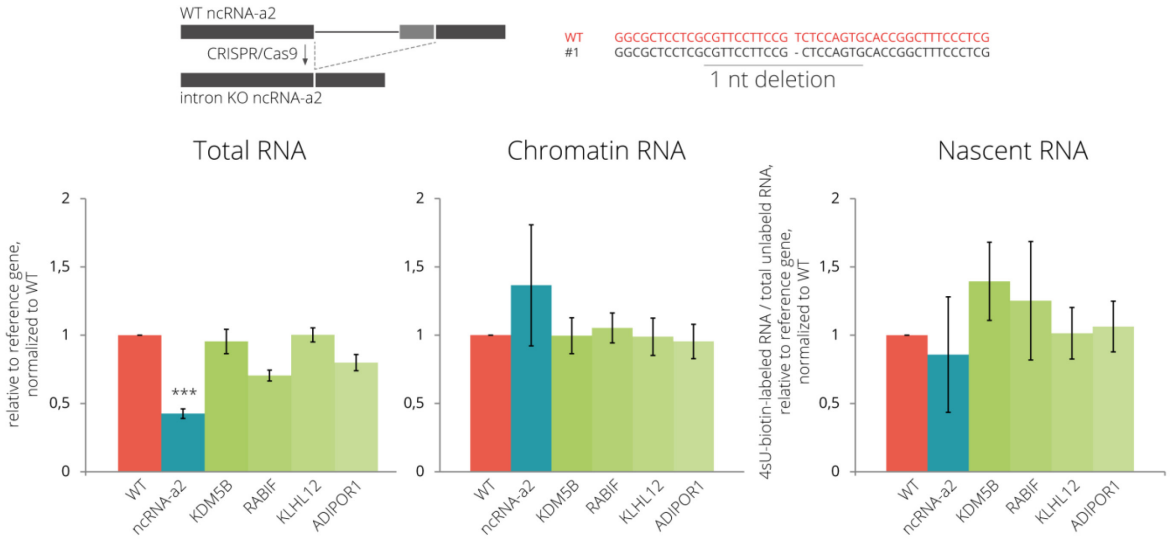

Figure 47. The ncRNA-a2 intron is not essential for the neighboring gene activation. Top schemes show CRISPR/Cas9-mediated intron removal (left) and details of the sequence at the exon/exon boundary after the intron deletion in #1 cell line (right). The expression of the ncRNA-a2 gene and PCGs located in its genomic vicinity was analyzed in #1 cell line. Bar plots show RNA levels as determined by RT-qPCR. The mean of at least three independent experiments is shown. Error bars indicate SEM; asterisks indicate the statistical significance levels calculated by two-tailed Student's T-test comparing the individual mutant with WT, *p<0.05, *p<0.01, ***p<0.001 (published in Krchňáková et al. 2018).

Discussion

Regulation of Splicing via Chromatin

It was previously suggested that chromatin modifications modulate alternative splicing (Nogués et al. 2002; Sims et al. 2007; Schor et al. 2009; Luco et al. 2010; Hnilicová et al. 2011; Saint-Andre et al. 2011; Dušková et al. 2014) and also that RNA splicing reciprocally feeds back on histone modifications (Almeida et al. 2011; Kim et al. 2011; Bieberstein et al. 2012). However, most of these studies have relied on global depletion/inhibition or overexpression of chromatin modifying enzymes and even though, the experiments are thoroughly controlled, one can never exclude the possibility that the observed changes in alternative splicing are caused by downstream secondary effects. Because of that, we have decided to change histone modifications locally using TALEs as a tool to navigate histone modification enzymes to a specific and unique sequence within the targeted gene. Given the specificity of TALE binding, secondary effects due to global changes in the transcriptional output can be eliminated, thus providing a tool to directly test the influence of local histone modifications on splicing.

In our study (Bieberstein et al. 2016), we have shown the modulation of histone marks directly on specific loci can lead to altered splicing. By changing H3K9 methylation, we observed altered inclusion of EDB exon of the *FN1* gene. Additionally, we have shown using publicly available ChIP-Seq data that H3K9me3 was enriched around internal exons in all tested cells suggesting this histone mark plays a general role in exon recognition and is a functional element in splicing regulation. Because of that, we tested if H3K9 methylation can also affect the splicing efficiency of a constitutive exon of *FOSL1* gene surrounded by weak splice sites. And indeed, depleted H3K9me3 resulted in an accumulation of unspliced transcripts (**Figure 18**) supporting the hypothesis that H3K9me3 at nucleosomes within the gene body affects exon recognition.

Two mechanistic models can explain the effect of H3K9me3 on splicing. First, methylated H3K9 is bound by Heterochromatin protein 1 (HP1), which in turn can recruit the splicing regulatory protein SRSF1 (Salton et al. 2014). Alternatively, H3K9 methylation might affect alternative splicing through kinetic coupling by slowing down RNA polymerase II elongation (Allo et al. 2009; Schor et al. 2009; Saint-Andre et al. 2011). Nucleosomes positioned over exons were previously proposed to act as speed

bumps for RNA polymerase II elongation and H3K9 methylation on these exon-centered nucleosomes may increase the effect on RNA polymerase II speed (Kwak et al. 2013). Furthermore, splicing of the EDB exon was shown to be sensitive to RNA polymerase II elongation rate (Pagani et al. 2003; Hnilicová et al. 2011; Dušková et al. 2014), rendering kinetic coupling the most likely mechanism by which TALE-directed H3K9 methylation changes EDB inclusion rates. Together, these results demonstrate the local chromatin environment modulates splicing outcome (Carrillo Oesterreich et al. 2010; Hnilicová et al. 2011).

In addition to chromatin modifications, a connection between the promoter and alternative splicing regulation has been proposed (Cramer et al. 1997; Cramer et al. 1999; Auboeuf et al. 2002; Pagani et al. 2003). However, it seems that alternative splicing regulation is not dependent on promoter strength or the amount of RNA transcribed from individual promoters (Pagani et al. 2003). On the contrary, it was suggested that factors regulating alternative splicing somehow act through specific promoter occupancy because different promoter-associated transcriptional co-activators affect alternative splicing only when tethered to their specific promoter (Monsalve et al. 2000; Kadener et al. 2001; Nogués et al. 2002; Hnilicová et al. 2013). Moreover, promoters were shown to contain elements that can control alternative splicing independently of transcription regulation and that this regulation involves chromatin acetylation. It was proposed that p300 histone acetyltransferase binds to specific promoter sequences (CRE – cAMP response elements) and it was shown the deletion or mutation of such sequences in promoter influences the alternative splicing of EDB exon (Dušková et al. 2014). In addition, the alternative splicing can also be influenced by transcription enhancer elements likely via affecting the speed of RNA polymerase II (Kadener et al. 2002). However, all these experiments were performed using minigenes and *in vivo*, the alternative exons are usually located several thousand bases away from these transcription regulatory sequences (e.g. EDB exon is ~42 kilobases downstream of its promoter).

Therefore, we decided to study a potential influence of predicted endogenous transcription enhancer on alternative splicing of the *FN1* gene. Using data about histone modifications, binding of regulatory proteins and DNaseI hypersensitivity, we predicted a FN1 transcription enhancer located ~23 kilobases upstream of FN1 transcription start site (Volek 2018). After the deletion of the whole enhancer element by CRISPR/Cas9, we observed lower overall transcription of the *FN1* gene and higher EDB exon inclusion

compared to WT (**Figure 20**). The lower transcription of the *FN1* gene in mutants is consistent with the deleted region to be indeed a functional transcription enhancer element.

Since it is known from previous studies (Dušková et al. 2014; Bieberstein et al. 2016) that the EDB exon inclusion is sensitive to chromatin marks, we analyzed changes in histone methylation along the *FN1* gene upon deletion of its putative transcription enhancer. Firstly, H3K4me2 levels were studied. We observed a partial increase of H3K4me2 over the promoter region (**Figure 21A**). This histone mark extensively overlaps with transcription factor binding regions and together with the H3K27 signal can greatly reduce false positive predictions of the transcription factor binding regions (Wang et al. 2014). However, whether the binding of transcription factors results in activation or repression of transcription is dependent on the particular transcription factor that is bound (Arnone and Davidson 1997). Therefore, in respect to our results, we propose that partially higher H3K4me2 levels over FN1 promoter upon enhancer element deletion lead to the binding of transcription repressor factors to the promoter, and thus lower overall FN1 transcription. However, whether enhancer deletion also modulates the binding of transcription factors onto promoter as well as the identity of such factors is a subject of additional experiments. Moreover, the significance of the H3K4me2 decrease over the alternative EDB exon for alternative splicing also waits for further investigation since it was observed only in one cell clone (**Figure 21A**).

Additionally, the H3K9me3 mark was also assayed. Since the deletion of the enhancer element resulted in ~2-fold increase in H3K9me3 immediately downstream of deleted enhancer element and significant depletion of H3K9me3 over EDB exon (**Figure 21B**), we hypothesize that region around enhancer element could become heterochromatized and thus no longer function as transcription enhancer. In organisms ranging from yeast to humans, constitutive heterochromatin is marked by H3K9me2 and H3K9me3 (Rea et al. 2000; Nakayama et al. 2001; Martens et al. 2005) which are bound by the Heterochromatin protein 1 (HP1). This protein self-oligomerizes and recruits repressive histone modifiers, contributing to heterochromatin compaction and spread (Bannister et al. 2001; Lachner et al. 2001; Canzio et al. 2011). Moreover, H3K9me3 occludes the DNA from binding by transcription factors with diverse DNA-binding domains (Soufi et al. 2012). We propose a model in which heterochromatization of enhancer could abolish the potential intrachromosomal looping between enhancer element and EDB exon which in normal situation spatially restrict the accessibility of the EDB

exon by splicing factors. However, such direct looping of FN1 enhancer with its exons was not studied yet, thus its presence as well as its potential role in splicing regulation is speculative. More importantly, these results are contradictory to our results obtained by TALEs when higher H3K9me3 lead to higher EDB exon inclusion (Bieberstein et al. 2016). However, in parallel, we observed reduced histone H4 acetylation over the EDB exon after enhancer deletion (**Figure 22**). These results are consistent with Hnilicová et al. (2011) and we speculate that histone acetylation or some other, not specified modification plays a superior role in splicing regulation than histone methylation.

Finally, we analyzed at H3K36me3 which is a general mark of the active transcription. And even though the overall FN1 transcription was substantially reduced after enhancer deletion (**Figure 20**), we did not observe any extensive decrease of H3K36me3 along the *FN1* gene upon enhancer deletion. The only exception was EDB exon where we observed a significant drop in H3K36me3 levels (**Figure 21C**). Although H3K36 methylation is most commonly associated with the transcription of active euchromatin, it has also been implicated in diverse processes, including alternative splicing, dosage compensation and transcriptional repression, as well as DNA repair and recombination. Even though it appears to act as both an activating and inhibitory signal, the overall biological readout might depend on the context of additional surrounding marks and their corresponding reader proteins (reviewed in Wagner and Carpenter 2012). Furthermore, H3K36me3 was shown to be preferentially positions to exonic over intronic sequences which are GC-content and exon-length independent but transcription-associated (Kolasinska-Zwierz et al. 2009). Moreover, alternative cassette exons have significantly reduced H3K36me3 exon signals relative to their constitutive neighbors (again GC-content and exon-length independently) indicating that this exon marking is related to splicing and is conserved in worms, mice and humans (Kolasinska-Zwierz et al. 2009). It is deposited on gene bodies concomitantly with RNA polymerase II transcriptional elongation (Li et al. 2002; Li et al. 2003; Xiao et al. 2003; Kizer et al. 2005; Sun et al. 2005; Yuan et al. 2009). A proposed function of such exon marking is to provide splicing machinery with a mechanism how to define exons via recruitment of specific splicing factors leading to inclusion or exclusion of particular exons. An example of such splicing regulation is altered exon inclusion of *FGFR2* (fibroblast growth factor receptor 2) gene in which H3K36me3 mark is specifically enriched over alternative exon. This modification is recognized by MRG15 (MORF-related gene 15) reader which also interacts with PTBP1

(polypyrimidine tract binding protein 1) protein. PTBP1 then binds to intronic splicing silencer sites which results in the lower exon inclusion (Carstens et al. 1998; Luco et al. 2010). In the case of our enhancer deletion mutants, the lower H3K36me3 can result in the less binding of negative splicing factor PTBP1 and thus leading to higher EDB exon inclusion (**Figure 21C**). Interestingly, H3K36me3 crosstalk was shown to work also in an opposite way since altered splicing can affect H3K36me3 levels (Kim et al. 2011). In this study, splicing abolition caused a shift in the relative distribution of H3K36me3 away from 5′ ends toward 3′ ends of a transcript and concomitantly a decrease in RNA polymerase II occupancy. Therefore, we hypothesize that the lower H3K36me3 levels over EDB exon can be a result of altered splicing upon enhancer deletion when elevated exon inclusion leads to lower H3K36me3. However, such decrease in H3K36me3 was not observed in the downstream regions of EDB exon (exon 38; **Figure 21C**) which would argue against this hypothesis. Nevertheless, such H3K36me3 deposition regulated by splicing was observed within one kilobase downstream of altered splice site (Kim et al. 2011) which can be mitigated in the case of exon 38 *FN1* gene which is located ~17 kilobases downstream of EDB exon. Overall, we think that H3K36me3 mark can be a cause as well as result of altered splicing of EDB exon. However, which option is true has to be investigated in the future.

Additionally to methylation, we also checked H4 acetylation in enhancer-deleted mutants since histone deacetylation was shown to modulate alternative splicing globally as well as locally in the *FN1* gene (Hnilicová et al. 2011). Consistently with the previous study, we observed a correlation between lower H4 acetylation over EDB exon (**Figure 22**) and EDB inclusion (**Figure 20**). The possible mechanism includes the lower RNA polymerase II rate over EDB exon. According to the kinetic coupling model, this could provide more time for splicing factors to recognize and bind the alternative exon and thus lead to higher exon inclusion. However, to definitely conclude this, the RNA polymerase II occupancy and/or rate should be examined by additional experiments. Interestingly, H4 acetylation was slightly elevated immediately downstream of deleted enhancer element and over promoter region (**Figure 22**). Since transcription enhancer regions were shown to be bound by histone acetyltransferases (HATs) such as p300 and CBP (CREB-binding protein) (Heintzman et al. 2007; Heintzman et al. 2009; Visel et al. 2009; Wang et al. 2009; Blow et al. 2010), by deletion of enhancer element, we would expect to see the

decrease in binding of HATs to this region and therefore acetylation. The additional experiments are needed to shed more light on the mechanism of regulation of this locus.

Taken together, our results suggest that our predicted DNA enhancer element is truly a transcription enhancer since upon its deletion the overall FN1 transcription was significantly reduced. Additionally, this deletion also led to altered alternative splicing of FN1 EDB exon, and at the same time, several histone modifications levels over this exon were affected. However, we cannot currently decide whether changes in histone modifications are a result or a cause of the altered splicing. It is possible that the deletion of enhancer leads to different splicing outcome which subsequently affects histone modification marks or *vice versa*. Nevertheless, we propose that alternative splicing can also be modified by enhancer element located several kilobases away of the particular exon including the modulation of chromatin. To specifically determine how splicing of EDB exon, transcription enhancer and chromatin regulation are associated, further experiments have to be done, e.g. check if a chromosomal looping between transcription enhancer region and the promoter/EDB exon of *FN1* gene takes place, and find which if any splicing factors are contributing to EDB splicing regulation.

Splicing of Long Intergenic Non-Coding RNAs and its Importance

Long intergenic non-coding RNAs (lincRNAs) have been extensively studied in recent years, and previous studies have shown that lncRNAs are less efficiently spliced and polyadenylated in comparison to PCGs (Seidl et al. 2006; Tilgner et al. 2012; Mukherjee et al. 2016; Lagarde et al. 2017; Melé et al. 2017; Schlackow et al. 2017), but the reason for this remained unknown. In our study (Krchňáková et al. 2018), we calculated splicing indices for lincRNAs and PCGs as a proxy for splicing efficiencies of expressed lincRNAs. Our findings show less efficient splicing of lincRNAs in comparison to PCGs in all studied cell lines. The bioinformatic results are supported by splicing efficiency analyses of several lincRNAs by quantitative and semi-quantitative RT-PCRs (**Figure 25**, **Figure 26**, **Figure 27**).

To determine factors affecting lincRNA splicing we transiently expressed several lincRNAs from a CMV-driven promoter and did not detect any significant changes in their splicing (**Figure 26**, **Figure 36**). These results indicate that the promoter and the genomic context do not significantly influence lincRNA splicing profile and that inefficient splicing is an intrinsic property of lincRNA transcripts. To identify potential sequences inhibiting splicing, we created a series of deletion mutants that lack different parts of the ncRNA-a2 intron (**Figure 28**, **Figure 29**). We did not find any strong splicing silencers, which is consistent with a bioinformatic analysis that did not reveal any specific accumulation of splicing inhibitory sequences in lincRNAs with respect to PCGs (Krchňáková et al. 2018). However, we found that lincRNAs have longer introns and exons than PCGs (Krchňáková et al. 2018), which might partially explain their less efficient splicing (Klinz and Gallwitz 1985; Sterner et al. 1996; Bell et al. 1998; Fox-Walsh et al. 2005; Dewey et al. 2006; Louloupi et al. 2018). Finally, we analyzed sequences of 5’ and 3’ splice sites and found a positive correlation between the strength of 5’ss and PPT and lincRNA splicing efficiencies (Krchňáková et al. 2018). This finding was further supported by experimental evidence showing that increasing the strength of 5’ and 3’ splice sites significantly improved splicing of model lincRNAs (**Figure 32**, **Figure 33**, **Figure 35**, **Figure 36**, and **Figure 37**).

The 5’ss and PPT sequences are crucial factors for the splicing efficiency in general, but our data suggest that lincRNA requires stronger 5’ss and PPT containing a high number of thymidines to be effectively spliced (Krchňáková et al. 2018). To

understand why lincRNAs are more dependent on basic splice site sequences, we analyzed the presence of known SR protein exonic binding motifs because binding of SR proteins to exons promotes splicing (reviewed in Graveley 2000; Long and Caceres 2009). We found that the majority of analyzed SR-binding sequences are more abundant in PCG exons while only one motif out of 29 analyzed motifs is enriched in lincRNAs (**Figure 39**). Consistently, we show that all analyzed SR proteins exhibited a clear binding preference for PCGs even when we compared lincRNAs with expression-matched PCGs (Krchňáková et al. 2018). This result provides experimental evidence that lincRNAs are unable to secure productive binding of SR proteins. Based on our data we propose a model (**Figure 48**) that lincRNAs lack the cooperative network of positive signals that efficiently navigates the splicing machinery to splice sites. For most lincRNAs, U1 and U2 snRNPs and their auxiliary factors thus have to find splice sites without the help of splicing enhancers, rendering the sequences around exon/intron boundaries more important.

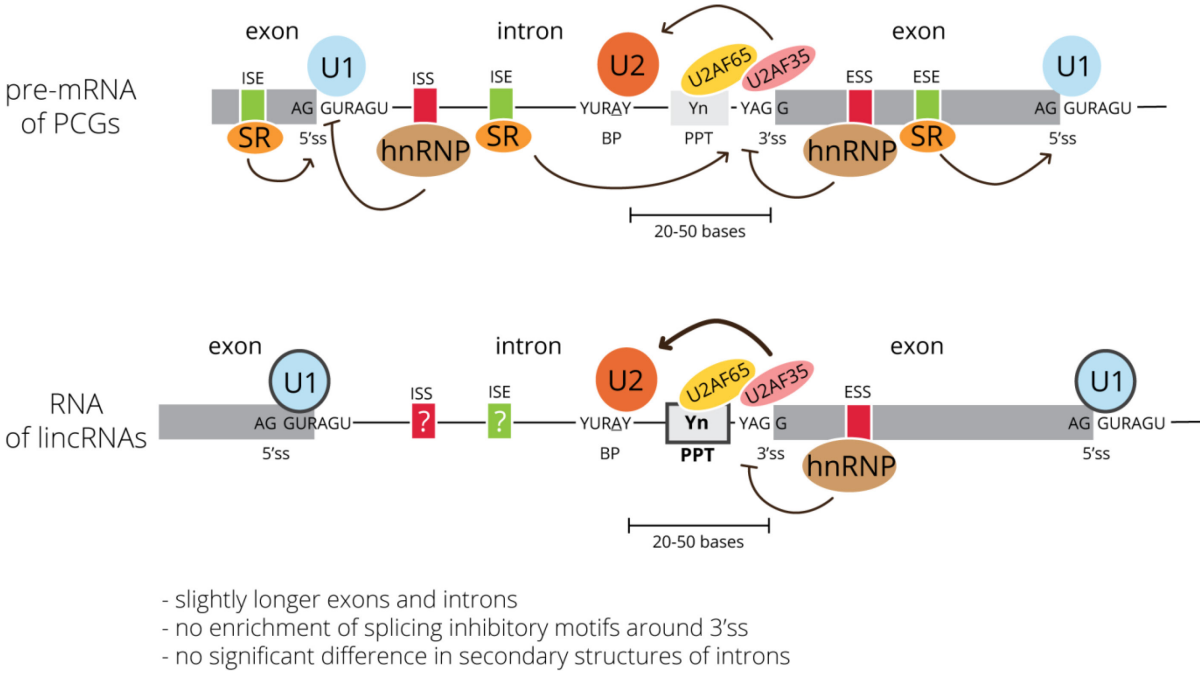

Figure 48. A model of splicing regulation of lincRNAs. We showed that lincRNAs do not contain more splicing inhibitory motifs around 3'ss than PCGs as well as there was no significant difference in the probability of formation of intron secondary structures in lincRNAs than in PCGs. An only slight difference was detected in the lengths of exons and introns between lincRNAs and PCGs. Together with the inefficient binding of SR proteins to lincRNAs, we propose lincRNAs are more dependent on the strengths of 5' and 3'ss, specifically PPT. 5'and 3'ss – 5'and 3'splice sites, BP – branch point, subscribed A represents the base mediating branching of intron lariat, PPT – polypyrimidine tract, ISS/ISE – intron splicing silencer/enhancer, ESS/ESE – exon splicing silencer/enhancer, R – purine, Y – pyrimidine, U1 – U1 snRNP, U2 – U2 snRNP, U2AF1/2 – U2 snRNP auxiliary factor 1/2, hnRNP – hnRNP proteins, SR – SR proteins.

However, it should be noted that the insertion of the ncRNA-a2 intron between PCG exons did not improve splicing efficiency (**Figure 27**). This suggests that complete sequence and context is important for correct splicing, which was recently shown for splicing of various 5'ss sequences (Wong et al. 2018). This is also consistent with studies proposing that the local environment and the continuous sequence of exons and introns information are critical for correct intron definition and removal (Lagarde et al. 2017). During evolution, sequences of PCGs were fine-tuned to ensure a robust recognition of intron/exon boundaries and efficient splicing. In contrast to PCGs, where splicing is an essential component of gene expression, our study (**Figure 47**) and the results of Engreitz et al. (2016a) and Yin et al. (2015) indicate that introns are not essential for the function of lincRNAs. On the contrary, our results of overexpression of ncRNA-a2 (**Figure 43**) and CRISPRa/i (**Figure 44** and **Figure 45**) suggest that the DNA sequence rather the RNA product is important for regulation of neighouring PCGs. Unfortunately, we were not able to provide further experimental evidence for this hypothesis because we were not able to downregulate *ncRNA-a2* expression by insertion of poly(A) sites (**Figure 46**).

However, the observed unimportance of introns and splicing in the function of lincRNAs (**Figure 47**) may result in a lower evolutionary pressure on some lincRNAs to promote efficient splicing. This is quite surprising because the great majority of purifying selection operating on lincRNAs in humans was shown to be splicing-related together with the ability of splicing to modify chromatin through recruitment of splice-coupled chromatin modifiers, which in turn might modulate neighbor gene activity (Schüler et al. 2014). Interestingly, a recent study of Tan et al. (2018) shows that lincRNAs are efficiently and rapidly spliced and that their processing rate is strongly associated with their cognate enhancer activity. The authors propose that lincRNA splicing enhances their transcription and directly impacts the *cis*-regulatory function of their cognate enhancers. This suggests that intron presence or its splicing might play some function in at least some lincRNAs.

There are three most common processes of creating new lncRNAs including (i) duplication of existing lncRNAs which seems to only rarely contribute to the evolution of new lncRNAs, (ii) the utilization of already established transcription via losing the coding potential of PCGs through mutations, insertion of transposable elements or genomic rearrangements without losing the transcription activity or triggering the nonsense-mediated decay (Duret et al. 2006; Romito and Rougeulle 2011), and (iii) exaptation of

previously non-transcribed locus through a series of mutations creating a favourable combination of promoters, splice sites and polyadenylation signals.

U1 snRNP splicing signals are enriched in the sense direction in the case of a divergent transcription and these U1-binding sites can promote transcription via stimulation of RNA polymerase II initiation and elongation (Fong and Zhou 2001; Furger et al. 2002; Kwek et al. 2002; Kotovic et al. 2003; Lin et al. 2008). However, when antisense strand of such divergent locus gains splicing signals through mutations or the insertion of transposable elements, the splicing is favored over polyadenylation and transcription is even further promoted. Additionally, this leads to the suppression of early polyadenylation, and thus stabilization of the transcription product. This can easily transform cryptic transcripts into stable RNAs which can then acquire new functions (Gotea et al. 2013; Wu and Sharp 2013). This is supported by the observation that transposable elements with a functional promoter (such as ERVs – endogenous retroviruses) were shown to be sufficient to drive transcription initiation at a previously non-transcribed locus (Peaston et al. 2004; Faulkner et al. 2009; Jacques et al. 2013; Kapusta et al. 2013; Fort et al. 2014; Lu et al. 2014; Consortium 2015; Göke et al. 2015; Melé et al. 2015) and heavily contribute to the evolution of lncRNAs (Kelley and Rinn 2012; Kapusta et al. 2013; Karlic et al. 2017).

Interestingly, there are also proposed mechanisms how pre-mRNAs protect transposable elements inserted into deep intronic regions from being recognized and spliced in mammals. In the first mechanism, human cells safeguard transcriptome from the aberrant and potentially detrimental exonization of *Alu* transposable elements by the competing binding of hnRNP C and U2AF2 at cryptic splice sites (Zarnack et al. 2013). This mechanism uses the ability of hnRNP C to displace U2AF2 from continuous U-tracts and act as a splicing repressor. However, U2AF2 can also recognize cytidines while hnRNP C preferentially bind continuous uridines. Thus, in the case of alternative splicing, the disruption of U-tract but preserving polypyrimidine tract (PPT) leads to lower hnRNP C binding and higher exon inclusion. In the case of two opposing *Alu* elements located in introns, weakening of hnRNP C binding is sufficient to promote exonization. Indeed, genuine *Alu*-derived exons are estimated to contribute to 5% of all internal alternative exons (Sela et al. 2007; Sorek 2007; Vorechovsky 2010). Although the sudden incorporation of *Alu* elements into mature transcripts is deleterious in the vast majority of cases, this process of creating new exons can also be advantageous. In the presence of

hnRNP C, *Alu* elements are repressed instead of being removed from the genome through selection, allowing them to evolve near-neutrally for longer evolutionary times. Mutations to the U-tracts that change the binding balance of hnRNP C and U2AF2 may allow low levels of exonization, which allows even stronger evolutionary testing by selection. However, even though lincRNAs have higher thymidines over cytidines ratios than PCGs, hnRNP C-bound lincRNAs are spliced as efficiently as unbound lincRNAs (Krchnakova 2018) suggesting that hnRNP C does not play a strong role in splicing repression of lincRNAs.

Another mechanism of protecting transcriptome from aberrant exonization is specific for LINE retrotransposons (Attig et al. 2018) since they contains strong cryptic splice sites in both sense and antisense orientation (Belancio et al. 2006; Merkin et al. 2015) which can also disrupt expression of their host gene and cause a number of hereditary human diseases (Schwahn et al. 1998; Yoshida et al. 1998; Meischl et al. 2000). This protection includes the recognition of a transposable element by MATR3 (Matrin-3) which promotes the binding of PTBP1 and represses splicing and 3'end processing within and around this transposable element. This repression of splicing occurs preferentially on evolutionary young, primate-specific transposable elements, which are depleted in the vicinity of exons. However, the gradual loss of such insulation of transposable elements during evolution leads to diversification of the roles of transposable elements. Older transposable elements are located closer to exons, are a source of new mammalian exons, and are increasingly bound by RNA binding proteins that enhance RNA processing. Thus, as authors propose (Attig et al. 2018), LINEs facilitate evolutionary innovations and contribute to the evolution of new lineage-specific transcripts in mammals. However, how much this mechanism contributes to the evolution and shaping of new lincRNAs is currently not known.

Evolutionary new lincRNAs transcripts may acquire functional 5'ss and 3'ss over time and splicing may change their functional output. The most common outcome of mutations affecting splice sites is exon skipping, followed by cryptic splice site activation (Krawczak et al. 1992; Nakai and Sakamoto 1994). Earlier analysis revealed a higher prevalence of cryptic 5'ss over 3'ss, likely reflecting the less complex nature of the 5'ss sequence (Nakai and Sakamoto 1994). This is probably due to a more complex molecular mechanism involved in recognition of 3'ss over 5'ss. While YAG motif of 3'ss is recognized by the U2AF1 and branch point by U2AF2 which subsequently attracts

U2snRNP to 3’ss, 5’ss is recognized by the base complementarity of U1 snRNA with intron which was shown to be necessary for efficient splicing but not for unique definition of the 5' cleavage site (Séraphin et al. 1988). Indeed, 5'ss can be provided by endogenous retrotransposons (Franke et al. 2017). Analysis of cryptic 3'ss revealed the importance of the 3'ss sequences and showed that intronic *de novo* 3′ss arose mainly by AG-creating mutations in existing functional PPTs. In contrast, exonic *de novo* 3′ss were often induced by mutations improving the PPT, branch-point sequence or distant auxiliary signals (Královicová et al. 2005). Interestingly, a group of disease-causing mutations that create AG dinucleotides in the PPT and activate aberrant 3’ss upstream of branch points shares a similar distance between predicted branch points and newly introduced AGs. Moreover, the reduction of this distance and/or the strength of the new PPT enhanced the expression of natural transcripts (Královicová et al. 2005). We found the strongest correlation between lincRNA splicing and PPT sequence, which suggests that in the absence of functional splicing enhancers, weak PPT sequence and inefficient U2AF binding represent the major barriers that evolutionary new transcripts have to overcome to become efficiently spliced.

Conclusion

In my thesis, I focused on two major topics, namely how gene chromatin and genomic context affects pre-mRNA splicing, and how long non-coding RNAs are spliced and how transcription and splicing affects their activating function.

In the first project, we observed altered both alternative and constitutive splicing upon chromatin modulation. Together with global enrichment of H3K9me3 around internal exons, we hypothesize that it can either be bound by specific proteins and recruit splicing factors or can affect alternative splicing through kinetic coupling by slowing down RNA polymerase II elongation.

In the second project, we showed that transcription enhancer of the *FN1* gene modulates inclusion of the alternative EDB exon. Additionally, we also observed changes in histone methylation and acetylation either immediately downstream of transcription enhancer, promoter sequence or over alternative EDB exon. Therefore, we propose that alternative splicing can also be modified by enhancer element located several kilobases away of the particular exon, likely via the modulating histone modification marks.

Following up the splicing regulation under different conditions, we elucidated the less efficient splicing of long intergenic non-coding RNAs (lincRNAs) when compared to protein-coding genes (PCGs). Based on our results, we propose that lincRNAs are not efficiently bound by SR proteins and lack the cooperative network of positive signals that efficiently navigates the splicing machinery to splice sites. Therefore, their splicing efficiency depends more on the strength of splice sites, mainly 5'ss and polypyrimidine tract.

Finally, we showed that the intron of ncRNA-a2 is not essential for its activating function which is in contrast to PCGs that were fine-tuned to ensure a robust recognition of intron/exon boundaries and efficient splicing during evolution. This indicates the lower evolutionary pressure on some lincRNAs to promote efficient splicing.

Supplementary Material

Supplementary List of Primers

Primers for Quantitative PCR – Co-transcriptional Splicing of FOSL1

Target	sequence
RT primer downstream of poly(A)	5‘ - GGGACCTAGGGCTCCAAATA - 3‘
FOSL1 pre-mRNA (intron 3 – exon 4)	5’- CCTCAGAACCCTGAGTCCAA- 3’
	5’- CTTCTGCTTCTGCAGCTCCT - 3’
FOSL1 mRNA (exon 3 – exon 4)	5’ - CAGGCGGAGACTGACAAACT - 3’
	5’ - CTTCCAGCACCAGCTCTAGG - 3’

Primers for Quantitative PCR – Chromatin Immunoprecipitation

Target	sequence
FN1 downstream of enhancer	5’- ATCCAAATTCATTTTCAAAATTTTGT -3’
	5’- AACTCAGGCTCAGAAAATATGGAC -3’
FN1 promoter (-505 bp[2])	5’- TTGATGACCGCAAAGGAAAC -3’
	5’- TCGCAGCGAACAAAAGAGAT -3’
FN1 exon 24 (upstream of EDB)	5’- GGAAGAAGTGGTCCATGCTG -3’
	5’- GGGACACTTTCCTTGTCATCC -3’
FN1 exon 25 (EDB)	5’- AGGTGCCCCAACTCACTGACC -3’
	5’- TGCCGCAACTACTGTGATGCGGTA -3’
FN1 exon 38	5’- CACCCAATTCCTTGCTGGTA -3’
	5’- GGACCACTTCTCTGGGAGGA -3’
intergenic region	5’- GGCTAATCCTCTATGGGAGTCTGTC -3’
	5’- CCAGGTGCTCAAGGTCAACATC -3’
FOSL1 intron 1 – exon 2	5’ - ACTGCCAAGCTGTGCTCTTT - 3’
	5’ - ACTGCCACTCATGGTGTTGA - 3’

Primers for Quantitative PCR – Splicing Efficiency in Fractions

Target	sequence
ncRNA-a2 spliced variant 1	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- ACCTGGAGCCCGGAAGGAAC - 3’
ncRNA-a2 spliced variant 2	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- GCACTGGAGACGGAAGGAAC - 3’
ncRNA-a2 unspliced	5’- CTCATTCGGTCCATCCAACT - 3’
	5’- GGAGGGGAGGGAGGAAGT - 3’
ncRNA-a5 spliced	5’- CAGGTCAGGTCCTCTGGGTA - 3’
	5’- CATCCCTTTCCTGGGGTAGT - 3’
ncRNA-a5 unspliced	5’- GCTTGAGTCCTCCCAAGGTT - 3’
	5’- CATCCCTTTCCTGGGGTAGT - 3’
GAPDH spliced	5’- ACATCGCTCAGACACCATGG - 3’
	5’- GTTAAAAGCAGCCCTGGTGA - 3’
GAPDH unspliced	5’- CAGGGAAGCTCAAGGGAGAT - 3’
	5’- GTTAAAAGCAGCCCTGGTGA - 3’
LDHA spliced	5’- TGGCAGCCTTTTCCTTAGAA - 3’
	5’- CTTTCTCCCTCTTGCTGACG - 3’

[2] Distance from transcription start site

LDHA unspliced	5'- TGGCAGCCTTTTCCTTAGAA - 3'
	5'- TGTGCAACTGCACTCTACCC - 3'

Primers for Semiquantitative PCR

Target	sequence
ncRNA-a2 semi	5'- CTCATTCGGTCCATCCAACT - 3'
	5'- TGATGGCATTGAATTGGAGA - 3'

Primers for Quantitative PCR – Intron Swap

Target	sequence
MID3-4 RT	5'- CTACAGTGCTGAGTGCGTCT - 3'
ncRNA-a2 spliced variant 1	5'- CTCATTCGGTCCATCCAACT - 3'
	5'- ACCTGGAGCCCGGAAGGAAC - 3'
ncRNA-a2 spliced variant 2	5'- CTCATTCGGTCCATCCAACT - 3'
	5'- GCACTGGAGACGGAAGGAAC - 3'
ncRNA-a2 unspliced	5'- CTGCCGGGTTGCAAACTTG - 3'
	5'- TGGTAGAAGCACGAGCAAGG- 3'
ncRNA-a2 MID3/4	5'- GTGCCCTGGAATGTTTGCTG - 3'
	5'- AGTGCTGAGTGCGTCTCTGA - 3'
HBB spliced	5'- TTGGACCCAGAGGTTCTTTG - 3'
	5'- TGCCCAGGAGCCTGAAGTTC - 3'
HBB unspliced	5'- TAGCAGCTACAATCCAGCTACC - 3'
	5'- CACACAGACCAGCACGTTG - 3'
HBB total	5'- TTGGACCCAGAGGTTCTTTG - 3'
	5'- CCTGAAGTTCTCAGGATCCA - 3'

Primers for Quantitative PCR – Splicing Efficiency of All NcRNA-a2 Mutants

Target	sequence
MID3-4 RT	5'- CTACAGTGCTGAGTGCGTCT - 3'
ncRNA-a2 spliced variant 1	5'- CTCATTCGGTCCATCCAACT - 3'
	5'- ACCTGGAGCCCGGAAGGAAC - 3'
ncRNA-a2 spliced variant 2	5'- CTCATTCGGTCCATCCAACT - 3'
	5'- GCACTGGAGACGGAAGGAAC - 3'
ncRNA-a2 MID3/4	5'- GTGCCCTGGAATGTTTGCTG - 3'
	5'- AGTGCTGAGTGCGTCTCTGA - 3'

Primers for Quantitative PCR – Splicing Efficiency after hnRNP H siRNA KD

Target	sequence
ncRNA-a2 spliced variant 1	5'- CTCATTCGGTCCATCCAACT - 3'
	5'- ACCTGGAGCCCGGAAGGAAC - 3'
ncRNA-a2 spliced variant 2	5'- CTCATTCGGTCCATCCAACT - 3'
	5'- GCACTGGAGACGGAAGGAAC - 3'
ncRNA-a2 unspliced 2	5'- CTGCCGGGTTGCAAACTTG- 3'
	5'- TGGTAGAAGCACGAGCAAGG - 3'
GAPDH unspliced	5'- CAGGGAAGCTCAAGGGAGAT - 3'
	5'- GTTAAAAGCAGCCCTGGTGA - 3'
UBB	5'- GCTTTGTTGGGTGAGCTTGT - 3'
	5'- TCACGAAGATCTGCATTTTGA - 3'

Primers for Quantitative PCR – Splicing Efficiency of other LincRNAs

Target	sequence
MID5-6 RT	5'- CTCGCGATATCGTGTCTGAT- 3'
SNHG8 spliced	5'- TTAGGTGAAAGTCGCCGGGC - 3'
	5'- TCAAACTGACGGTTCTCGGG - 3'
SNHG8 unspliced	5'- GGCTTTGGAAACCCTTAAGT - 3'
	5'- TCAAACTGACGGTTCTCGGG - 3'
BX088651.4spliced	5'- GAAAGCTCAGACTCAGGGCC- 3'
	5'- AAGCACTAGACTGCGCACGC- 3'
BX088651.4unspliced	5'- GAAAGCTCAGACTCAGGGCC- 3'
	5'- AAAACCAGAGGCGGTGGAAG- 3'
BX005266.2spliced	5'- GAAAGCTCAGACTCAGGGCC- 3'
	5'- GGATCGAGATCTGCGCACGC- 3'
BX005266.2unspliced	5'- GAAAGCTCAGACTCAGGGCC- 3'
	5'- AAAACCAGAGGCGGTGGAAG- 3'
AC005840.2 spliced	5'- TGGCAAGTTTACCACCCTGA- 3'
	5'- ACAGTGCTTCCGTCCTCATG- 3'
AC005840.2 unspliced	5'- TGGCAAGTTTACCACCCTGA- 3'
	5'- GAAAGCAGCCATCCCCTTAC- 3'
AC116021.1spliced	5'- CGAGCTCTTGGAAACCCGGG- 3'
	5'- GGGTTGACATGAGGATGGCA- 3'
AC116021.1unspliced	5'- CTTTGCAGGCAACAACCCTC- 3'
	5'- GGGTTGACATGAGGATGGCA- 3'

Primers for Quantitative PCR – Splicing Efficiency of NcRNA-a2 with HBB PPT

Target	sequence
MID3-4 RT	5'- CTACAGTGCTGAGTGCGTCT - 3'
ncRNA-a2 spliced variant 1	5'- CTCATTCGGTCCATCCAACT - 3'
	5'- ACCTGGAGCCCGGAAGGAAC - 3'
ncRNA-a2 spliced variant 2	5'- CTCATTCGGTCCATCCAACT - 3'
	5'- GCACTGGAGACGGAAGGAAC - 3'
ncRNA-a2 MID3/4	5'- GTGCCCTGGAATGTTTGCTG - 3'
	5'- AGTGCTGAGTGCGTCTCTGA - 3'

Primers for Quantitative PCR – U2AF2 RIP

Target	sequence
MID3-4 RT	5'- CTACAGTGCTGAGTGCGTCT - 3'
ncRNA-a2 unspliced 2	5'- CTGCCGGGTTGCAAACTTG- 3'
	5'- TGGTAGAAGCACGAGCAAGG - 3'
ncRNA-a2 MID3/4	5'- GTGCCCTGGAATGTTTGCTG - 3'
	5'- AGTGCTGAGTGCGTCTCTGA - 3'

Primers for Quantitative PCR – siRNA KDs of NcRNA-as

Target	sequence
ncRNA-a2	5'- CTCATTCGGTCCATCCAACT - 3'
	5'- TGGTAGAAGCACGAGCAAGG - 3'
KDM5B	5'- TCAGTGCAGAGAGCCAGAGA - 3'
	5'- GGATAGATCGGCCTCGTGTA - 3'
RABIF	5'- AGGGACCGCTCTCTTCTCTC - 3'
	5'- AGTGTTCCTGGAGGAGATCG - 3'
ADIPOR1	5'- ACATCTGGACCCATCTGCTT - 3'

	5’- CCCAAAAACCACCTTCTCCT - 3’
KLHL12	5’- TCAAGTGCGACGAAATTCAG - 3’
	5’- GCTCTTTCTTGGCATGCTTC - 3’
GAPDH	5’- ACATCGCTCAGACACCATGG - 3’
	5’- GTTAAAAGCAGCCCTGGTGA - 3’
LDHA	5’- TGGCAGCCTTTTCCTTAGAA - 3’
	5’- CTTTCTCCCTCTTGCTGACG - 3’
ACTB	5’- GGCATCCTCACCCTGAAGTA - 3’
	5’- AGGTGTGGTGCCAGATTTTC - 3’
UBB	5’- GCTTTGTTGGGTGAGCTTGT - 3’
	5’- TCACGAAGATCTGCATTTTGA - 3’
ncRNA-a5	5’- CAGGTCAGGTCCTCTGGGTA - 3’
	5’- CATCCCTTTCCTGGGGTAGT - 3’
PQLC3	5’- CCTCAGCCTTCCGAGTTTAC - 3’
	5’- GAGGATGGGGTACTCCAGGT - 3’
E2F6	5’- AGGAATGGGCTCCAGAGAGA - 3’
	5’- TCGGACTCCCAGTTTCGTTG - 3’
ROCK2	5’- TGAAGCCTGACAACATGCTC - 3’
	5’- AATCCGGTGTTCCAACTGCT - 3’
GAPDH	5’- ACATCGCTCAGACACCATGG - 3’
	5’- GTTAAAAGCAGCCCTGGTGA - 3’

Primers for Quantitative PCR – Ectopic Overexpression

Target	sequence
ncRNA-a2	5’- CCTTACTCTTGGACAACACTCC - 3’
	5’- TTGGATGGACCGAATGAGGATG - 3’
KLHL12	5’- ACAAGCCTGCTGTGAGTTCT - 3’
	5’- TCCACCTCTCCTTGACTCAGA - 3’
GAPDH	5’- ATTTGGTCGTATTGGGCGCC - 3’
	5’- TGAGGTCAATGAAGGGGTCA - 3’

Primers for Quantitative PCR – CRISPRa and CRISPRi

Target	sequence
ncRNA-a2	5’- CGCAGTCCATCTCAGCTCAT - 3’
	5’- CAGGGGACATCTGACAGCAA - 3’
KDM5B	5’- TCAGTGCAGAGAGCCAGAGA - 3’
	5’- GGATAGATCGGCCTCGTGTA - 3’
RABIF	5’- AGGGACCGCTCTCTTCTCTC - 3’
	5’- AGTGTTCCTGGAGGAGATCG - 3’
ADIPOR1	5’- ACATCTGGACCCATCTGCTT - 3’
	5’- CCCAAAAACCACCTTCTCCT - 3’
KLHL12	5’- ACAAGCCTGCTGTGAGTTCT - 3’
	5’- TCCACCTCTCCTTGACTCAGA - 3’
UBB	5’- GCTTTGTTGGGTGAGCTTGT - 3’
	5’- TCACGAAGATCTGCATTTTGA - 3’

Primers for Genotyping CRISPR Mutants

Target	sequence
ncRNA-a2 2x pA KI 1st round	5’- CCCTTGCCCTTTATGATGTTTAC - 3’
	5’- CAGGGGACATCTGACAGCAA - 3’
ncRNA-a2 2x pA KI 2nd round	5’- ACTGAGTCTTCACAACCACCATATT - 3’

5’- CAGCAGTAGACGATAGCATAGGAGG - 3’

Primers for Quantitative PCR – NcRNA-a2 2x pA KI

Target	sequence
ncRNA-a2	5’- CCTTACTCTTGGACAACACTCC - 3’
	5’- TTGGATGGACCGAATGAGGATG - 3’
KDM5B	5’- TCAGTGCAGAGAGCCAGAGA - 3’
	5’- GGATAGATCGGCCTCGTGTA - 3’
RABIF	5’- AGGGACCGCTCTCTTCTCTC - 3’
	5’- AGTGTTCCTGGAGGAGATCG - 3’
ADIPOR1	5’- ACATCTGGACCCATCTGCTT - 3’
	5’- CCCAAAAACCACCTTCTCCT - 3’
KLHL12	5’- ACAAGCCTGCTGTGAGTTCT - 3’
	5’- TCCACCTCTCCTTGACTCAGA - 3’
UBB	5’- GCTTTGTTGGGTGAGCTTGT - 3’
	5’- TCACGAAGATCTGCATTTTGA - 3’

Primers for Quantitative PCR – NcRNA-a2 Intron KO

Target	sequence
ncRNA-a2	5’- CCTTACTCTTGGACAACACTCC - 3’
	5’- TTGGATGGACCGAATGAGGATG - 3’
KDM5B	5’- TCAGTGCAGAGAGCCAGAGA - 3’
	5’- GGATAGATCGGCCTCGTGTA - 3’
RABIF	5’- AGGGACCGCTCTCTTCTCTC - 3’
	5’- AGTGTTCCTGGAGGAGATCG - 3’
ADIPOR1	5’- ACATCTGGACCCATCTGCTT - 3’
	5’- CCCAAAAACCACCTTCTCCT - 3’
KLHL12	5’- ACAAGCCTGCTGTGAGTTCT - 3’
	5’- TCCACCTCTCCTTGACTCAGA - 3’
GAPDH	5’- ATTTGGTCGTATTGGGCGCC - 3’
	5’- TGAGGTCAATGAAGGGGTCA - 3’

Supplementary List of Used ISE Motifs

Name of the mutant	sequence *(mutated nucleotides are underlined)*
1xISE	5’- TTTGGGC - 3’
2xISE	5’- TTTGGGCTATTGG - 3’
3xISE	5’- TTTGGGCTTTGGGCTATTGG - 3’
ISEctrl	5’- TTCGCGC - 3’

Supplementary List of Modified PPTs

Name of the mutant	sequence *(mutated nucleotides are underlined)*
ncRNAa-2 WT	5’- TCGCCGCCTCTGACAACTTTT - 3’
ncRNAa-2 T21	5’- TTTTTTTTTTTTTTTTTTTT- 3’
ncRNAa-2 CtoT	5’- TTGTTGTTTTTGATAATTTTT- 3’
ncRNAa-2 GAtoT	5’- TCTCCTCCTCTTTCTTCTTTT- 3’
SNHG8 WT	5’- GCATGCGCGGACTTGAGTGCTCAT - 3’
SNHG8 T25	5’- TTTTTTTTTTTTTTTTTTTTTTTTT- 3’
BX088651.4 WT	5’- CATCGCGTCCTCTTC - 3’
BX088651.4 T15	5’- TTTTTTTTTTTTTT- 3’
BX005266.2 WT	5’-CATCGCGTCCTCTTCCAGTCTAGTGCTTTTTT-3’
BX005266.2 T32	5’- TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT- 3’
AC005840.2 WT	5’- GAGACTCACTCTAGTCTTTTCCCG - 3’
AC005840.2 T24	5’- TTTTTTTTTTTTTTTTTTTTTTTT- 3’
AC116021.1 WT	5’- TCCCTATTTTG - 3’
AC116021.1 T11	5’- TTTTTTTTTTT- 3’

Supplementary List of MaxEnt Scores

Name of the mutant	5‘ss		3‘ss	
	sequence	MaxEnt score	sequence	MaxEnt score
ncRNA-a2	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a5 intron 2	GAGgtaagc	9.85	ttgagatccttcttttttcagGTG	9.33
GAPDH intron 2	CGGgtgagt	9.89	accctcacgtattcccccagGTT	8.32
LDHA intron 3	AAGgttgat	4.19	attattcccctttttctctagACT	8.08
ncRNA-a2 HBB intron	CCGgtgagt	10.90	ctcttatcttcctcccacagGGC	12.56
ncRNA-a2 HBB intron ncRNA-a2 5‘ss	CCGgtaacc	5.28	ctcttatcttcctcccacagGGC	12.56
HBB WT intron 2	AGGgtgagt	9.25	ctcttatcttcctcccacagCTC	11.43
HBB ncRNA-a2 intron	AGGgtaacc	5.40	cgcctctgacaacttttcagCTC	4.97
HBB ncRNA-a2 intron HBB 5‘ss	AGGgtgagt	9.25	cgcctctgacaacttttcagCTC	4.97
ncRNA-a2 FΔ1=Δ1	CCGgtaacg	7.34	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 FΔ2	CCGgtaacg	7.34	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 FΔ3	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 FΔ4	CCGgtaaca	5.92	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 FΔ5	CCGgtaact	6.39	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 FΔ6	CCGgtaaca	5.92	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 FΔ7	CCGgtaaca	5.92	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 FΔ8	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 RΔ1=Δ8	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 RΔ2	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 RΔ3	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 RΔ4	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 RΔ5	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 RΔ6	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 RΔ7	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 Δ2	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31

ncRNA-a2 Δ3	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 Δ4	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 Δ5	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 Δ6	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 Δ7	CCGgtaacc	5.28	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 7.53 5‘ss	ACGgtaact	7.53	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 9.6 5’ss	CAGgtgagc	9.6	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 11.08 5‘ss	CAGgtaagg	11.08	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 FΔ1 7.66 5‘ss	CAGgtgacg	7.66	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 FΔ1 9.35 5’ss	CCGgtgagg	9.35	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 FΔ1 11.08 5‘ss	CAGgtaagg	11.08	cgcctctgacaacttttcagGGC gatcctctcgtctcccccagTCT	6.13 9.31
ncRNA-a2 T21 7.53 5‘ss	ACGgtaact	7.53	ttttttttttttttttcagGGC gatcctctcgtctcccccagTCT	13.12 9.31
ncRNA-a2 T21 9.6 5’ss	CAGgtgagc	9.6	ttttttttttttttttcagGGC gatcctctcgtctcccccagTCT	13.12 9.31
ncRNA-a2 T21 11.08 5‘ss	CAGgtaagg	11.08	ttttttttttttttttcagGGC gatcctctcgtctcccccagTCT	13.12 9.31
ncRNA-a2 T21	CCGgtaacc	5.28	ttttttttttttttttcagGGC	13.12
ncRNA-a2 CtoT	CCGgtaacc	5.28	tgtttttgataatttttcagGGC	8.62
ncRNA-a2 GAtoT	CCGgtaacc	5.28	ctcctctttcttcttttcagGGC	13.39
ncRNA-a2 ΔPPT	CCGgtaacc	5.28	tcgccgcctctgacaaccagGGC	4.76
ncRNA-a2 3xISE-T21	CCGgtaacc	5.28	ttttttttttttttttcagGGC	13.12
ncRNA-a2 with HBB PPT	CCGgtaacc	5.28	ctcttatcttcctcccacagGGC	12.56
SNHG8 WT	AAAgtacgt	7.99	cggacttgagtgctcattagGTC	-0.49
SNHG8 T25	AAAgtacgt	7.99	ttttttttttttttttagGTC	13.77
BX088651.4 WT	CAGgtaaag	9.65	tccatcgcgtcctcttccagTCT	8.29
BX088651.4 T15	CAGgtaaag	9.65	tctttttttttttttcagTCT	10.63
BX005266.2 WT	CAGgtaaag	9.65	agtctagtgctttttccagATC	5.60
BX005266.2 T32	CAGgtaaag	9.65	ttttttttttttttccagATC	12.67
AC005840.2 WT	ACGgtaagg	10.38	actctagtctttcccgcagGAA	9.46
AC005840.2 T24	ACGgtaagg	10.38	ttttttttttttttttcagGAA	13.15
AC116021.1 WT	TTGgtaaaa	3.23	ttatcatccctattttgcagGAA	10.48
AC116021.1 T11	TTGgtaaaa	3.23	ttatcatttttttttttcagGAA	12.22

Supplementary List of Intron Sequences

Name	Sequence	Length (n)
ncRNA-a2 intron	5’- gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctcccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctc cccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctcgccgcct ctgacaacttttcag - 3’	204
ncRNA-a2 HBB intron 2 ncRNA-a2 5‘ss	5’- gtaaccgcatgggacgcttgatgttttctttccccttcttttctatggttaagttcatgtcataggaagg ggataagtaacagggtacagtttagaatgggaaacagacgaatgattgcatcagtgtggaagtct caggatcgttttagtttcttttatttgctgttcataacaattgttttcttttgtttaattcttgctttctttttttttc ttctccgcaatttttactattatacttaatgccttaacattgtgtataacaaaaggaaatatctctgagat acattaagtaacttaaaaaaaaactttacacagtctgcctagtacattactatttggaatatatgtgtg cttatttgcatattcataatctccctactttatttttctttttatttttaattgatacataatcattatacatatttat gggttaaagtgtaatgttttaatatgtgtacacatattgaccaaatcagggtaattttgcatttgtaattt taaaaaatgctttcttcttttaatatactttttttgtttatcttatttctaatactttccctaatctctttctttcag ggcaataatgatacaatgtatcatgcctctttgcaccattctaaagaataacagtgataatttctgggt taaggcaatagcaatatctctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttca tattgctaatagcagctacaatccagctaccattctgcttttatttttatggttgggataaggctggatta ttctgagtccaagctaggccctttgctaatcatgttcatacctcttatcttcctcccacag - 3’	850
ncRNA-a2 intron with HBB PPT	5’- gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctcccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctc cccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctgttcatacct cttatcttcctcccacag - 3’	209
ncRNA-a2 HBB 5‘ss	5’- gtgagtgcgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctcccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctc cccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctcgccgcct ctgacaacttttcag - 3’	204
ncRNA-a2 T21 HBB 5‘ss	5’- gtgagtgcgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctcccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctc cccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccttttttttttttttt ttttttcag - 3’	204
HBB intron 2	5’- gtgagtctatgggacgcttgatgttttctttccccttcttttctatggttaagttcatgtcataggaagg ggataagtaacagggtacagtttagaatgggaaacagacgaatgattgcatcagtgtggaagtct caggatcgttttagtttcttttatttgctgttcataacaattgttttcttttgtttaattcttgctttctttttttttc ttctccgcaatttttactattatacttaatgccttaacattgtgtataacaaaaggaaatatctctgagat acattaagtaacttaaaaaaaaactttacacagtctgcctagtacattactatttggaatatatgtgtg cttatttgcatattcataatctccctactttatttttctttttatttttaattgatacataatcattatacatatttat gggttaaagtgtaatgttttaatatgtgtacacatattgaccaaatcagggtaattttgcatttgtaattt taaaaaatgctttcttcttttaatatactttttttgtttatcttatttctaatactttccctaatctctttctttcag ggcaataatgatacaatgtatcatgcctctttgcaccattctaaagaataacagtgataatttctgggt taaggcaatagcaatatctctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttca tattgctaatagcagctacaatccagctaccattctgcttttatttttatggttgggataaggctggatta ttctgagtccaagctaggccctttgctaatcatgttcatacctcttatcttcctcccacag - 3’	850
HBB ncRNA-a2 intron HBB	5’- gtgagtctgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctcccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctc	204

5‘ss		cccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctcgccgcct ctgacaacttttcag - 3’	
HBB intron 2 with ncRNA-a2 PPT	5’-	gtgagtctatgggacgcttgatgttttctttcccccttctttttctatggttaagttcatgtcataggaagg ggataagtaacagggtacagtttagaatgggaaacagacgaatgattgcatcagtgtggaagtct caggatcgttttagtttctttttatttgctgttcataacaattgttttctttgtttaattcttgctttctttttttttc ttctccgcaatttttactattatacttaatgccttaacattgtgtataacaaaaggaaatatctctgagat acattaagtaacttaaaaaaaaactttacacagtctgcctagtacattactatttggaatatatgtgtg cttatttgcatattcataatctccctactttatttttctttttatttttaattgatacataatcattatacatatttat gggttaaagtgtaatgttttaatatgtgtacacatattgaccaaatcagggtaattttgcatttgtaattt taaaaaatgctttcttctttttaatatactttttgtttatcttatttctaatactttccctaatctctttctttcag ggcaataatgatacaatgtatcatgcctctttgcaccattctaaagaataacagtgataatttctgggt taaggcaatagcaatatctctgcatataaatatttctgcatataaattgtaactgatgtaagaggtttca tattgctaatagcagctacaatccagctaccattctgcttttattttatggttgggataaggctggatta ttctgagtccaagctaggccctttgctaatcatcgccgcctctgacaacttttcag - 3’	845
WT	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctcccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctc cccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctcgccgcct ctgacaacttttcag - 3’	204
FΔ1= Δ1	5’-	gtaacgccgggttgcaaacttgggggggacttcctccctcccctcccccctgggcgccgtgcaact gccctgggaccgggttctgggatgaggggggcagaccgggctccccagcggccggcgcagc acgtagcgcacgtgtagggtccgctccccacccccctcgccgcctctgacaacttttcag - 3’	184
FΔ2	5’-	gtaacggacttcctccctcccctcccccctgggcgccgtgcaactgccctgggaccgggttctgg gatgaggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggt ccgctccccacccccctcgccgcctctgacaacttttcag - 3’	164
FΔ3	5’-	gtaacccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccg ggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctcgc cgcctctgacaacttttcag - 3’	144
FΔ4	5’-	gtaacccctgggaccgggttctgggatgaggggggcagaccgggctccccagcggccggcg cagcacgtagcgcacgtgtagggtccgctccccacccccctcgccgcctctgacaacttttcag - 3’	124
FΔ5	5’-	gtaacatgaggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgta gggtccgctccccacccccctcgccgcctctgacaacttttcag - 3’	104
FΔ6	5’-	gtaactccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctcg ccgcctctgacaacttttcag - 3’	84
FΔ7	5’-	gtaacacgtagcgcacgtgtagggtccgctccccacccccctcgccgcctctgacaacttttcag - 3’	64
FΔ8	5’-	gtaacccgctccccacccccctcgccgcctctgacaacttttcag - 3’	44
RΔ1= Δ8	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctcccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctc cccagcggccggcgcagcccgctccccacccccctcgccgcctctgacaacttttcag - 3’	184

RΔ2	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggacttcctccctcccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggcc cgctccccacccccctcgccgcctctgacaacttttcag - 3’	164
RΔ3	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggacttcctccctcccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggccgctccccacccccctcgccgcc tctgacaacttttcag - 3’	144
RΔ4	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggacttcctccctcccctcc ccctgggcgccgtgcaactgccgctccccacccccctcgccgcctctgacaacttttcag - 3’	124
RΔ5	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggacttcctccctcccctcc ccgctccccacccccctcgccgcctctgacaacttttcag - 3’	104
RΔ6	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggccgctccccacccccctcg ccgcctctgacaacttttcag - 3’	84
RΔ7	5’-	gtaaccgcgttgcgaagaccacgctccgctccccacccccctcgccgcctctgacaacttttcag - 3’	64
Δ2	5’-	gtaaccgcgttgcgaagaccacgctggacttcctccctcccctcccccctgggcgccgtgcaact gccctgggaccgggttctgggatgaggggggcagaccgggctccccagcggccggcgcagc acgtagcgcacgtgtagggtccgctccccacccccctcgccgcctctgacaacttttcag - 3’	184
Δ3	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttggggccctgggcgccgtgcaac tgccctgggaccgggttctgggatgaggggggcagaccgggctccccagcggccggcgcag cacgtagcgcacgtgtagggtccgctccccacccccctcgccgcctctgacaacttttcag - 3’	184
Δ4	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggacttcctccctcccctcc ccctgggaccgggttctgggatgaggggggcagaccgggctccccagcggccggcgcagca cgtagcgcacgtgtagggtccgctccccacccccctcgccgcctctgacaacttttcag - 3’	184
Δ5	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggacttcctccctcccctcc ccctgggcgccgtgcaactgatgaggggggcagaccgggctccccagcggccggcgcagca cgtagcgcacgtgtagggtccgctccccacccccctcgccgcctctgacaacttttcag - 3’	184
Δ6	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggacttcctccctcccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggtccccagcggccggcgcagcac gtagcgcacgtgtagggtccgctccccacccccctcgccgcctctgacaacttttcag - 3’	184
Δ7	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggacttcctccctcccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggca cgtagcgcacgtgtagggtccgctccccacccccctcgccgcctctgacaacttttcag - 3’	184
1xISE	5’-	gtaaccgcgttgcgaagaccacgcttttgggcgccgggttgcaaacttgggggacttcctccct ccccctcccccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagac cgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctc gccgcctctgacaacttttcag - 3’	211
2xISE	5’-	gtaaccgcgttgcgaagaccacgcttttgggctattgggccgggttgcaaacttgggggacttc ctccctcccctcccccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggg cagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacc ccctcgccgcctctgacaacttttcag - 3’	217

3xISE	5’-	gtaaccgcgttgcgaagaccacgcttttgggctttgggctattgggccgggttgcaaacttgggg ggacttcctccctcccctccccctgggcgccgtgcaactgccctgggaccgggttctgggatga gggggggcagaccgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgct ccccacccccctcgccgcctctgacaacttttcag - 3’	224
ISEctrl	5’-	gtaaccgcgttgcgaagaccacgctttcgcgcgccgggttgcaaacttgggggggacttcctccct ccccctcccccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagac cgggctccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctc gccgcctctgacaacttttcag - 3’	211
Δ60	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctccccctcc tccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctcgccgc ctctgacaacttttcag - 3’	144
3xISE-Δ60	5’-	gtaaccgcgttgcgaagaccacgcttttgggctttgggctattgggccgggttgcaaacttgggg ggacttcctccctccccctcctccccagcggccggcgcagcacgtagcgcacgtgtagggtccg ctccccacccccctcgccgcctctgacaacttttcag - 3’	164
T21	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctccccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctc ccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccttttttttttttt ttttttcag - 3’	204
CtoT	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctccccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctc ccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccttgttgttttttg ataatttttcag - 3’	204
GAtoT	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctccccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctc ccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctctcctcctct ttcttctttttcag - 3’	204
ΔPPT	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctccccctcc ccctgggcgccgtgcaactgccctgggaccgggttctgggatgaggggggcagaccgggctc ccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccctcgccgcct ctgacaaccag - 3’	200
Δ60-T21	5’-	gtaaccgcgttgcgaagaccacgctgccgggttgcaaacttgggggggacttcctccctccccctcc tccccagcggccggcgcagcacgtagcgcacgtgtagggtccgctccccacccccttttttttttt ttttttttcag - 3’	144

Supplementary List of SRSF Binding Motifs

Protein	Motif	Reference
SRSF1	GARGARGARG	(Müller-McNicoll et al. 2016)
	TGRWG	(Mueller and Hertel 2011)
	RGAAGAAC	(Mueller and Hertel 2011)
	AGGACRRAGC	(Mueller and Hertel 2011)
	SRSASGA	(Mueller and Hertel 2011)
SRSF2	RTCTGWAGA	(Müller-McNicoll et al. 2016)
	GRYYMCYR	(Paz et al. 2010)
	TGCYGYY	(Paz et al. 2010)
	AGSAGAGTA	(Mueller and Hertel 2011)
	TGTTCSAGWT	(Mueller and Hertel 2011)
	AGGAGAT	(Mueller and Hertel 2011)
	GRYYCSYR	(Mueller and Hertel 2011)
SRSF3	TCWTCHTC	(Müller-McNicoll et al. 2016)
	WCWWC	(Mueller and Hertel 2011)
	CTCKTCY	(Mueller and Hertel 2011)
SRSF4	AWGAWGAWGA	(Müller-McNicoll et al. 2016)
	GAAGGA	(Mueller and Hertel 2011)
SRSF5	YDTCTGWAGACA	(Müller-McNicoll et al. 2016)
	YYWCWSG	(Paz et al. 2010)
	ACDGS	(Mueller and Hertel 2011)
SRSF6	RARGAWGA	(Müller-McNicoll et al. 2016)
	GAAGAAGA	(Müller-McNicoll et al. 2016)
	TSCGKM	(Mueller and Hertel 2011)
	YRCRKM	(Paz et al. 2010)
SRSF7	WRAWGAHRA	(Müller-McNicoll et al. 2016)
	TCAACA	(Mueller and Hertel 2011)
	ACGAGAGAY	(Mueller and Hertel 2011)
SRSF9	GACGAC	(Mueller and Hertel 2011)
	CTGGATT	(Mueller and Hertel 2011)

Supplementary List of sgRNAs and Guide Target Sequences

CRISPRa and CRISPRi

Guide	sequence
ncRNA-a2 CRISPRa/i	5'- GGGCCGCATTGTCTCGTCTAGGG- 3'

CRISPR 2x pA KI

Guide	sequence
ncRNA-a2 2xpA KI	5'- GGCCTCGCACCCAGCCCTCTGGG- 3'

Guide target	sequence
ncRNA-a2 2xpA KI	5'- CCCAGAGGGCTGGGTGCGAGGCC - 3'

Inserted sequence
5'-
aattcactcctcaggtgcaggctgcctatcagaaggtggtggctggtgtggccaatgccctggctcacaaataccactgagatctttttccctctg
ccaaaaattatggggacatcatgaagccccttgagcatctgacttctggctaataaaggaaatttattttcattgcaatagtgtgttggaatttttgtg
tctctcactcggaaggacatatgggagggcaaatcatttaaaacatcagaatgagtatttggtttagagtttggcaacatatgcccatatgctggct
gccatgaacaaaggttggctataaagaggtcatcagtatatgaaacagccccctgctgtccattccttattccatagaaaagccttgacttgaggtt
agatttttttttatattttgttttgtgttatttttttctttaacatccctaaaatttttccttacatgttttactagccagatttttcctcctctcctgactactcccagt
catagctgtccctcttctcttatggagaagcttcgactgtgccttctagttgccagccatctgttgtttgcccctccccccgtgccttccttgaccctgg
aaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggggc
aggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatgaattctaagctt -3'

- rabbit β-globin polyadenylation signal from Addgene plasmid #13777
- bovine growth hormone polyadenylation signal from Addgene plasmid 13445

References

Akerman M, David-Eden H, Pinter RY, Mandel-Gutfreund Y. 2009. A computational approach for genome-wide mapping of splicing factor binding sites. *Genome Biol* **10**(3): R30.

Alam T, Medvedeva YA, Jia H, Brown JB, Lipovich L, Bajic VB. 2014. Promoter Analysis Reveals Globally Differential Regulation of Human Long Non-Coding RNA and Protein-Coding Genes. *PLoS ONE* **9**(10): e109443.

Alekseyenko AV, Kim N, Lee CJ. 2007. Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. *RNA* **13**(5): 661-670.

Allo M, Buggiano V, Fededa JP, Petrillo E, Schor I, Mata M. 2009. Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat Struct Mol Biol* **16**.

Almada AE, Wu X, Kriz AJ, Burge CB, Sharp PA. 2013. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**(7458): 360-363.

Almeida SF, Grosso AR, Koch F, Fenouil R, Carvalho S, Andrade J. 2011. Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat Struct Mol Biol* **18**.

Alvarez-Dominguez JR, Hu W, Yuan B, Shi J, Park SS, Gromatzky AA, Oudenaarden Av, Lodish HF. 2014. Global discovery of erythroid long noncoding RNAs reveals novel regulators of red cell maturation. *Blood* **123**(4): 570-581.

Amit M, Donyo M, Hollander D, Goren A, Kim E, Gelfman S, Lev-Maor G, Burstein D, Schwartz S, Postolsky B et al. 2012. Differential GC Content between Exons and Introns Establishes Distinct Strategies of Splice-Site Recognition. *Cell Rep* **1**(5): 543-556.

An S, Yeo KJ, Jeon YH, Song J-J. 2011. Crystal Structure of the Human Histone Methyltransferase ASH1L Catalytic Domain and Its Implications for the Regulatory Mechanism. *J Biol Chem* **286**(10): 8369-8374.

Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T et al. 2014a. An atlas of active enhancers across human cell types and tissues. *Nature* **507**(7493): 455-461.

Andersson R, Refsing Andersen P, Valen E, Core LJ, Bornholdt J, Boyd M, Heick Jensen T, Sandelin A. 2014b. Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nature Communications* **5**: 5336.

Änkö M-L, Müller-McNicoll M, Brandl H, Curk T, Gorup C, Henry I, Ule J, Neugebauer KM. 2012. The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol* **13**(3): R17.

Ares M, Grate L, Pauling MH. 1999. A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* **5**(9): 1138-1139.

Arnone MI, Davidson EH. 1997. The hardwiring of development: organization and function of genomic regulatory systems. *Development* **124**(10): 1851-1864.

Ashe HL, Monks J, Wijgerde M, Fraser P, Proudfoot NJ. 1997. Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev* **11**.

Aspden JL, Eyre-Walker YC, Phillips RJ, Amin U, Mumtaz MAS, Brocard M, Couso J-P. 2014. Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. In *Elife*, Vol 3, p. e03528.

Ast G. 2004. How did alternative splicing evolve? *Nature Reviews Genetics* **5**: 773.

Athanasiadis A, Rich A, Maas S. 2004. Widespread A-to-I RNA Editing of Alu-Containing mRNAs in the Human Transcriptome. *PLoS Biol* **2**(12): e391.

Attig J, Agostini F, Gooding C, Singh A, Chakrabarti AM, Haberman N, Emmett W, Smith CW, Luscombe NM, Ule J. 2018. Heteromeric RNP assembly at LINEs controls lineage-specific RNA processing. *bioRxiv*.

Auboeuf D, Hönig A, Berget SM, O'Malley BW. 2002. Coordinate Regulation of Transcription and Splicing by Steroid Receptor Coregulators. *Science* **298**(5592): 416-419.

Babushok DV, Ostertag EM, Kazazian HH. 2007. Current topics in genome evolution: Molecular mechanisms of new gene formation. *Cellular and Molecular Life Sciences* **64**(5): 542-554.

Banerjee AR, Kim YJ, Kim TH. 2014. A novel virus-inducible enhancer of the interferon-β gene with tightly linked promoter and enhancer activities. *Nucleic Acids Res* **42**(20): 12537-12554.

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**(6824): 120-124.

Barbosa-Morais NL, Carmo-Fonseca M, Aparício S. 2006. Systematic genome-wide annotation of spliceosomal proteins reveals differential gene family expansion. *Genome Res* **16**(1): 66-77.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129**.

Bass BL. 2002. RNA Editing by Adenosine Deaminases That Act on RNA. *Annu Rev Biochem* **71**(1): 817-846.

Batsché E, Yaniv M, Muchardt C. 2005. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nature Structural &Amp; Molecular Biology* **13**: 22.

Bazzini AA, Johnstone TG, Christiano R, Mackowiak SD, Obermayer B, Fleming ES, Vejnar CE, Lee MT, Rajewsky N, Walther TC et al. 2014. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. *EMBO J* **33**(9): 981-993.

Belancio VP, Hedges DJ, Deininger P. 2006. LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res* **34**(5): 1512-1521.

Bell MV, Cowper AE, Lefranc M-P, Bell JI, Screaton GR. 1998. Influence of Intron Length on Alternative Splicing of CD44. *Mol Cell Biol* **18**(10): 5930-5941.

Berget SM. 1995. Exon Recognition in Vertebrate Splicing. *J Biol Chem* **270**(6): 2411-2414.

Berget SM, Moore C, Sharp PA. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci USA* **74**(8): 3171-3175.

Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S et al. 2004. Global Identification of Human Transcribed Sequences with Genome Tiling Arrays. *Science* **306**(5705): 2242-2246.

Beyer AL, Osheim YN. 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev* **2**(6): 754-765.

Bhatt Dev M, Pandya-Jones A, Tong A-J, Barozzi I, Lissner Michelle M, Natoli G, Black Douglas L, Smale Stephen T. 2012. Transcript Dynamics of Proinflammatory Genes Revealed by Sequence Analysis of Subcellular RNA Fractions. *Cell* **150**(2): 279-290.

Bieberstein Nicole I, Carrillo Oesterreich F, Straube K, Neugebauer Karla M. 2012. First Exon Length Controls Active Chromatin Signatures and Transcription. *Cell Rep* **2**(1): 62-68.

Bieberstein NI, Kozáková E, Huranová M, Thakur PK, Krchňáková Z, Krausová M, Carrillo Oesterreich F, Staněk D. 2016. TALE-directed local modulation of H3K9 methylation shapes exon recognition. *Sci Rep* **6**: 29961.

Birney E Stamatoyannopoulos JA Dutta A Guigo R Gingeras TR Margulies EH Weng Z Snyder M Dermitzakis ET Thurman RE et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**(7146): 799-816.

Blackwood EM, Kadonaga JT. 1998. Going the Distance: A Current View of Enhancer Action. *Science* **281**(5373): 60-63.

Blencowe BJ, Bowman JAL, McCracken S, Rosonina E. 1999. SR-related proteins and the processing of messenger RNA precursors. *Biochem Cell Biol* **77**(4): 277-291.

Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F et al. 2010. ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet* **42**: 806.

Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczyński B, Riddell A, Furlong EEM. 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* **44**: 148.

Botti V, McNicoll F, Steiner MC, Richter FM, Solovyeva A, Wegener M, Schwich OD, Poser I, Zarnack K, Wittig I et al. 2017. Cellular differentiation state modulates the mRNA export activity of SR proteins. *J Cell Biol* **216**(7): 1993-2009.

Boucher L, Ouzounis CA, Enright AJ, Blencowe BJ. 2001. A genome-wide survey of RS domain proteins. *RNA* **7**(12): 1693-1701.

Bourque G. 2009. Transposable elements in gene regulation and in the evolution of vertebrate genomes. *Curr Opin Genet Dev* **19**(6): 607-612.

Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS. 2012. High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. *Science* **335**(6068): 552-557.

Braunschweig U, Gueroussov S, Plocik AM, Graveley Brenton R, Blencowe Benjamin J. 2013. Dynamic Integration of Splicing within Gene Regulatory Pathways. *Cell* **152**(6): 1252-1269.

Brent MR, Guigó R. 2004. Recent advances in gene structure prediction. *Curr Opin Struct Biol* **14**(3): 264-272.

Brillen A-L, Schöneweis K, Walotka L, Hartmann L, Müller L, Ptok J, Kaisers W, Poschmann G, Stühler K, Buratti E et al. 2017. Succession of splicing regulatory elements determines cryptic 5´ss functionality. *Nucleic Acids Res* **45**(7): 4202-4216.

Brody Y, Neufeld N, Bieberstein N, Causse SZ, Böhnlein E-M, Neugebauer KM, Darzacq X, Shav-Tal Y. 2011. The In Vivo Kinetics of RNA Polymerase II Elongation during Co-Transcriptional Splicing. *PLoS Biol* **9**(1): e1000573.

Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, Willard HF. 1992. The human *XIST* gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**(3): 527-542.

Brugiolo M, Herzel L, Neugebauer KM. 2013. Counting on co-transcriptional splicing. *F1000Prime Rep* **5**: 9.

Bulger M, Groudine M. 2011. Functional and Mechanistic Diversity of Distal Transcription Enhancers. *Cell* **144**(3): 327-339.

Buratti E, Baralle FE. 2004. Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process. *Mol Cell Biol* **24**(24): 10505-10514.

Buratti E, Muro AF, Giombi M, Gherbassi D, Iaconcig A, Baralle FE. 2004. RNA Folding Affects the Recruitment of SR Proteins by Mouse and Human Polypurinic Enhancer Elements in the Fibronectin EDA Exon. *Mol Cell Biol* **24**(3): 1387-1400.

Buratti E, Stuani C, De Prato G, Baralle FE. 2007. SR protein-mediated inhibition of CFTR exon 9 inclusion: molecular characterization of the intronic splicing silencer. *Nucleic Acids Res* **35**(13): 4359-4368.

Burge CB, Padgett RA, Sharp PA. 1998. Evolutionary Fates and Origins of U12-Type Introns. *Mol Cell* **2**(6): 773-785.

Busch A, Hertel KJ. 2012. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdisciplinary Reviews: RNA* **3**(1): 1-12.

Cabili MN, Dunagin MC, McClanahan PD, Biaesch A, Padovan-Merhar O, Regev A, Rinn JL, Raj A. 2015. Localization and abundance analysis of human lncRNAs at single-cell and single-molecule resolution. *Genome Biol* **16**(1): 20.

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* **25**(18): 1915-1927.

Cáceres JF, Misteli T, Screaton GR, Spector DL, Krainer AR. 1997. Role of the Modular Domains of SR Proteins in Subnuclear Localization and Alternative Splicing Specificity. *The Journal of Cell Biology* **138**(2): 225-238.

Califice S, Baurain D, Hanikenne M, Motte P. 2012. A Single Ancient Origin for Prototypical Serine/Arginine-Rich Splicing Factors. *Plant Physiol* **158**(2): 546-560.

Calo E, Wysocka J. 2013. Modification of Enhancer Chromatin: What, How, and Why? *Mol Cell* **49**(5): 825-837.

Canzio D, Chang EY, Shankar S, Kuchenbecker KM, Simon MD, Madhani HD, Narlikar GJ, Al-Sady B. 2011. Chromodomain-Mediated Oligomerization of HP1 Suggests a Nucleosome-Bridging Mechanism for Heterochromatin Assembly. *Mol Cell* **41**(1): 67-81.

Cao W, Jamison SF, Garcia-Blanco MA. 1997. Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro. *RNA* **3**(12): 1456-1467.

Carlevaro-Fita J, Rahim A, Guigó R, Vardy LA, Johnson R. 2016. Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA* **22**(6): 867-882.

Carlini DB, Genut JE. 2006. Synonymous SNPs Provide Evidence for Selective Constraint on Human Exonic Splicing Enhancers. *J Mol Evol* **62**(1): 89-98.

Carmel I, Tal S, Vig I, Ast G. 2004. Comparative analysis detects dependencies among the 5′ splice-site positions. *RNA* **10**(5): 828-840.

Carmel L, Wolf YI, Rogozin IB, Koonin EV. 2007. Three distinct modes of intron dynamics in the evolution of eukaryotes. *Genome Res* **17**(7): 1034-1044.

Carninci P Kasukawa T Katayama S Gough J Frith MC Maeda N Oyama R Ravasi T Lenhard B Wells C et al. 2005. The Transcriptional Landscape of the Mammalian Genome. *Science* **309**(5740): 1559-1563.

Carrillo Oesterreich F, Preibisch S, Neugebauer KM. 2010. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol Cell* **40**.

Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR et al. 2005. Chromosome-Wide Mapping of Estrogen Receptor Binding Reveals Long-Range Regulation Requiring the Forkhead Protein FoxA1. *Cell* **122**(1): 33-43.

Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoute J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF et al. 2006. Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* **38**: 1289.

Carstens RP, McKeehan WL, Garcia-Blanco MA. 1998. An Intronic Sequence Element Mediates Both Activation and Repression of Rat Fibroblast Growth Factor Receptor 2 Pre-mRNA Splicing. *Mol Cell Biol* **18**(4): 2205-2217.

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res* **31**(13): 3568-3571.

Cavaloc Y, Bourgeois CF, Kister L, Stévenin J. 1999. The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers. *RNA* **5**(3): 468-483.

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, Baller JA, Somia NV, Bogdanove AJ, Voytas DF. 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**(12): e82-e82.

Cermak T, Starker CG, Voytas DF. 2015. Efficient Design and Assembly of Custom TALENs Using the Golden Gate Platform. In *Chromosomal Mutagenesis*, (ed. SM Pruett-Miller), pp. 133-159. Springer New York, New York, NY.

Clark MB, Johnston RL, Inostroza-Ponta M, Fox AH, Fortini E, Moscato P, Dinger ME, Mattick JS. 2012. Genome-wide analysis of long noncoding RNA stability. *Genome Res* **22**(5): 885-898.

Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB. 2009. An Architectural Role for a Nuclear Noncoding RNA: *NEAT1* RNA Is Essential for the Structure of Paraspeckles. *Mol Cell* **33**(6): 717-726.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA et al. 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**(6121): 819-823.

Consortium G. 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**(6235): 648-660.

Cordin O, Hahn D, Beggs JD. 2012. Structure, function and regulation of spliceosomal RNA helicases. *Curr Opin Cell Biol* **24**(3): 431-438.

Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46**: 1311.

Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science* **322**(5909): 1845-1848.

Corvelo A, Hallegger M, Smith CWJ, Eyras E. 2010. Genome-Wide Association between Branch Point Properties and Alternative Splicing. *PLoS Comp Biol* **6**(11): e1001016.

Cramer P, Bushnell DA, Kornberg RD. 2001. Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. *Science* **292**(5523): 1863-1876.

Cramer P, Cáceres JF, Cazalla D, Kadener S, Muro AF, Baralle FE, Kornblihtt AR. 1999. Coupling of Transcription with Alternative Splicing: RNA Pol II Promoters Modulate SF2/ASF and 9G8 Effects on an Exonic Splicing Enhancer. *Mol Cell* **4**(2): 251-258.

Cramer P, Pesce CG, Baralle FE, Kornblihtt AR. 1997. Functional association between promoter structure and transcript alternative splicing. *Proceedings of the National Academy of Sciences* **94**(21): 11456-11460.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences* **107**(50): 21931-21936.

Csűrös M, Rogozin IB, Koonin EV. 2008. Extremely Intron-Rich Genes in the Alveolate Ancestors Inferred with a Flexible Maximum-Likelihood Approach. *Mol Biol Evol* **25**(5): 903-911.

Cui K, Zang C, Roh T-Y, Schones DE, Childs RW, Peng W, Zhao K. 2009. Chromatin Signatures in Multipotent Human Hematopoietic Stem Cells Indicate the Fate of Bivalent Genes during Differentiation. *Cell Stem Cell* **4**(1): 80-93.

Curado J, Iannone C, Tilgner H, Valcárcel J, Guigó R. 2015. Promoter-like epigenetic signatures in exons displaying cell type-specific splicing. *Genome Biol* **16**(1): 236.

Danko Charles G, Hah N, Luo X, Martins André L, Core L, Lis John T, Siepel A, Kraus WL. 2013. Signaling Pathways Differentially Affect RNA Polymerase II Initiation, Pausing, and Elongation Rate in Cells. *Mol Cell* **50**(2): 212-222.

Das R, Yu J, Zhang Z, Gygi MP, Krainer AR, Gygi SP, Reed R. 2007. SR Proteins Function in Coupling RNAP II Transcription to Pre-mRNA Splicing. *Mol Cell* **26**(6): 867-881.

David CJ, Manley JL. 2010. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev* **24**(21): 2343-2364.

David CJ, Manley JL. 2011. The RNA polymerase C-terminal domain: a new role in spliceosome assembly. *Transcription* **2**(5): 221-225.

Davis CA, Ares M. 2006. Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **103**(9): 3262-3267.

De Conti L, Baralle M, Buratti E. 2013. Exon and intron definition in pre-mRNA splicing. *Wiley Interdisciplinary Reviews: RNA* **4**(1): 49-60.

de la Mata M, Lafaille C, Kornblihtt AR. 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**(5): 904-912.

De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei C-L, Natoli G. 2010. A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers. *PLoS Biol* **8**(5): e1000384.

de Souza FSJ, Franchini LF, Rubinstein M. 2013. Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? *Mol Biol Evol* **30**(6): 1239-1251.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res* **22**(9): 1775-1789.

Deutsch M, Long M. 1999. Intron-exon structures of eukaryotic model organisms. *Nucleic Acids Res* **27**(15): 3219-3228.

Deveson IW, Brunck ME, Blackburn J, Tseng E, Hon T, Clark TA, Clark MB, Crawford J, Dinger ME, Nielsen LK et al. 2018. Universal Alternative Splicing of Noncoding Exons. *Cell Systems* **6**(2): 245-255.e245.

Dewey CN, Rogozin IB, Koonin EV. 2006. Compensatory relationship between splice sites and exonic splicing signals depending on the length of vertebrate introns. *Bmc Genomics* **7**(1): 311.

Dietrich RC, Incorvaia R, Padgett RA. 1997. Terminal Intron Dinucleotide Sequences Do Not Distinguish between U2- and U12-Dependent Introns. *Mol Cell* **1**(1): 151-160.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F et al. 2012. Landscape of transcription in human cells. *Nature* **489**(7414): 101-108.

Dominski Z, Kole R. 1991. Selection of splice sites in pre-mRNAs with short internal exons. *Mol Cell Biol* **11**(12): 6075-6083.

Doolittle RF. 1995. The Multiplicity of Domains in Proteins. *Annu Rev Biochem* **64**(1): 287-314.

Drummond DA, Raval A, Wilke CO. 2006. A single determinant dominates the rate of yeast protein evolution. *Mol Biol Evol* **23**(2): 327-337.

Dujardin G, Buratti E, Charlet-Berguerand N, Martins de Araujo M, Mbopda A, Le Jossic-Corcos C, Pagani F, Ferec C, Corcos L. 2010. CELF proteins regulate CFTR pre-mRNA splicing: essential role of the divergent domain of ETR-3. *Nucleic Acids Res* **38**(20): 7273-7285.

Dujardin G, Lafaille C, Mata M, Marasco LE, Munoz MJ, Jossic-Corcos C. 2014. How slow RNA polymerase II elongation favors alternative exon skipping. *Mol Cell* **54**.

Dujardin G, Lafaille C, Petrillo E, Buggiano V, Gómez Acuña LI, Fiszbein A, Godoy Herz MA, Nieto Moreno N, Muñoz MJ, Alló M et al. 2013. Transcriptional elongation and alternative splicing. *Biochim Biophys Acta* **1829**(1): 134-140.

Duret L, Chureau C, Samain S, Weissenbach J, Avner P. 2006. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* **312**.

Dušková E, Hnilicová J, Staněk D. 2014. CRE promoter sites modulate alternative splicing via p300-mediated histone acetylation. *RNA biology* **11**(7): 865-874.

Dutertre M, Sanchez G, De Cian M-C, Barbier J, Dardenne E, Gratadou L, Dujardin G, Le Jossic-Corcos C, Corcos L, Auboeuf D. 2010. Cotranscriptional exon skipping in the genotoxic stress response. *Nature Structural &Amp; Molecular Biology* **17**: 1358.

Encode Project Consortium T Dunham I Kundaje A Aldred SF Collins PJ Davis CA Doyle F Epstein CB Frietze S Harrow J et al. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57.

Engreitz JM, Haines JE, Perez EM, Munson G, Chen J, Kane M, McDonel PE, Guttman M, Lander ES. 2016a. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**(7629): 452-455.

Engreitz JM, Ollikainen N, Guttman M. 2016b. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat Rev Mol Cell Biol* **17**(12): 756-770.

Eperon LP, Graham IR, Griffiths AD, Eperon IC. 1988. Effects of RNA secondary structure on alternative splicing of Pre-mRNA: Is folding limited to a region behind the transcribing RNA polymerase? *Cell* **54**(3): 393-401.

Erkelenz S, Mueller WF, Evans MS, Busch A, Schöneweis K, Hertel KJ, Schaal H. 2013. Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA* **19**(1): 96-102.

Fairbrother WG, Holste D, Burge CB, Sharp PA. 2004. Single Nucleotide Polymorphism–Based Validation of Exonic Splicing Enhancers. *PLoS Biol* **2**(9): e268.

Fairbrother WG, Yeh R-F, Sharp PA, Burge CB. 2002. Predictive Identification of Exonic Splicing Enhancers in Human Genes. *Science* **297**(5583): 1007-1013.

Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T et al. 2009. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* **41**(5): 563-571.

Fedorov A, Merican AF, Gilbert W. 2002. Large-scale comparison of intron positions among animal, plant, and fungal genes. *Proceedings of the National Academy of Sciences* **99**(25): 16128-16133.

Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD. 2006. The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev* **20**.

Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME et al. 2010. Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences* **107**(19): 8689-8694.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics* **9**: 397.

Fica SM, Tuttle N, Novak T, Li N-S, Lu J, Koodathingal P, Dai Q, Staley JP, Piccirilli JA. 2013. RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**: 229.

Fong N, Kim H, Zhou Y, Ji X, Qiu J, Saldi T, Diener K, Jones K, Fu X-D, Bentley DL. 2014. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev* **28**(23): 2663-2676.

Fong YW, Zhou Q. 2001. Stimulatory effect of splicing factors on transcriptional elongation. *Nature* **414**: 929.

Fort A, Hashimoto K, Yamada D, Salimullah M, Keya CA, Saxena A, Bonetti A, Voineagu I, Bertin N, Kratz A et al. 2014. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat Genet* **46**(6): 558-566.

Fox-Walsh KL, Dou Y, Lam BJ, Hung S-p, Baldi PF, Hertel KJ. 2005. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc Natl Acad Sci USA* **102**(45): 16176-16181.

Franke V, Ganesh S, Karlic R, Malik R, Pasulka J, Horvat F, Kuzman M, Fulka H, Cernohorska M, Urbanova J et al. 2017. Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. *Genome Res* **27**(8): 1384-1394.

Fu X-D, Ares Jr M. 2014. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat Rev Genet* **15**(10): 689-701.

Fu Y, Bannach O, Chen H, Teune J-H, Schmitz A, Steger G, Xiong L, Barbazuk WB. 2009. Alternative splicing of anciently exonized 5S rRNA regulates plant transcription factor TFIIIA. *Genome Res* **19**(5): 913-921.

Furger A, Binnie JMOS, Alexandra, Lee BA, Proudfoot NJ. 2002. Promoter proximal splice sites enhance transcription. *Genes Dev* **16**(21): 2792-2799.

Gal-Mark N, Schwartz S, Ram O, Eyras E, Ast G. 2009. The Pivotal Roles of TIA Proteins in 5′ Splice-Site Selection of Alu Exons and Across Evolution. *Plos Genetics* **5**(11): e1000717.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**(10): R80.

Germann S, Gratadou L, Dutertre M, Auboeuf D. 2012. Splicing Programs and Cancer. *Journal of Nucleic Acids* **2012**: 9.

Gilbert Luke A, Larson Matthew H, Morsut L, Liu Z, Brar Gloria A, Torres Sandra E, Stern-Ginossar N, Brandman O, Whitehead Evan H, Doudna Jennifer A et al. 2013. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **154**(2): 442-451.

Gilbert W. 1978. Why genes in pieces? *Nature*.

Girard C, Will CL, Peng J, Makarov EM, Kastner B, Lemm I, Urlaub H, Hartmuth K, Lührmann R. 2012. Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nature Communications* **3**: 994.

Giudice J, Cooper TA. 2014. RNA-Binding Proteins in Heart Development. In *Systems Biology of RNA Binding Proteins*, (ed. GW Yeo), pp. 389-429. Springer New York, New York, NY.

Goguel V, Rosbash M. 1993. Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. *Cell* **72**(6): 893-901.

Göke J, Lu X, Chan Y-S, Ng H-H, Ly L-H, Sachs F, Szczerbinska I. 2015. Dynamic Transcription of Distinct Classes of Endogenous Retroviral Elements Marks Specific Populations of Early Human Embryonic Cells. *Cell Stem Cell* **16**(2): 135-141.

Göke J, Ng HH. 2016. CTRL+INSERT: retrotransposons and their contribution to regulation and innovation of the transcriptome. *EMBO reports* **17**(8): 1131-1144.
Gómez Acuña LI, Fiszbein A, Alló M, Schor IE, Kornblihtt AR. 2013. Connections between chromatin signatures and splicing. *Wiley Interdisciplinary Reviews: RNA* **4**(1): 77-91.
Görnemann J, Kotovic KM, Hujer K, Neugebauer KM. 2005. Cotranscriptional Spliceosome Assembly Occurs in a Stepwise Fashion and Requires the Cap Binding Complex. *Mol Cell* **19**(1): 53-63.
Gotea V, Petrykowska HM, Elnitski L. 2013. Bidirectional Promoters as Important Drivers for the Emergence of Species-Specific Transcripts. *PLoS ONE* **8**(2): e57323.
Graveley BR. 2000. Sorting out the complexity of SR protein functions. *RNA* **6**(9): 1197-1211.
Graveley BR. 2005. Mutually Exclusive Splicing of the Insect Dscam Pre-mRNA Directed by Competing Intronic RNA Secondary Structures. *Cell* **123**(1): 65-73.
Graveley BR, Hertel KJ, Maniatis T. 2001. The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *RNA* **7**(6): 806-818.
Guil S, Esteller M. 2012. Cis-acting noncoding RNAs: friends and foes. *Nat Struct Mol Biol* **19**(11): 1068-1075.
Guo M, Lo PC, Mount SM. 1993. Species-specific signals for the splicing of a short Drosophila intron in vitro. *Mol Cell Biol* **13**(2): 1104-1118.
Guo R, Zheng L, Park Juw W, Lv R, Chen H, Jiao F, Xu W, Mu S, Wen H, Qiu J et al. 2014. BS69/ZMYND11 Reads and Connects Histone H3.3 Lysine 36 Trimethylation-Decorated Chromatin to Regulated Pre-mRNA Processing. *Mol Cell* **56**(2): 298-310.
Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP et al. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**.
Guttman M, Russell P, Ingolia Nicholas T, Weissman Jonathan S, Lander Eric S. 2013. Ribosome Profiling Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins. *Cell* **154**(1): 240-251.
Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, McClanahan P, Hendrickson DG, Sauvageau M, Kelley DR et al. 2014. Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nature Structural &Amp; Molecular Biology* **21**: 198.
Haerty W, Ponting CP. 2015. Unexpected selection to retain high GC content and splicing enhancers within exons of multiexonic lncRNA loci. *RNA* **21**(3): 333-346.
Hager GL, Nagaich AK, Johnson TA, Walker DA, John S. 2004. Dynamics of nuclear receptor movement and transcription. *Biochim Biophys Acta* **1677**(1): 46-51.
Hah N, Murakami S, Nagari A, Danko CG, Kraus WL. 2013. Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res* **23**(8): 1210-1223.
Harlen KM, Churchman LS. 2017. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature Reviews Molecular Cell Biology* **18**: 263.
Hawkins JD. 1988. A survey on intron and exon lengths. *Nucleic Acids Res* **16**(21): 9893-9908.
Haygood R, Babbitt CC, Fedrigo O, Wray GA. 2010. Contrasts between adaptive coding and noncoding changes during human evolution. *Proceedings of the National Academy of Sciences* **107**(17): 7853-7857.
Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**(7243): 108-112.
Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**(3): 311-318.
Henriques T, Scruggs BS, Inouye MO, Muse GW, Williams LH, Burkholder AB, Lavender CA, Fargo DC, Adelman K. 2018. Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev* **32**(1): 26-41.
Hezroni H, Koppstein D, Schwartz Matthew G, Avrutin A, Bartel David P, Ulitsky I. 2015. Principles of Long Noncoding RNA Evolution Derived from Direct Comparison of Transcriptomes in 17 Species. *Cell Rep* **11**(7): 1110-1122.

Hicks MJ, Mueller WF, Shepard PJ, Hertel KJ. 2010. Competing Upstream 5′ Splice Sites Enhance the Rate of Proximal Splicing. *Mol Cell Biol* **30**(8): 1878-1886.
Hirose Y, Manley JL. 2000. RNA polymerase II and the integration of nuclear events. *Genes Dev* **14**(12): 1415-1429.
Hnilicová J, Hozeifi S, Dusková E, Icha J, Tománková T, Stanek D. 2011. Histone Deacetylase Activity Modulates Alternative Splicing. *PLoS ONE* **6**(2): 1-11.
Hnilicová J, Hozeifi S, Stejskalová E, Dušková E, Poser I, Humpolíčková J, Hof M, Staněk D, Matera AG. 2013. The C-terminal domain of Brd2 is important for chromatin interaction and regulation of transcription and alternative splicing. *Molecular Biology of the Cell* **24**(22): 3557-3568.
Ho Y, Elefant F, Liebhaber SA, Cooke NE. 2006. Locus Control Region Transcription Plays an Active Role in Long-Range Gene Activation. *Mol Cell* **23**(3): 365-375.
House AE, Lynch KW. 2006. An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. *Nature Structural &Amp; Molecular Biology* **13**: 937.
Hsieh C-L, Fei T, Chen Y, Li T, Gao Y, Wang X, Sun T, Sweeney CJ, Lee G-SM, Chen S et al. 2014. Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proceedings of the National Academy of Sciences* **111**(20): 7319-7324.
Hsin J-P, Manley JL. 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* **26**(19): 2119-2137.
Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotech* **31**(9): 827-832.
Huang H, Yu S, Liu H, Sun X. 2012. Nucleosome organization in sequences of alternative events in human genome. *BioSyst* **109**(2): 214-219.
Hurst LD, Smith NGC. 1999. Molecular evolutionary evidence that H19 mRNA is functional. *Trends in Genetics* **15**(4): 134-135.
Cheah MT, Wachter A, Sudarsan N, Breaker RR. 2007. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* **447**: 497.
Chen W, Luo L, Zhang L. 2010. The organization of nucleosomes around splice sites. *Nucleic Acids Res* **38**(9): 2788-2798.
Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G et al. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**.
Chew G-L, Pauli A, Rinn JL, Regev A, Schier AF, Valen E. 2013. Ribosome profiling reveals resemblance between long non-coding RNAs and 5′ leaders of coding RNAs. *Development* **140**(13): 2828-2834.
Chodavarapu RK, Feng S, Bernatavichute YV, Chen P-Y, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ et al. 2010. Relationship between nucleosome positioning and DNA methylation. *Nature* **466**: 388.
Chodroff RA, Goodstadt L, Sirey TM, Oliver PL, Davies KE, Green ED, Molnár Z, Ponting CP. 2010. Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome Biol* **11**(7): R72.
Chou M-Y, Rooke N, Turck CW, Black DL. 1999. hnRNP H Is a Component of a Splicing Enhancer Complex That Activates a c-src Alternative Exon in Neuronal Cells. *Mol Cell Biol* **19**(1): 69-77.
Chow LT, Gelinas RE, Broker TR, Roberts RJ. 1977. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**(1): 1-8.
Ibrahim EC, Schaal TD, Hertel KJ, Reed R, Maniatis T. 2005. Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. *Proc Natl Acad Sci USA* **102**(14): 5002-5007.
Ilott NE, Heward JA, Roux B, Tsitsiou E, Fenwick PS, Lenzi L, Goodhead I, Hertz-Fowler C, Heger A, Hall N et al. 2014. Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes. *Nature Communications* **5**: 3979.

Ingolia Nicholas T, Brar Gloria A, Stern-Ginossar N, Harris Michael S, Talhouarne Gaëlle JS, Jackson Sarah E, Wills Mark R, Weissman Jonathan S. 2014. Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Rep* **8**(5): 1365-1379.

International Human Genome Sequencing C Lander ES Linton LM Birren B Nusbaum C Zody MC Baldwin J Devon K Dewar K Doyle M et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860.

Ip JY, Schmidt D, Pan Q, Ramani AK, Fraser AG, Odom DT, Blencowe BJ. 2011. Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome Res* **21**(3): 390-401.

Irimia M, Blencowe BJ. 2012. Alternative splicing: decoding an expansive regulatory layer. *Curr Opin Cell Biol* **24**(3): 323-332.

Irimia M, Rukov JL, Penny D, Roy SW. 2007. Functional and evolutionary analysis of alternatively spliced genes is consistent with an early eukaryotic origin of alternative splicing. *BMC Evol Biol* **7**(1): 188.

Jacques P-É, Jeyakani J, Bourque G. 2013. The Majority of Primate-Specific Regulatory Sequences Are Derived from Transposable Elements. *Plos Genetics* **9**(5): e1003504.

Jaillon O, Bouhouche K, Gout J-F, Aury J-M, Noel B, Saudemont B, Nowacki M, Serrano V, Porcel BM, Ségurens B et al. 2008. Translational control of intron splicing in eukaryotes. *Nature* **451**: 359.

Jasnovidova O, Klumpler T, Kubicek K, Kalynych S, Plevka P, Stefl R. 2017a. Structure and dynamics of the RNAPII CTDsome with Rtt103. *Proceedings of the National Academy of Sciences* **114**(42): 11133-11138.

Jasnovidova O, Krejcikova M, Kubicek K, Stefl R. 2017b. Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p. *EMBO reports* **18**(6): 906-913.

Jonkers I, Kwak H, Lis JT. 2014. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* **3**: e02407.

Kadener S, Cramer P, Nogués G, Cazalla D, de la Mata M, Fededa JP, Werbajh SE, Srebrow A, Kornblihtt AR. 2001. Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing. *EMBO J* **20**(20): 5759-5768.

Kadener S, Fededa JP, Rosbash M, Kornblihtt AR. 2002. Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. *Proceedings of the National Academy of Sciences* **99**(12): 8185-8190.

Kaikkonen Minna U, Spann Nathanael J, Heinz S, Romanoski Casey E, Allison Karmel A, Stender Joshua D, Chun Hyun B, Tough David F, Prinjha Rab K, Benner C et al. 2013. Remodeling of the Enhancer Landscape during Macrophage Activation Is Coupled to Enhancer Transcription. *Mol Cell* **51**(3): 310-325.

Kandul NP, Noor MA. 2009. Large introns in relation to alternative splicing and gene evolution: a case study of Drosophila bruno-3. *BMC Genet* **10**(1): 67.

Kanopka A, Mühlemann O, Akusjärvi G. 1996. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* **381**: 535.

Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL et al. 2007. RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription. *Science* **316**(5830): 1484-1488.

Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C. 2013. Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. *Plos Genetics* **9**(4): e1003470.

Karlic R, Ganesh S, Franke V, Svobodova E, Urbanova J, Suzuki Y, Aoki F, Vlahovicek K, Svoboda P. 2017. Long non-coding RNA exchange during the oocyte-to-embryo transition in mice. *DNA research : an international journal for rapid publication of reports on genes and genomes* **24**(2): 129-141.

Kasparek P, Krausova M, Haneckova R, Kriz V, Zbodakova O, Korinek V, Sedlacek R. 2014. Efficient gene targeting of the Rosa26 locus in mouse zygotes using TALE nucleases. *FEBS Letters* **588**(21): 3982-3988.

Ke S, Zhang XH-F, Chasin LA. 2008. Positive selection acting on splicing motifs reflects compensatory evolution. *Genome Res* **18**(4): 533-543.
Kelley D, Rinn J. 2012. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol* **13**(11): R107.
Kent LB, Robertson HM. 2009. Evolution of the sugar receptors in insects. *BMC Evol Biol* **9**(1): 41.
Keren H, Lev-Maor G, Ast G. 2010. Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews Genetics* **11**: 345.
Kerényi Z, Mérai Z, Hiripi L, Benkovics A, Gyula P, Lacomme C, Barta E, Nagy F, Silhavy D. 2008. Inter-kingdom conservation of mechanism of nonsense-mediated mRNA decay. *EMBO J* **27**(11): 1585-1595.
Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A et al. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences* **106**(28): 11667-11672.
Khodor YL, Menet JS, Tolan M, Rosbash M. 2012. Cotranscriptional splicing efficiency differs dramatically between Drosophila and mouse. *RNA* **18**(12): 2174-2186.
Khodor YL, Rodriguez J, Abruzzi KC, Tang CH, Marr MT, Rosbash M. 2011. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev* **25**.
Kim E, Magen A, Ast G. 2007. Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res* **35**(1): 125-131.
Kim H, Klein R, Majewski J, Ott J. 2004. Estimating rates of alternative splicing in mammals and invertebrates. *Nat Genet* **36**: 915.
Kim S, Kim H, Fong N, Erickson B, Bentley DL. 2011. Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proceedings of the National Academy of Sciences* **108**(33): 13564-13569.
Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Jing W, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**(7295): 182-187.
Kizer KO, Phatnani HP, Shibata Y, Hall H, Greenleaf AL, Strahl BD. 2005. A Novel Domain in Set2 Mediates RNA Polymerase II Interaction and Couples Histone H3 K36 Methylation with Transcript Elongation. *Mol Cell Biol* **25**(8): 3305-3316.
Klinz F-J, Gallwitz D. 1985. Size and position of intervening sequences are critical for the splicing efficiency of pre-mRNA in the yeast Saccharomyces cerevisiae. *Nucleic Acids Res* **13**(11): 3791-3804.
Kohtz JD, Jamison SF, Will CL, Zuo P, Lührmann R, Garcia-Blanco MA, Manley JL. 1994. Protein–protein interactions and 5' and 3' splice-site recognition in mammalian mRNA precursors. *Nature* **368**: 119.
Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, Spicuglia S, de la Chapelle AL, Heidemann M, Hintermair C et al. 2011. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature Structural &Amp; Molecular Biology* **18**: 956.
Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet* **41**.
Kolkman JA, Stemmer WPC. 2001. Directed evolution of proteins by exon shuffling. *Nat Biotechnol* **19**: 423.
Konarska MM, Vilardell J, Query CC. 2006. Repositioning of the Reaction Intermediate within the Catalytic Center of the Spliceosome. *Mol Cell* **21**(4): 543-553.
Kondrashov FA, Koonin EV. 2001. Origin of alternative splicing by tandem exon duplication. *Hum Mol Genet* **10**(23): 2661-2669.
Kondrashov FA, Koonin EV. 2003. Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences. *Trends in Genetics* **19**(3): 115-119.

König J, Zarnack K, Rot G, Curk T, Kayikci M, Zupan B, Turner DJ, Luscombe NM, Ule J. 2010. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol* **17**: 909.

Konkel DA, Tilghman SM, Leder P. 1978. The sequence of the chromosomal mouse β-globin major gene: Homologies in capping, splicing and poly(A) sites. *Cell* **15**(4): 1125-1132.

Kornblihtt AR, Vibe-Pedersen K, Baralle FE. 1984. Human fibronectin: molecular cloning evidence for two mRNA species differing by an internal segment coding for a structural domain. *EMBO J* **3**(1): 221-226.

Kornienko AE, Guenzl PM, Barlow DP, Pauler FM. 2013. Gene regulation by the act of long non-coding RNA transcription. *BMC Biol* **11**(1): 59.

Kotovic KM, Lockshon D, Boric L, Neugebauer KM. 2003. Cotranscriptional Recruitment of the U1 snRNP to Intron-Containing Genes in Yeast. *Mol Cell Biol* **23**(16): 5768-5779.

Kouzarides T. 2007. Chromatin Modifications and Their Function. *Cell* **128**(4): 693-705.

Královicová J, Christensen MB, Vorechovský I. 2005. Biased exon/intron distribution of cryptic and de novo 3' splice sites. *Nucleic Acids Res* **33**(15): 4882-4898.

Krawczak M, Reiss J, Cooper DN. 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet* **90**(1-2): 41-54.

Krecic AM, Swanson MS. 1999. hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol* **11**(3): 363-371.

Krchňáková Z, Thakur PK, Krausová M, Bieberstein N, Haberman N, Müller-McNicoll M, Staněk D. 2018. Splicing of long non-coding RNAs primarily depends on polypyrimidine tract and 5′ splice-site sequences due to weak interactions with SR proteins. *Nucleic Acids Res*: doi: 10.1093/nar/gky1114.

Krull M, Brosius Jr, Schmitz Jr. 2005. Alu-SINE Exonization: En Route to Protein-Coding Function. *Mol Biol Evol* **22**(8): 1702-1711.

Kubicek K, Cerna H, Holub P, Pasulka J, Hrossova D, Loehr F, Hofr C, Vanacova S, Stefl R. 2012. Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev* **26**(17): 1891-1896.

Kubíček K, Pasulka J, Černá H, Löhr F, Štefl R. 2011. 1H, 13C, and 15N resonance assignments for the CTD-interacting domain of Nrd1 bound to Ser5-phosphorylated CTD of RNA polymerase II. *Biomolecular NMR Assignments* **5**(2): 203-205.

Kurihara Y, Matsui A, Hanada K, Kawashima M, Ishida J, Morosawa T, Tanaka M, Kaminuma E, Mochizuki Y, Matsushima A et al. 2009. Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in Arabidopsis. *Proceedings of the National Academy of Sciences* **106**(7): 2453-2458.

Kwak H, Fuda NJ, Core LJ, Lis JT. 2013. Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. *Science* **339**(6122): 950-953.

Kwek KY, Murphy S, Furger A, Thomas B, O'Gorman W, Kimura H, Proudfoot NJ, Akoulitchev A. 2002. U1 snRNA associates with TFIIH and regulates transcriptional initiation. *Nat Struct Biol* **9**: 800.

Lacadie SA, Rosbash M. 2005. Cotranscriptional Spliceosome Assembly Dynamics and the Role of U1 snRNA:5′ss Base Pairing in Yeast. *Mol Cell* **19**(1): 65-75.

Lacadie SA, Tardiff DF, Kadener S, Rosbash M. 2006. In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. *Genes Dev* **20**(15): 2055-2066.

Lagarde J, Uszczynska-Ratajczak B, Carbonell S, Pérez-Lluch S, Abad A, Davis C, Gingeras TR, Frankish A, Harrow J, Guigo R et al. 2017. High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. *Nat Genet* **49**: 1731.

Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**(6824): 116-120.

Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhattar R. 2013. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**(7438): 497-501.

Lam MTY, Cho H, Lesch HP, Gosselin D, Heinz S, Tanaka-Oishi Y, Benner C, Kaikkonen MU, Kim AS, Kosaka M et al. 2013. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498**: 511.

Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J et al. 2010. Dynamic changes in the human methylome during differentiation. *Genome Res* **20**(3): 320-331.

Lawrence M, Gentleman R, Carey V. 2009. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25**(14): 1841-1842.

Lear AL, Eperon LP, Wheatley IM, Eperon IC. 1990. Hierarchy for 5′ splice site preference determined in vivo. *J Mol Biol* **211**(1): 103-115.

Lee Y, Rio DC. 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu Rev Biochem* **84**: 291-323.

Lenhard B, Sandelin A, Carninci P. 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Reviews Genetics* **13**: 233.

Letunic I, Copley RR, Bork P. 2002. Common exon duplication in animals and its role in alternative splicing. *Hum Mol Genet* **11**(13): 1561-1567.

Lev-Maor G, Goren A, Sela N, Kim E, Keren H, Doron-Faigenboim A, Leibman-Barak S, Pupko T, Ast G. 2007. The "Alternative" Choice of Constitutive Exons throughout Evolution. *Plos Genetics* **3**(11): e203.

Lev-Maor G, Ram O, Kim E, Sela N, Goren A, Levanon EY, Ast G. 2008. Intronic Alus Influence Alternative Splicing. *Plos Genetics* **4**(9): e1000204.

Lev-Maor G, Sorek R, Shomron N, Ast G. 2003. The Birth of an Alternatively Spliced Exon: 3' Splice-Site Selection in Alu Exons. *Science* **300**(5623): 1288-1291.

Levine M. 2010. Transcriptional Enhancers in Animal Development and Evolution. *Curr Biol* **20**(17): R754-R763.

Li B, Howe L, Anderson S, Yates JR, Workman JL. 2003. The Set2 Histone Methyltransferase Functions through the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. *J Biol Chem* **278**(11): 8897-8903.

Li G, Ruan X, Auerbach Raymond K, Sandhu Kuljeet S, Zheng M, Wang P, Poh Huay M, Goh Y, Lim J, Zhang J et al. 2012. Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation. *Cell* **148**(1–2): 84-98.

Li J, Moazed D, Gygi SP. 2002. Association of the Histone Methyltransferase Set2 with RNA Polymerase II Plays a Role in Transcription Elongation. *J Biol Chem* **277**(51): 49383-49388.

Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X et al. 2013. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**: 516.

Li W, Notani D, Rosenfeld MG. 2016. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nature Reviews Genetics* **17**: 207.

Lin S, Coutinho-Mansfield G, Wang D, Pandit S, Fu X-D. 2008. The splicing factor SC35 has an active role in transcriptional elongation. *Nature Structural &Amp; Molecular Biology* **15**: 819.

Lin S, Fu X-D. 2007. SR proteins and related factors in alternative splicing. *Adv Exp Med Biol* **623**: 107-122.

Ling J, Ainol L, Zhang L, Yu X, Pi W, Tuan D. 2004. HS2 Enhancer Function Is Blocked by a Transcriptional Terminator Inserted between the Enhancer and the Promoter. *J Biol Chem* **279**(49): 51704-51713.

Lipshitz HD, Peattie DA, Hogness DS. 1987. Novel transcripts from the Ultrabithorax domain of the bithorax complex. *Genes Dev* **1**(3): 307-322.

Listerman I, Sapra AK, Neugebauer KM. 2006. Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nature Structural &Amp; Molecular Biology* **13**: 815.

Liu H-X, Zhang M, Krainer AR. 1998. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev* **12**(13): 1998-2012.

Liu M, Grigoriev A. 2004. Protein domains correlate strongly with exons in multiple eukaryotic genomes – evidence of exon shuffling? *Trends in Genetics* **20**(9): 399-403.

Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, Abugessaisa I, Fukuda S, Hori F, Ishikawa-Kato S et al. 2015. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol* **16**(1): 22.

Long Jennifer C, Caceres Javier F. 2009. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J* **417**(1): 15-27.

Louloupi A, Ntini E, Conrad T, Ørom UAV. 2018. Transient N-6-Methyladenosine Transcriptome Sequencing Reveals a Regulatory Role of m6A in Splicing Efficiency. *Cell Rep* **23**(12): 3429-3437.

Lu X, Sachs F, Ramsay L, Jacques P, Göke J, Bourque G, Ng HH. 2014. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol* **21**(4): 423-425.

Lubas M, Andersen Peter R, Schein A, Dziembowski A, Kudla G, Jensen Torben H. 2015. The Human Nuclear Exosome Targeting Complex Is Loaded onto Newly Synthesized RNA to Direct Early Ribonucleolysis. *Cell Rep* **10**(2): 178-192.

Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. 2010. Regulation of alternative splicing by histone modifications. *Science* **327**.

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**: 251.

Lunde BM, Reichow SL, Kim M, Suh H, Leeper TC, Yang F, Mutschler H, Buratowski S, Meinhart A, Varani G. 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nature Structural &Amp; Molecular Biology* **17**: 1195.

Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M. 2008. FoxA1 Translates Epigenetic Signatures into Enhancer-Driven Lineage-Specific Transcription. *Cell* **132**(6): 958-970.

Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. 2010. The Honey Bee Epigenomes: Differential Methylation of Brain DNA in Queens and Workers. *PLoS Biol* **8**(11): e1000506.

Mahajan MC, Narlikar GJ, Boyapaty G, Kingston RE, Weissman SM. 2005. Heterogeneous nuclear ribonucleoprotein C1/C2, MeCP1, and SWI/SNF form a chromatin remodeling complex at the β-globin locus control region. *Proc Natl Acad Sci USA* **102**(42): 15012-15017.

Mahiet C, Swanson CM. 2016. Control of HIV-1 gene expression by SR proteins. *Biochem Soc Trans* **44**(5): 1417-1425.

Makałowski W, Mitchell GA, Labuda D. 1994. Alu sequences in the coding regions of mRNA: a source of protein variability. *Trends in Genetics* **10**(6): 188-193.

Malik B, Feng F. 2016. Long noncoding RNAs in prostate cancer: overview and clinical implications. *Asian Journal of Andrology* **18**(4): 568-574.

Marques AC, Ponting CP. 2009. Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. *Genome Biol* **10**.

Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, Jenuwein T. 2005. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J* **24**(4): 800-812.

Martínez-Contreras RD, Cloutier P, Shkreta L, Fisette J-F, Revil T, Chabot B. 2007. hnRNP proteins and splicing control. In *J Advances in experimental medicine and biology*, Vol 623, pp. 123-147.

Maruyama A, Mimura J, Itoh K. 2014. Non-coding RNA derived from the region adjacent to the human HO-1 E2 enhancer selectively regulates HO-1 gene induction by modulating Pol II binding. *Nucleic Acids Res* **42**(22): 13599-13614.

Mas-Ponte D, Carlevaro-Fita J, Palumbo E, Hermoso Pulido T, Guigo R, Johnson R. 2017. LncATLAS database for subcellular localization of long noncoding RNAs. *RNA* **23**(7): 1080-1087.

Masternak K, Peyraud N, Krawczyk M, Barras E, Reith W. 2003. Chromatin remodeling and extragenic transcription at the MHC class II locus control region. *Nat Immunol* **4**: 132.
Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F. 2003. A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell* **12**.
Matera AG, Wang Z. 2014. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* **15**(2): 108-121.
Mattern KA, vanGoethem REM, deJong L, vanDriel R. 1997. Major internal nuclear matrix proteins are common to different human cell types. *J Cell Biochem* **65**(1): 42-52.
Matthew Michael W, Choi M, Dreyfuss G. 1995. A nuclear export signal in hnRNP A1: A signal-mediated, temperature-dependent nuclear protein export pathway. *Cell* **83**(3): 415-422.
Mattick JS, Amaral PP, Dinger ME, Mercer TR, Mehler MF. 2009. RNA regulation of epigenetic processes. *Bioessays* **31**(1): 51-59.
McCullough AJ, Berget SM. 1997. G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. *Mol Cell Biol* **17**(8): 4562-4571.
McCullough AJ, Berget SM. 2000. An Intronic Splicing Enhancer Binds U1 snRNPs To Enhance Splicing and Select 5′ Splice Sites. *Mol Cell Biol* **20**(24): 9225-9235.
McManus CJ, Graveley BR. 2011. RNA structure and the mechanisms of alternative splicing. *Curr Opin Genet Dev* **21**(4): 373-379.
Medstrand P, van de Lagemaat LN, Dunn CA, Landry JR, Svenback D, Mager DL. 2005. Impact of transposable elements on the evolution of mammalian gene regulation. *Cytogenet Genome Res* **110**(1-4): 342-352.
Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P. 2005. A structural perspective of CTD function. *Genes Dev* **19**(12): 1401-1415.
Meischl C, de Boer M, Åhlin A, Roos D. 2000. A new exon created by intronic insertion of a rearranged LINE-1 element as the cause of chronic granulomatous disease. *Europ J Hum Genet* **8**: 697.
Melamud E, Moult J. 2009. Stochastic noise in splicing machinery. *Nucleic Acids Res* **37**(14): 4873-4886.
Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ et al. 2015. The human transcriptome across tissues and individuals. *Science* **348**(6235): 660-665.
Melé M, Mattioli K, Mallard W, Shechner DM, Gerhardinger C, Rinn JL. 2017. Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res* **27**(1): 27-37.
Melgar MF, Collins FS, Sethupathy P. 2011. Discovery of active enhancers through bidirectional expression of short transcripts. *Genome Biol* **12**(11): R113.
Melo Carlos A, Drost J, Wijchers Patrick J, van de Werken H, de Wit E, Vrielink Joachim AFO, Elkon R, Melo Sónia A, Léveillé N, Kalluri R et al. 2013. eRNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription. *Mol Cell* **49**(3): 524-535.
Mendenhall EM, Williamson KE, Reyon D, Zou JY, Ram O, Joung JK, Bernstein BE. 2013. Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol* **31**: 1133.
Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. 2008. Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci USA* **105**.
Merkin Jason J, Chen P, Alexis Maria S, Hautaniemi Sampsa K, Burge Christopher B. 2015. Origins and Impacts of New Mammalian Exons. *Cell Rep* **10**(12): 1992-2005.
Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ et al. 2010. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**: 143.
Modrek B, Lee C. 2002. A genomic view of alternative splicing. *Nat Genet* **30**: 13.
Modrek B, Lee CJ. 2003. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat Genet* **34**: 177.
Mola G, Vela E, Fernández-Figueras MT, Isamat M, Muñoz-Mármol AM. 2007. Exonization of Alu-generated Splice Variants in the Survivin Gene of Human and Non-human Primates. *J Mol Biol* **366**(4): 1055-1063.

Monsalve M, Wu Z, Adelmant G, Puigserver P, Fan M, Spiegelman BM. 2000. Direct Coupling of Transcription and mRNA Processing through the Thermogenic Coactivator PGC-1. *Mol Cell* **6**(2): 307-316.

Montell C, Fisher EF, Caruthers MH, Berk AJ. 1982. Resolving the functions of overlapping viral genes by site-specific mutagenesis at a mRNA splice site. *Nature* **295**(5848): 380-384.

Moumen A, Masterson P, O'Connor MJ, Jackson SP. 2005. hnRNP K: An HDM2 Target and Transcriptional Coactivator of p53 in Response to DNA Damage. *Cell* **123**(6): 1065-1078.

Mousavi K, Zare H, Dell'Orso S, Grontved L, Gutierrez-Cruz G, Derfoul A, Hager Gordon L, Sartorelli V. 2013. eRNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. *Mol Cell* **51**(5): 606-617.

Mueller WF, Hertel KJ. 2011. The role of SR and SR-related proteins in pre-mRNA splicing. In *RNA Binding Proteins*, Vol I (ed. Z Lorkovic), pp. 1-21. Landes Bioscience and Springer Science+Business Media, New York.

Mukherjee N, Calviello L, Hirsekorn A, de Pretis S, Pelizzola M, Ohler U. 2016. Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nat Struct Mol Biol* **24**: 86.

Müller-McNicoll M, Botti V, de Jesus Domingues AM, Brandl H, Schwich OD, Steiner MC, Curk T, Poser I, Zarnack K, Neugebauer KM. 2016. SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev* **30**(5): 553-566.

Mulligan GJ, Guo W, Wormsley S, Helfman DM. 1992. Polypyrimidine tract binding protein interacts with sequences involved in alternative splicing of beta-tropomyosin pre-mRNA. *J Biol Chem* **267**(35): 25480-25487.

Naftelberg S, Schor IE, Ast G, Kornblihtt AR. 2015. Regulation of Alternative Splicing Through Coupling with Transcription and Chromatin Structure. *Annu Rev Biochem* **84**(1): 165-198.

Nakai K, Sakamoto H. 1994. Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene* **141**(2): 171-177.

Nakayama J-i, Rice JC, Strahl BD, Allis CD, Grewal SIS. 2001. Role of Histone H3 Lysine 9 Methylation in Epigenetic Control of Heterochromatin Assembly. *Science* **292**(5514): 110-113.

Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grutzner F, Kaessmann H. 2014. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**(7485): 635-640.

Neish AS, Anderson SF, Schlegel BP, Wei W, Parvin JD. 1998. Factors associated with the mammalian RNA polymerase II holoenzyme. *Nucleic Acids Res* **26**(3): 847-853.

Nekrutenko A, Li W-H. 2001. Transposable elements are found in a large number of human protein-coding genes. *Trends in Genetics* **17**(11): 619-621.

Nguyen HD, Yoshihama M, Kenmochi N. 2005. New Maximum Likelihood Estimators for Eukaryotic Intron Evolution. *PLoS Comp Biol* **1**(7): e79.

Nieto Moreno N, Giono LE, Cambindo Botto AE, Muñoz MJ, Kornblihtt AR. 2015. Chromatin, DNA structure and alternative splicing. *FEBS Letters* **589**(22): 3370-3378.

Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**.

Niranjanakumari S, Lasda E, Brazas R, Garcia-Blanco MA. 2002. Reversible cross-linking combined with immunoprecipitation to study RNA–protein interactions in vivo. *Methods* **26**(2): 182-190.

Nitsche A, Doose G, Tafer H, Robinson M, Saha NR, Gerdol M, Canapa A, Hoffmann S, Amemiya CT, Stadler PF. 2014. Atypical RNAs in the coelacanth transcriptome. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **322**(6): 342-351.

Noble CG, Hollingworth D, Martin SR, Ennis-Adeniran V, Smerdon SJ, Kelly G, Taylor IA, Ramos A. 2005. Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nature Structural &Amp; Molecular Biology* **12**: 144.

Nogués G, Kadener S, Cramer P, Bentley D, Kornblihtt AR. 2002. Transcriptional Activators Differ in Their Abilities to Control Alternative Splicing. *J Biol Chem* **277**(45): 43110-43114.

Nogués G, Muñoz MJ, Kornblihtt AR. 2003. Influence of Polymerase II Processivity on Alternative Splicing Depends on Splice Site Strength. *J Biol Chem* **278**(52): 52166-52171.
Nojima T, Gomes T, Grosso Ana Rita F, Kimura H, Dye Michael J, Dhir S, Carmo-Fonseca M, Proudfoot Nicholas J. 2015. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**(3): 526-540.
Nudler E, Mironov AS. 2004. The riboswitch control of bacterial metabolism. *Trends in Biochemical Sciences* **29**(1): 11-17.
Olson S, Blanchette M, Park J, Savva Y, Yeo GW, Yeakley JM, Rio DC, Graveley BR. 2007. A regulator of Dscam mutually exclusive splicing fidelity. *Nat Struct Mol Biol* **14**(12): 1134-1140.
Onodera CS, Underwood JG, Katzman S, Jacobs F, Greenberg D, Salama SR, Haussler D. 2012. Gene Isoform Specificity through Enhancer-Associated Antisense Transcription. *PLoS ONE* **7**(8): e43511.
Ørom UA, Derrien T, Beringer M, Gumireddy K, Gardini A, Bussotti G, Lai F, Zytnicki M, Notredame C, Huang Q et al. 2010. Long Noncoding RNAs with Enhancer-like Function in Human Cells. *Cell* **143**(1): 46-58.
Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S, Curina A, Prosperini E, Ghisletti S, Natoli G. 2013. Latent Enhancers Activated by Stimulation in Differentiated Cells. *Cell* **152**(1): 157-171.
Ounzain S, Pezzuto I, Micheletti R, Burdet F, Sheta R, Nemir M, Gonzales C, Sarre A, Alexanian M, Blow MJ et al. 2014. Functional importance of cardiac enhancer-associated noncoding RNAs in heart development and disease. *Journal of Molecular and Cellular Cardiology* **76**(0): 55-70.
Padgett RA. 2012. New connections between splicing and human disease. *Trends in Genetics* **28**(4): 147-154.
Pagani F, Stuani C, Zuccato E, Kornblihtt AR, Baralle FE. 2003. Promoter Architecture Modulates CFTR Exon 9 Skipping. *J Biol Chem* **278**(3): 1511-1517.
Pal C, Papp B, Hurst LD. 2001. Highly expressed genes in yeast evolve slowly. *Genetics* **158**(2): 927-931.
Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**: 1413.
Pandya-Jones A, Bhatt DM, Lin C-H, Tong A-J, Smale ST, Black DL. 2013. Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA* **19**(6): 811-827.
Pandya-Jones A, Black DL. 2009. Co-transcriptional splicing of constitutive and alternative exons. *RNA* **15**(10): 1896-1908.
Pang KC, Frith MC, Mattick JS. 2006. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in Genetics* **22**(1): 1-5.
Paralkar Vikram R, Taborda Cristian C, Huang P, Yao Y, Kossenkov Andrew V, Prasad R, Luan J, Davies James OJ, Hughes Jim R, Hardison Ross C et al. 2016. Unlinking an lncRNA from Its Associated cis Element. *Mol Cell* **62**(1): 104-110.
Parmley JL, Chamary JV, Hurst LD. 2006. Evidence for Purifying Selection Against Synonymous Mutations in Mammalian Exonic Splicing Enhancers. *Mol Biol Evol* **23**(2): 301-309.
Parmley JL, Urrutia AO, Potrzebowski L, Kaessmann H, Hurst LD. 2007. Splicing and the evolution of proteins in mammals. *PLoS Biol* **5**.
Patel AA, Steitz JA. 2003. Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol* **4**(12): 960-970.
Patrushev LI, Kovalenko TF. 2014. Functions of noncoding sequences in mammalian genomes. *Biochem (Mosc)* **79**(13): 1442-1469.
Patthy L. 1987. Intron-dependent evolution: Preferred types of exons and introns. *FEBS Letters* **214**(1): 1-7.
Patthy L. 1996. Exon shuffling and other ways of module exchange. *Matrix Biology* **15**(5): 301-310.

Patthy L. 1999. Genome evolution and the evolution of exon-shuffling — a review. *Gene* **238**(1): 103-114.

Pauli A, Norris ML, Valen E, Chew G-L, Gagnon JA, Zimmerman S, Mitchell A, Ma J, Dubrulle J, Reyon D et al. 2014. Toddler: An Embryonic Signal That Promotes Cell Movement via Apelin Receptors. *Science* **343**(6172).

Pauli A, Valen E, Schier AF. 2015. Identifying (non-)coding RNAs and small peptides: Challenges and opportunities. *Bioessays* **37**(1): 103-112.

Paz I, Akerman M, Dror I, Kosti I, Mandel-Gutfreund Y. 2010. SFmap: a web server for motif analysis and prediction of splicing factor binding sites. *Nucleic Acids Res* **38**(Web Server issue): W281-285.

Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB. 2004. Retrotransposons Regulate Host Genes in Mouse Oocytes and Preimplantation Embryos. *Dev Cell* **7**(4): 597-606.

Pefanis E, Wang J, Rothschild G, Lim J, Kazadi D, Sun J, Federation A, Chao J, Elliott O, Liu Z-P et al. 2015. RNA Exosome-Regulated Long Non-Coding RNA Transcription Controls Super-Enhancer Activity. *Cell* **161**(4): 774-789.

Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H, Imbert J, Andrau JC, Ferrier P, Spicuglia S. 2011. H3K4 tri-methylation provides an epigenetic signature of active enhancers. *EMBO J* **30**(20): 4198-4210.

Peng T, Li Y. 2009. Tandem exon duplication tends to propagate rather than to create de novo alternative splicing. *Biochem Biophys Res Commun* **383**(2): 163-166.

Perales R, Bentley D. 2009. "Cotranscriptionality": The Transcription Elongation Complex as a Nexus for Nuclear Transactions. *Mol Cell* **36**(2): 178-191.

Piñol-Roma S, Dreyfuss G. 1992. Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* **355**: 730.

Piñol-Roma S, Choi YD, Matunis MJ, Dreyfuss G. 1988. Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes Dev* **2**(2): 215-227.

Plass M, Agirre E, Reyes D, Camara F, Eyras E. 2008. Co-evolution of the branch site and SR proteins in eukaryotes. *Trends in Genetics* **24**(12): 590-594.

Pnueli L, Rudnizky S, Yosefzon Y, Melamed P. 2015. RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin α-subunit gene. *Proceedings of the National Academy of Sciences* **112**(14): 4369-4374.

Ponjavic J, Ponting CP, Lunter G. 2007. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res* **17**.

Pradeepa MM, Sutherland HG, Ule J, Grimes GR, Bickmore WA. 2012. Psip1/Ledgf p52 Binds Methylated Histone H3K36 and Splicing Factors and Contributes to the Regulation of Alternative Splicing. *Plos Genetics* **8**(5): e1002717.

Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH. 2008. RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. *Science* **322**(5909): 1851-1854.

Pulakanti K, Pinello L, Stelloh C, Blinka S, Allred J, Milanovich S, Kiblawi S, Peterson J, Wang A, Yuan G-C et al. 2013. Enhancer transcribed RNAs arise from hypomethylated, Tet-occupied genomic regions. *Epigenetics* **8**(12): 1303-1320.

Quinn JJ, Chang HY. 2016. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet* **17**(1): 47-62.

Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA. 2010. c-Myc Regulates Transcriptional Pause Release. *Cell* **141**(3): 432-445.

Ransohoff JD, Wei Y, Khavari PA. 2017. The functions and unique features of long intergenic non-coding RNA. *Nat Rev Mol Cell Biol* **19**: 143.

Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun Z-W, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD et al. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**: 593.

Resch A, Xing Y, Alekseyenko A, Modrek B, Lee C. 2004. Evidence for a subpopulation of conserved alternative splicing events under selection pressure for protein reading frame preservation. *Nucleic Acids Res* **32**(4): 1261-1269.

Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E et al. 2007. Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* **129**(7): 1311-1323.

Rivera Chloe M, Ren B. 2013. Mapping Human Epigenomes. *Cell* **155**(1): 39-55.

Robberson BL, Cote GJ, Berget SM. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* **10**(1): 84-94.

Roberts GC, Gooding C, Mak HY, Proudfoot NJ, Smith CW. 1998. Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Res* **26**.

Roca X, Akerman M, Gaus H, Berdeja A, Bennett CF, Krainer AR. 2012. Widespread recognition of 5′ splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. *Genes Dev* **26**(10): 1098-1109.

Roca X, Krainer AR, Eperon IC. 2013. Pick one, but be quick: 5′ splice sites and the problems of too many choices. *Genes Dev* **27**(2): 129-144.

Roca X, Sachidanandam R, Krainer AR. 2003. Intrinsic differences between authentic and cryptic 5' splice sites. *Nucleic Acids Res* **31**(21): 6321-6333.

Rodriguez J, Menet Jerome S, Rosbash M. 2012. Nascent-Seq Indicates Widespread Cotranscriptional RNA Editing in *Drosophila*. *Mol Cell* **47**(1): 27-37.

Rogan DF, Cousins DJ, Santangelo S, Ioannou PA, Antoniou M, Lee TH, Staynov DZ. 2004. Analysis of intergenic transcription in the human IL-4/IL-13 gene cluster. *Proc Natl Acad Sci USA* **101**(8): 2446-2451.

Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV. 2003. Remarkable Interkingdom Conservation of Intron Positions and Massive, Lineage-Specific Intron Loss and Gain in Eukaryotic Evolution. *Curr Biol* **13**(17): 1512-1517.

Rocha EPC. 2008. The Organization of the Bacterial Genome. *Annu Rev Genet* **42**(1): 211-233.

Romito A, Rougeulle C. 2011. Origin and evolution of the long non-coding genes in the X-inactivation center. *Biochimie* **93**(11): 1935-1942.

Rose D, Hiller M, Schutt K, Hackermüller J, Backofen R, Stadler PF. 2011. Computational discovery of human coding and non-coding transcripts with conserved splice sites. *Bioinformatics* **27**(14): 1894-1900.

Rot G, Wang Z, Huppertz I, Modic M, Lenče T, Hallegger M, Haberman N, Curk T, von Mering C, Ule J. 2017. High-Resolution RNA Maps Suggest Common Principles of Splicing and Polyadenylation Regulation by TDP-43. *Cell Rep* **19**(5): 1056-1067.

Roy SW, Gilbert W. 2005. Rates of intron loss and gain: Implications for early eukaryotic evolution. *Proc Natl Acad Sci USA* **102**(16): 5773-5778.

Roy SW, Irimia M. 2009. Splicing in the eukaryotic ancestor: form, function and dysfunction. *Trends Ecol Evol* **24**(8): 447-455.

Ruf S, Symmons O, Uslu VV, Dolle D, Hot C, Ettwiller L, Spitz F. 2011. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat Genet* **43**: 379.

Ruiz-Orera J, Messeguer X, Subirana JA, Alba MM. 2014. Long non-coding RNAs as a source of new peptides. In *Elife*, Vol 3, p. e03523.

Saint-Andre V, Batsche E, Rachez C, Muchardt C. 2011. Histone H3 lysine 9 trimethylation and HP1gamma favor inclusion of alternative exons. *Nat Struct Mol Biol* **18**.

Sakharkar MK, Perumal BS, Sakharkar KR, Kangueane P. 2005. An Analysis on Gene Architecture in Human and Mouse Genomes. *In Silico Biology* **5**(4): 347-365.

Salton M, Voss TC, Misteli T. 2014. Identification by high-throughput imaging of the histone methyltransferase EHMT2 as an epigenetic regulator of VEGFA alternative splicing. *Nucleic Acids Res* **42**(22): 13662-13673.

Sanford JR, Coutinho P, Hackett JA, Wang X, Ranahan W, Caceres JF. 2008. Identification of Nuclear and Cytoplasmic mRNA Targets for the Shuttling Protein SF2/ASF. *PLoS ONE* **3**(10): e3369.

Sanjana NE, Shalem O, Zhang F. 2014. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**: 783.

Sanyal A, Lajoie BR, Jain G, Dekker J. 2012. The long-range interaction landscape of gene promoters. *Nature* **489**: 109.

Seidl CI, Stricker SH, Barlow DP. 2006. The imprinted *Air* ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *EMBO J* **25**(15): 3565-3575.

Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. 2008. Divergent Transcription from Active Promoters. *Science* **322**(5909): 1849-1851.

Sela N, Mersch B, Gal-Mark N, Lev-Maor G, Hotz-Wagenblatt A, Ast G. 2007. Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. *Genome Biol* **8**(6): R127.

Séraphin B, Kretzner L, Rosbash M. 1988. A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *EMBO J* **7**(8): 2533-2538.

Sharifi-Zarchi A, Gerovska D, Adachi K, Totonchi M, Pezeshk H, Taft RJ, Schöler HR, Chitsaz H, Sadeghi M, Baharvand H et al. 2017. DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. *Bmc Genomics* **18**(1): 964.

Shen H, Green MR. 2006. RS domains contact splicing signals and promote splicing by a common mechanism in yeast through humans. *Genes Dev* **20**(13): 1755-1765.

Shen H, Kan JLC, Green MR. 2004. Arginine-Serine-Rich Domains Bound at Splicing Enhancers Contact the Branchpoint to Promote Prespliceosome Assembly. *Mol Cell* **13**(3): 367-376.

Sheth N, Roca X, Hastings ML, Roeder T, Krainer AR, Sachidanandam R. 2006. Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res* **34**(14): 3955-3967.

Shlyueva D, Stampfel G, Stark A. 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews Genetics* **15**: 272.

Schaub MC, Lopez SR, Caputi M. 2007. Members of the Heterogeneous Nuclear Ribonucleoprotein H Family Activate Splicing of an HIV-1 Splicing Substrate by Promoting Formation of ATP-dependent Spliceosomal Complexes. *J Biol Chem* **282**(18): 13617-13626.

Schaukowitch K, Joo J-Y, Liu X, Watts Jonathan K, Martinez C, Kim T-K. 2014. Enhancer RNA Facilitates NELF Release from Immediate Early Genes. *Mol Cell* **56**(1): 29-42.

Schlackow M, Nojima T, Gomes T, Dhir A, Carmo-Fonseca M, Proudfoot NJ. 2017. Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. *Mol Cell* **65**(1): 25-38.

Schlesinger F, Smith AD, Gingeras TR, Hannon GJ, Hodges E. 2013. De novo DNA demethylation and noncoding transcription define active intergenic regulatory elements. *Genome Res* **23**(10): 1601-1614.

Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL. 2000. Drosophila Dscam Is an Axon Guidance Receptor Exhibiting Extraordinary Molecular Diversity. *Cell* **101**(6): 671-684.

Schor IE, Fiszbein A, Petrillo E, Kornblihtt AR. 2013. Intragenic epigenetic changes modulate NCAM alternative splicing in neuronal differentiation. *EMBO J* **32**(16): 2264-2274.

Schor IE, Rascovan N, Pelisch F, Alló M, Kornblihtt AR. 2009. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proceedings of the National Academy of Sciences* **106**(11): 4325-4330.

Schüler A, Ghanbarian AT, Hurst LD. 2014. Purifying Selection on Splice-Related Motifs, Not Expression Level nor RNA Folding, Explains Nearly All Constraint on Human lincRNAs. *Mol Biol Evol* **31**(12): 3164-3183.

Schwahn U, Lenzner S, Dong J, Feil S, Hinzmann B, van Duijnhoven G, Kirschner R, Hemberger M, Bergen AAB, Rosenberg T et al. 1998. Positional cloning of the gene for X-linked retinitis pigmentosa 2. *Nat Genet* **19**: 327.

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

Schwartz S, Silva J, Burstein D, Pupko T, Eyras E, Ast G. 2008. Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res* **18**(1): 88-103.

Sigova AA, Abraham BJ, Ji X, Molinie B, Hannett NM, Guo YE, Jangi M, Giallourakis CC, Sharp PA, Young RA. 2015. Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**(6263): 978-981.

Sigova AA, Mullen AC, Molinie B, Gupta S, Orlando DA, Guenther MG, Almada AE, Lin C, Sharp PA, Giallourakis CC et al. 2013. Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proceedings of the National Academy of Sciences* **110**(8): 2876-2881.

Sims RJ, III, Millhouse S, Chen C-F, Lewis BA, Erdjument-Bromage H, Tempst P, Manley JL, Reinberg D. 2007. Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing. *Mol Cell* **28**(4): 665-676.

Sironi M, Menozzi G, Riva L, Cagliani R, Comi GP, Bresolin N, Giorda R, Pozzoli U. 2004. Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic Acids Res* **32**(5): 1783-1791.

Smith E, Lin C, Shilatifard A. 2011. The super elongation complex (SEC) and MLL in development and disease. *Genes Dev* **25**(7): 661-672.

Smith Jenna E, Alvarez-Dominguez Juan R, Kline N, Huynh Nathan J, Geisler S, Hu W, Coller J, Baker Kristian E. 2014. Translation of Small Open Reading Frames within Unannotated RNA Transcripts in Saccharomyces cerevisiae. *Cell Rep* **7**(6): 1858-1866.

Smith MA, Gesell T, Stadler PF, Mattick JS. 2013. Widespread purifying selection on RNA structure in mammals. *Nucleic Acids Res* **41**(17): 8220-8236.

Solier S, Barb J, Zeeberg BR, Varma S, Ryan MC, Kohn KW, Weinstein JN, Munson PJ, Pommier Y. 2010. Genome-wide Analysis of Novel Splice Variants Induced by Topoisomerase I Poisoning Shows Preferential Occurrence in Genes Encoding Splicing Factors. *Cancer Res* **70**(20): 8055-8065.

Sone M, Hayashi T, Tarui H, Agata K, Takeichi M, Nakagawa S. 2007. The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. *J Cell Sci* **120**.

Sorek R. 2007. The birth of new exons: Mechanisms and evolutionary consequences. *RNA* **13**(10): 1603-1608.

Sorek R, Ast G. 2003. Intronic Sequences Flanking Alternatively Spliced Exons Are Conserved Between Human and Mouse. *Genome Res* **13**(7): 1631-1637.

Sorek R, Ast G, Graur D. 2002. Alu-Containing Exons are Alternatively Spliced. *Genome Res* **12**(7): 1060-1067.

Sorek R, Lev-Maor G, Reznik M, Dagan T, Belinky F, Graur D, Ast G. 2004a. Minimal Conditions for Exonization of Intronic Sequences: 5′ Splice Site Formation in Alu Exons. *Mol Cell* **14**(2): 221-231.

Sorek R, Shamir R, Ast G. 2004b. How prevalent is functional alternative splicing in the human genome? *Trends in Genetics* **20**(2): 68-71.

Soufi A, Donahue G, Zaret Kenneth S. 2012. Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome. *Cell* **151**(5): 994-1004.

Spies N, Nielsen CB, Padgett RA, Burge CB. 2009. Biased chromatin signatures around polyadenylation sites and exons. *Mol Cell* **36**.

Spiluttini B, Gu B, Belagal P, Smirnova AS, Nguyen VT, Hébert C, Schmidt U, Bertrand E, Darzacq X, Bensaude O. 2010. Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. *J Cell Sci* **123**(12): 2085-2093.

Spitz F, Furlong EEM. 2012. Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics* **13**: 613.

Srivastava R, Ahn SH. 2015. Modifications of RNA polymerase II CTD: Connections to the histone code and cellular function. *Biotechnol Adv* **33**(6, Part 1): 856-872.

Stamm S, Zhang MQ, Marr TG, Helfman DM. 1994. A sequence compilation and comparison of exons that are alternatively spliced in neurons. *Nucleic Acids Res* **22**(9): 1515-1526.

Sterner DA, Carlo T, Berget SM. 1996. Architectural limits on split genes. *Proc Natl Acad Sci USA* **93**(26): 15081-15085.

Struhl K. 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* **14**(2): 103-105.

Sun X-J, Wei J, Wu X-Y, Hu M, Wang L, Wang H-H, Zhang Q-H, Chen S-J, Huang Q-H, Chen Z. 2005. Identification and Characterization of a Novel Human Histone H3 Lysine 36-specific Methyltransferase. *J Biol Chem* **280**(42): 35261-35271.

Tacke R, Chen Y, Manley JL. 1997. Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: Creation of an SRp40-specific splicing enhancer. *Proc Natl Acad Sci USA* **94**(4): 1148-1153.

Tacke R, Manley JL. 1995. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J* **14**(14): 3540-3551.

Taft RJ, Pheasant M, Mattick JS. 2007. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29**(3): 288-299.

Tachibana M, Sugimoto K, Fukushima T, Shinkai Y. 2001. SET Domain-containing Protein, G9a, Is a Novel Lysine-preferring Mammalian Histone Methyltransferase with Hyperactivity and Specific Selectivity to Lysines 9 and 27 of Histone H3. *J Biol Chem* **276**(27): 25309-25317.

Talbert PB, Henikoff S. 2010. Histone variants — ancient wrap artists of the epigenome. *Nature Reviews Molecular Cell Biology* **11**: 264.

Talerico M, Berget SM. 1994. Intron definition in splicing of small Drosophila introns. *Mol Cell Biol* **14**(5): 3434-3445.

Tan JY, Biasini A, Young RS, Marques A. 2018. An unexpected contribution of lincRNA splicing to enhancer function. *bioRxiv*.

Tanenbaum Marvin E, Gilbert Luke A, Qi Lei S, Weissman Jonathan S, Vale Ronald D. 2014. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell* **159**(3): 635-646.

Tange TØ, Damgaard CK, Guth S, Valcárcel J, Kjems J. 2001. The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J* **20**(20): 5748-5758.

Tani H, Mizutani R, Salam KA, Tano K, Ijiri K, Wakamatsu A, Isogai T, Suzuki Y, Akimitsu N. 2012. Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res* **22**(5): 947-956.

Tappino B, Regis S, Corsolini F, Filocamo M. 2008. An Alu insertion in compound heterozygosity with a microduplication in GNPTAB gene underlies Mucolipidosis II. *Mol Genet Metab* **93**(2): 129-133.

Tazi J, Kornstädt U, Rossi F, Jeanteur P, Cathala G, Brunel C, Lührmann R. 1993. Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing. *Nature* **363**: 283.

Thompson Peter J, Macfarlan Todd S, Lorincz Matthew C. 2016. Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. *Mol Cell* **62**(5): 766-776.

Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, Curado J, Snyder M, Gingeras TR, 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 Res* **22**(9): 1616-1625.

Tilgner H, Nikolaou C, Althammer S, Sammeth M, Beato M, Valcarcel J. 2009. Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol* **16**.

Tolstorukov Michael Y, Goldman Joseph A, Gilbert C, Ogryzko V, Kingston Robert E, Park Peter J. 2012. Histone Variant H2A.Bbd Is Associated with Active Transcription and mRNA Processing in Human Cells. *Mol Cell* **47**(4): 596-607.

Tuan D, Kong S, Hu K. 1992. Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proceedings of the National Academy of Sciences* **89**(23): 11219-11223.

Tucker BJ, Breaker RR. 2005. Riboswitches as versatile gene control elements. *Curr Opin Struct Biol* **15**(3): 342-348.

Turunen JJ, Niemelä EH, Verma B, Frilander MJ. 2013. The significant other: splicing by the minor spliceosome. *Wiley interdisciplinary reviews RNA* **4**(1): 61-76.

Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB. 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**(7119): 580-586.

Ulitsky I. 2016. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet* **17**(10): 601-614.

Ulitsky I, Bartel David P. 2013. lincRNAs: Genomics, Evolution, and Mechanisms. *Cell* **154**(1): 26-46.

Ulitsky I, Shkumatava A, Jan Calvin H, Sive H, Bartel David P. 2011. Conserved Function of lincRNAs in Vertebrate Embryonic Development despite Rapid Sequence Evolution. *Cell* **147**(7): 1537-1550.

Van Nostrand EL, Freese P, Pratt GA, Wang X, Wei X, Blue SM, Dominguez D, Cody NAL, Olson S, Sundararaman B et al. 2017. A Large-Scale Binding and Functional Map of Human RNA Binding Proteins. *bioRxiv*.

van Rijk A, Bloemendal H. 2003. Molecular mechanisms of exon shuffling: illegitimate recombination. *Genetica* **118**(2-3): 245-249.

Veloso A, Kirkconnell KS, Magnuson B, Biewen B, Paulsen MT, Wilson TE, Ljungman M. 2014. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res* **24**(6): 896-905.

Villar D, Berthelot C, Aldridge S, Rayner Tim F, Lukk M, Pignatelli M, Park Thomas J, Deaville R, Erichsen Jonathan T, Jasinska Anna J et al. 2015. Enhancer Evolution across 20 Mammalian Species. *Cell* **160**(3): 554-566.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**: 854.

Volek M. 2018. Influence of transcription regulatory elements on pre-mRNA splicing. In *Faculty of Natural Sciences, Department of Genetics and Microbiology*, p. 78. Charles Univerzity in Prage, Prague.

Vorechovsky I. 2010. Transposable elements in disease-associated cryptic exons. *Hum Genet* **127**(2): 135-154.

Vučićević D, Corradin O, Ntini E, Scacheri PC, Ørom UA. 2015. Long ncRNA expression associates with tissue-specific enhancers. *Cell Cycle* **14**(2): 253-260.

Wagner EJ, Carpenter PB. 2012. Understanding the language of Lys36 methylation at histone H3. *Nature Reviews Molecular Cell Biology* **13**: 115.

Wagner EJ, Garcia-Blanco MA. 2001. Polypyrimidine Tract Binding Protein Antagonizes Exon Definition. *Mol Cell Biol* **21**(10): 3281-3288.

Wang E, Dimova N, Cambi F. 2007. PLP/DM20 ratio is regulated by hnRNPH and F and a novel G-rich enhancer in oligodendrocytes. *Nucleic Acids Res* **35**(12): 4164-4178.

Wang E, Mueller WF, Hertel KJ, Cambi F. 2011a. G Run-mediated Recognition of Proteolipid Protein and DM20 5 ' Splice Sites by U1 Small Nuclear RNA Is Regulated by Context and Proximity to the Splice Site. *J Biol Chem* **286**(6): 4059-4071.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**.

Wang J, Zhang J, Zheng H, Li J, Liu D, Li H, Samudrala R, Yu J, Wong GK-S. 2004a. Neutral evolution of 'non-coding' complementary DNAs. *Nature* **431**.

Wang Kevin C, Chang Howard Y. 2011. Molecular Mechanisms of Long Noncoding RNAs. *Mol Cell* **43**(6): 904-914.

Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA et al. 2011b. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**(7341): 120-124.

Wang W, Kirkness EF. 2005. Short interspersed elements (SINEs) are a major source of canine genomic diversity. *Genome Res* **15**(12): 1798-1808.

Wang W, Zheng H, Yang S, Yu H, Li J, Jiang H, Su J, Yang L, Zhang J, McDermott J et al. 2005. Origin and evolution of new exons in rodents. *Genome Res* **15**(9): 1258-1264.

Wang Y, Li X, Hu H. 2014. H3K4me2 reliably defines transcription factor binding regions in different cells. *Genomics* **103**(2): 222-228.

Wang Y, Ma M, Xiao X, Wang Z. 2012. Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules. *Nat Struct Mol Biol* **19**(10): 1044-1052.
Wang Z, Burge CB. 2008. Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA* **14**(5): 802-813.
Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. 2004b. Systematic Identification and Analysis of Exonic Splicing Silencers. *Cell* **119**(6): 831-845.
Wang Z, Zang C, Cui K, Schones DE, Barski A, Peng W, Zhao K. 2009. Genome-wide Mapping of HATs and HDACs Reveals Distinct Functions in Active and Inactive Genes. *Cell* **138**(5): 1019-1031.
Washietl S, Kellis M, Garber M. 2014. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res* **24**(4): 616-628.
Wei C-C, Zhang S-L, Chen Y-W, Guo D-F, Ingelfinger JR, Bomsztyk K, Chan JSD. 2006. Heterogeneous Nuclear Ribonucleoprotein K Modulates Angiotensinogen Gene Expression in Kidney Cells. *J Biol Chem* **281**(35): 25344-25355.
Whetstine JR, Nottke A, Lan F, Huarte M, Smolikov S, Chen Z, Spooner E, Li E, Zhang G, Colaiacovo M et al. 2006. Reversal of Histone Lysine Trimethylation by the JMJD2 Family of Histone Demethylases. *Cell* **125**(3): 467-481.
Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.
Wieringa B, Meyer F, Reiser J, Weissmann C. 1983. Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit β-globin gene. *Nature* **301**: 38.
Will CL, Lührmann R. 2011. Spliceosome Structure and Function. *Cold Spring Harb Perspect Biol* **3**(7): 3:a003707.
Wilson CJ, Chao DM, Imbalzano AN, Schnitzler GR, Kingston RE, Young RA. 1996. RNA Polymerase II Holoenzyme Contains SWI/SNF Regulators Involved in Chromatin Remodeling. *Cell* **84**(2): 235-244.
Wilusz JE, Sunwoo H, Spector DL. 2009. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* **23**(13): 1494-1504.
Wong MS, Kinney JB, Krainer AR. 2018. Quantitative Activity Profile and Context Dependence of All Human 5' Splice Sites. *Mol Cell* **71**(6): 1012-1026.e1013.
Wong MS, Wright WE, Shay JW. 2014. Alternative splicing regulation of telomerase: a new paradigm? *Trends in Genetics* **30**(10): 430-438.
Wu H, Min J, Lunin VV, Antoshenko T, Dombrovski L, Zeng H, Allali-Hassani A, Campagna-Slater V, Vedadi M, Arrowsmith CH et al. 2010. Structural Biology of Human H3K9 Methyltransferases. *PLoS ONE* **5**(1): e8570.
Wu H, Nord AS, Akiyama JA, Shoukry M, Afzal V, Rubin EM, Pennacchio LA, Visel A. 2014. Tissue-Specific RNA Expression Marks Distant-Acting Developmental Enhancers. *Plos Genetics* **10**(9): e1004610.
Wu JY, Maniatis T. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**(6): 1061-1070.
Wu X, Sharp PA. 2013. Divergent Transcription: A Driving Force for New Gene Origination? *Cell* **155**(5): 990-996.
Wyers F, Rougemaille M, Badis G, Rousselle J-C, Dufour M-E, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B et al. 2005. Cryptic Pol II Transcripts Are Degraded by a Nuclear Quality Control Pathway Involving a New Poly(A) Polymerase. *Cell* **121**(5): 725-737.
Xiao SH, Manley JL. 1997. Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev* **11**(3): 334-344.
Xiao T, Hall H, Kizer KO, Shibata Y, Hall MC, Borchers CH, Strahl BD. 2003. Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes Dev* **17**(5): 654-663.
Xiao X, Wang Z, Jang M, Burge CB. 2007. Coevolutionary networks of splicing cis-regulatory elements. *Proc Natl Acad Sci USA* **104**(47): 18583-18588.
Xiao X, Wang Z, Jang M, Nutiu R, Wang ET, Burge CB. 2009. Splice site strength–dependent activity and genetic buffering by poly-G runs. *Nat Struct Mol Biol* **16**: 1094.

Xue Y, Zhou Y, Wu T, Zhu T, Ji X, Kwon Y-S, Zhang C, Yeo G, Black DL, Sun H et al. 2009. Genome-wide Analysis of PTB-RNA Interactions Reveals a Strategy Used by the General Splicing Repressor to Modulate Exon Inclusion or Skipping. *Mol Cell* **36**(6): 996-1006.

Yang L, Lin C, Liu W, Zhang J, Ohgi Kenneth A, Grinstein Jonathan D, Dorrestein Pieter C, Rosenfeld Michael G. 2011. ncRNA- and Pc2 Methylation-Dependent Gene Relocation between Nuclear Structures Mediates Gene Activation Programs. *Cell* **147**(4): 773-788.

Yin Y, Yan P, Lu J, Song G, Zhu Y, Li Z, Zhao Y, Shen B, Huang X, Zhu H et al. 2015. Opposing Roles for the lncRNA Haunt and Its Genomic Locus in Regulating HOXA Gene Activation during Embryonic Stem Cell Differentiation. *Cell Stem Cell* **16**(5): 504-516.

Yoshida K, Nakamura A, Yazaki M, Ikeda S-i, Takeda Si. 1998. Insertional mutation by transposable element, L1, in the DMD gene results in X-linked dilated cardiomyopathy. *Hum Mol Genet* **7**(7): 1129-1132.

Yuan W, Xie J, Long C, Erdjument-Bromage H, Ding X, Zheng Y, Tempst P, Chen S, Zhu B, Reinberg D. 2009. Heterogeneous Nuclear Ribonucleoprotein L Is a Subunit of Human KMT3a/Set2 Complex Required for H3 Lys-36 Trimethylation Activity in Vivo. *J Biol Chem* **284**(23): 15701-15707.

Zarnack K, König J, Tajnik M, Martincorena I, Eustermann S, Stévant I, Reyes A, Anders S, Luscombe Nicholas M, Ule J. 2013. Direct Competition between hnRNP C and U2AF65 Protects the Transcriptome from the Exonization of Alu Elements. *Cell* **152**(3): 453-466.

Zerbino DR, Wilder SP, Johnson N, Juettemann T, Flicek PR. 2015. The Ensembl Regulatory Build. *Genome Biol* **16**(1): 56.

Zhang MQ. 1998. Statistical features of human exons and their flanking regions. *Hum Mol Genet* **7**(5): 919-932.

Zhang XH-F, Leslie CS, Chasin LA. 2005. Dichotomous splicing signals in exon flanks. *Genome Res* **15**(6): 768-779.

Zhou H-L, Hinman MN, Barron VA, Geng C, Zhou G, Luo G, Siegel RE, Lou H. 2011. Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner. *Proceedings of the National Academy of Sciences* **108**(36): E627-E635.

Zhu J, Mayeda A, Krainer AR. 2001. Exon Identity Established through Differential Antagonism between Exonic Splicing Silencer-Bound hnRNP A1 and Enhancer-Bound SR Proteins. *Mol Cell* **8**(6): 1351-1361.

Zhu Y, Sun L, Chen Z, Whitaker JW, Wang T, Wang W. 2013. Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Res* **41**(22): 10032-10043.

Zhuang Y, Ma F, Li-Ling J, Xu X, Li Y. 2003. Comparative Analysis of Amino Acid Usage and Protein Length Distribution Between Alternatively and Non-alternatively Spliced Genes Across Six Eukaryotic Genomes. *Mol Biol Evol* **20**(12): 1978-1985.