TESI DOCTORAL UPF 2013

Epigenetics in alternative splicing: links between chromatin structure, transcription and non-coding RNA mediated regulation

Eneritz Agirre Ortiz de Guzmán

Department of Experimental and Health Sciences

Regulatory Genomics Group, Research Programme on Biomedical Informatics (GRIB), IMIM-UPF

Dr. Eduardo Eyras

Director

Universitat Pompeu Fabra

Institució Catalana de Recerca i Estudis Avançats (ICREA)

Barcelona, 2013

Para aita.

"On the other side of the screen, it all looks so easy."
Kevin Flynn, 1982.

Acknowledgments

Hace ya tiempo que empecé la tésis y muchísimas cosas han cambiado desde entonces, para bien y para mal. A pesar de quejarme de vez en cuando (muchas veces) estoy segura de que volvería a empezarla de nuevo si echase atrás. Cambiaría algunas cosas, pero probablemente el resultado no sería muy diferente. Ante todo creo que he aprendido mucho a nivel de trabajo, porque al principio casi no tenía ni idea de nada aparte de mi frikismo innato. Además de mis orígenes como bióloga de bata y monte... Todos estos aos me han hecho descubrir muchas cosas y a mucha gente, que ahora son parte de mi. Lo adelanto, si me olvido de alguien, soys muchos y a todos los que me habeis ayudado alguna vez, gracias.

En primer lugar quiero agradecer a Eduardo Eyras el haberme dado la oportunidad de empezar esta tésis, confió en mí y yo he intentado hacerlo lo mejor que he podido. Durante todo este largo y duro proceso, porque ha sido un poco complicado, me ha acompaado mucha gente. Muchos han estado en etapas diferentes pero todavía queda alguno que ha aguantado hasta el final. Por eso, primero quiero agradecer mis comienzos a Andre y Mireya, cuando sólo eramos tres en el grupo. Nada mas llegar nos fuimos de conferencia a Polonia y todavía me acuerdo de como Mireya me iba comentado quienes eran todos los PI, super cracks de splicing. Como Andre y yo aprovechabamos al máximo todos los pica-picas de la conferencia, esas noches yendo al mismo bar (y creo que único de krakovia) para que Andre

pudiese practicar su polaco con el barman pidiendo zubruska (creo que no se escribe así), Nunca lo olvidaré y fue una de las mejores maneras de empezar. Aparte les debo a los dos mucho porque casi todo lo que se de R (hasta los nombre de los colores) es por Andre y todas mis dudas de Perl y de montones de tonterías fueron resueltas avidamente por Mireya.

Tambin quiero agradecer al resto de la gente del despacho. A el grupillo de la esquina con Alice, Macarena, Andre y Mireya. Se notó mucho cuando Alice se fue, que además siempre estaba ahí preparada para ayudarte con cualquier pregunta de trabajo o cualquier otra cosa. A Macarena, con la que siempre es un placer cotillear y más si también se apunta Alice.

Al principio cuando llegué, así de causualidad caí en el grupo de Lorena, Albert,Andrea,Christian, Elena,Alice que luego se uniria Eli y también Sergi y Raquel. Y ese viaje que nos pegamos a Berlin! Que hasta vino Iñaki. Y ahora cosas mas recientes, thanks a lot Sonja!! My roomie!! All these years side by side and at the end we finish living together!! They were very fun these months. Mi grupillo de chicas de ahora al que se le suma Steve. Muchas gracias Inma por ayudarme con Latex, con formulas... pero esto casi que es lo de menos, gracias. Y a Nuria, que junto con Inma, soys las mas entrañables del despacho. Muchas gracias por estar ahí, aunque luego tenga que ser yo la que tenga que explicar cosas de modernos. Y Steve, que ultimamente eras mi compi de sufrimientos con la tésis y con el que siempre me gusta comentar cual es la película que merece la pena en los cines. Y gracias a la gente del despacho que me ha ayudado cuando lo he necesitado. A Jaume y los de estructural, aunque no siempre se acuerden de pasarse a buscarme.

Nunca hay que olvidarse de agradecer a los sys admins todo (Miguel Angel y Alfons), porque sin ellos...mal. A mi admin particular para darle chapas sobre que ordenador me convendría mas, Judit.

Del grupo especialmente quiero agradecerle a Amadis, mi compaero de cafes y amigo. Que además en el trabajo me has ayudado mucho. Y Nico, en lo personal y en el trabajo muchas gracias, te debo de está última etapa. A ver si vuelves o tengo que ir a visitarte a Bariloche.

No me olvido de las cenas en porvenir y la montãna de pulpo.

A mis amigos... suerte que te encontré en el prbb Lorena. Lorena a tí te

debo tanto, que aquí no entraría y no hace falta que lo ponga porque ya lo sabes. Lucia, que haría sin tí. Muchas gracias por todo de verdad. A Iñaki, porque si le digo que lo necesito está, gracias. Carlos, siempre animándome, gracias. Anna, por aguantar chapas y porque siempre estas ahí también. Y porque soys Annacarlos, je y siempre estais ahí cuando os necesito. Judit otra vez, gracias y nor por lo del ordenador. A Marco, porque el si que está desde el principio desde la primera vez que pisé Barcelona casi. Gracias por los submarinos y más cosas. Y las noches frikis y el cine forum... La gente del máster y la UB y últimamente sobretodo a Olatz, lo vas a conseguir! Os voy a echar tanto de monos a todos.

Y a mi familia, porque si no me hubieran apoyado para hacer lo que siempre he querido hacer no estaría aqui, por confiar en mí (aunque no sepais exactamente lo que hago). A mi ama por que la quiero mucho y el estar haciendo esto, me supuso estar lejos. A mi hermana por que la quiero tanto y encima le rallo para que me haga portadas y dibujos... Y a mi aita, que lo que mas querría es que pudieses ver como por fin acabo. Pero hay cosas que son imposibles. Espero que estuvieses orgulloso. Eta nik txoria nuen maite.

Para los que no estan pero también para los que quedan, simplemente gracias.

Ekaitzaren ostean dator barealdia, edo hori esaten dute.

Eneritz Agirre
Barcelona, May 2013

Abstract

The regulation of alternative splicing has been generally thought of being primarily controlled by the interaction of splicing factors with the RNA molecule and by the elongation rate of the RNA polymerase II (RNAPII). There is an emerging understanding of the complexity of how alternative splicing is regulated which now involves the activity of non-coding RNAs and the chromatin state. Different experiments have shown that histone modifications can regulate the inclusion of alternative exons and that the elongation rate of the RNAPII could be influenced by different chromatin states. In this sense, small RNAs (sRNAs), which are a family of non-coding RNAs associated with members of the Argonaute family of proteins, that are effectors of the silencing pathway, which can participate in an alternative pathway known as transcriptional gene silencing (TGS). Experimental evidence shows that siRNAs targeting introns can induce chromatin marks that affect the rate of transcriptional elongation, affecting the splicing of pre-mRNAs, which is called transcriptional gene silencing alternative splicing (TGS-AS) (Allo et al., 2009). Thus, we proposed that the Argonaute protein (AGO1) could trigger heterochromatin formation and affect splicing by affecting RNAPII elongation.

In order to perform a genome-wide analysis of the regulation of alternative splicing we used new high-throughput sequencing technologies as ChIP-Seq and RNA-Seq. We found that there is AGO1 dependent alternative

splicing regulation, and our results suggest that endogenous sRNAs could be involved. Additionally, in the last part of the thesis we show a cell specific alternative splicing chromatin code, which also involves AGO1. Even though AGO1 regulation of alternative splicing was related to some specific cases, we found that other effectors, CTCF and HP1α were also important for the splicing changes decisions. This thesis and other recent reports show the regulation of alternative splicing as an integrated process, which involves many nuclear components and probably more that still need to be uncovered.

Resum

Generalment s'ha pensat que la regulació de l'splicing alternatiu està controlada principalment per la interacció entre els factors reguladors de l'splicing i la taxa d'elongació de la ARN polimerasa II (RNAPII). Hi ha un evidencia emergent de la complexitat de la regulació de l'splicing alternatiu, que ara també inclou l'activitat d'ARNs no codificants i l'estat de la cromatina. Diverses experiments han demostrat que modificacions en les histones poden regular la inclusió d'exons alternatius, i que la taxa d'elongació de la RNAPII pot estar influenciada pels diferents estats de la cromatina. Els ARNs petits (sRNAs) són una família d'ARNs no codificants associats amb membres de la família de proteïnes Argonauta i són efectors de la via de silenciació gènica. Alguns sRNAs participen en una via alternativa anomenada via de silenciació gènica transcripcional (TGS). Evidències experimentals ha mostrat que els sRNAs interferents que s'uneixen a introns poden promoure l'aparició de modificacions en les histones que alteren la taxa de elongació de la transcripció provocant canvis en l'splicing alternatiu. Aquesta via és coneguda com via de silenciació gènica transcripcional acoplada a splicing alternatiu (TGS-AS) (Allo et al., 2009). Tenint això en compte, nosaltres vam proposar que la proteïna Argonauta 1 (AGO1), podria induir la formació d'heterocromatina i canviar l'splicing alternatiu alterant l'elongaci de la RNAPII.

Per tal de realitzar una análisi a escala genómica de la regulaci de l'splicing

alternatiu, hem utilitzat dades provinents de noves tècniques de seqüenciació a gran escala, com ChIP-Seq i RNA-Seq. Hem trobat que hi ha regulació d'splicing alternatiu depenent d'AGO1. Els nostres resultats suggereixen que ARNs interferents endógens podrien estar relacionats amb aquesta regulació. A més, a la part final de la tesi demostrem que hi ha un codi de cromatina que requereix AGO1 que regula l'splicing alternatiu i que és específic per diferents tipus cellulars. Adicionalment hem trobat que altres efectors, com CTCF i HP1α, també sòn importants per explicar els canvis en l'splicing dels pre-ARNs. Conjunatment amb altres treballs, aquesta tesis demostra que la regulació de l'splicing alternatiu implica la funció de molts components nuclears i probablement de molts altres que encara han de ser descoberts.

Preface

When Phillip A Sharp and Richard Roberts described in 1977 the mechanism now known as splicing, we probably couldn't expect how this would impact the future of molecular genetics. During the 80s the details on the knowledge about the transcription mechanism started emerging with the discovery of the RNA polymerase in 1969 by Roeder and Rutter. This enzyme is the responsible of producing the primary transcripts including the pre-mRNAs in a process that requires several proteins for completion. Now we know that the elongation rate of the polymerase II is essential for the process and that the chromatin organization plays an unexpected role in the mechanism. Alternative splicing provides a way to explain proteomic complexity from a limited number of genes. With more than the 90% of the human genes being affected by alternative splicing, alternative splicing contributes to functional diversity and tissue specificity and it is responsible of different diseases. Nowadays alternative splicing regulation is known not only to depend on the interaction of splicing factors, but also in the coupling with the RNA transcription by RNA polymerase II (RNAPII), the role of chromatin structure and the emerging role of non-coding RNAs (ncRNAs). The study of how all these components are integrated, will lead to the better understanding of the regulation of alternative splicing.

The work in this thesis provides a step forward to the link between some of the regulators of alternative splicing. For long, splicing was thought to

occur mostly through splicing enhancers and silencers, modulating the use of alternative splice sites through the binding of specific regulatory proteins, now we know that this is extremely more complex. During the process of this thesis, the evidence of a chromatin structure that marks the exons has been described, the emerging roles and classes of sRNAs and later long non-coding RNAs have been found, histone modifications have appeared to be crucial for the splicing changes decisions, transcriptional gene silencing has been found to regulate alternative splicing and more recently, evidence of association between splicing machinery and components of the sRNA pathways was found. All this advances are closely related to the technological revolution of high throughput sequencing. In this context, these are still many things to be soved, like data processing approaches, experimental technical issues, biases in the processing and the interpretation of the high amounts of data that they are coming. This thesis analysis and results come from deep sequencing data which supposed in many of cases to start from a new approach for analyzing them.

In summary, here we show further evidence that histone modifications can regulate the inclusion of alternative exons and that the elongation rate of the RNAPII could be influenced by different chromatin states and also focus in the connection with non coding small RNAS. The first part of this thesis is based on the bioinformatics results we published in collaboration with Alberto Kornblihtt's group in (Allo et al., 2009), where experimental evidence showed that siRNAs targeting introns can induce chromatin marks that affect the rate of transcriptional elongation, affecting the splicing of pre-mRNAs, which is called transcriptional gene silencing alternative splicing (TGS-AS) (Allo et al., 2009). Our proposal is that the Argonaute protein (AGO1) could trigger heterochromatin formation and affect splicing by affecting RNAPII elongation. The second part of this thesis shows a continuation of the results from (Allo et al., 2009), where we used high troughput sequencing data in order to look for evidence of a genome wide connection between AGO1 and alternative splicing. The third part of the thesis describes a chromatin RNA code for alternative splicing that involves AGO1, CTCF, HP1α, RNAPII and different histone modifications. The three parts of the thesis are connected and provide as a whole a view of the epigenetics of alternative splicing regulation.

Contents

CHAPTER 1

Introduction

Contents

1.1 Overview

There is an emerging understanding of the complexity of how alternative splicing is regulated, which not only involves the activity of the spliceosome and RNA regulatory elements but probably also of the activity of non-coding RNAs and the chromatin state. Recently, several reports explain the role of epigenetics in the regulation of alternative splicing that also includes a possible role of non-coding RNAs acting on the chromatin structure and leading into a new mechanism that can affect the regulation of alternative splicing. High-throughput sequencing, carried out in consortia like ENCODE and in individual labs, is providing large amounts of data (RNA-seq, ChIP-seq, CLIP-seq, DNA-seq) in different tissues and cell lines. These new methodologies facilitate the understanding of the regulation of alternative splicing as an integrated process, which involves many nuclear components and probably more that still need to be uncovered.

1.2 The history of splicing

In 1977, Phillip A Sharp and Richard Roberts described the mechanism that we now know as splicing, where introns are removed from precursor messenger RNAs to create the mature transcripts. A simple comparison of the sequence of an mRNA and its corresponding nuclear DNA, revealed sequences that were removed by splicing during the processing leading to a non exact complementarity between the DNA and the mRNA (Berget et al., 1977) and (Chow et al., 1977) . Additionally, nuclear long RNAs and shorter cytoplasmic mRNAs showed the same termini, a cap and a poly(A) tail, while they were different in length due to the removal of the middle introns. Phillip A Sharp and Richard Roberts got the Nobel prize in 1993 for their findings.

Introns and exons show different consensus sequences at their boundaries, which after mutation inactivate splicing (Breathnach and Chambon,

1981). These boundaries are defined by four different splicing signals; the 5'splice site (5'ss), which corresponds to the beginning of the intron, the 3'splice site (3'ss), which corresponds to the end of the intron, the branch site (BS) and the polypyrimide tract (PPT). Developments in biochemistry using soluble reactions led to realize that the intron was excised in a branch-structure with a lariat RNA (Padgett et al., 1986). Even before, several hypotheses based on the complementarity between intron consensus sequences with some small nuclear RNAs (snRNAs) (U1,U2,U4/U6 and U5), indicated that these snRNAs might be important for splicing (Lerner et al., 1980). These snRNAs containing particles, the small nuclear ribonucleoproteins (snRNPs), were found to form part of a macromolecular complex, the spliceosome.

1.3 The splicing reaction and the spliceosome

Splicing is carried out by the spliceosome, in which five snRNPs and a large number of auxiliary proteins cooperate to accurately recognize the splice sites and catalyse the two steps of the splicing reaction. Both the conformation and composition of the spliceosome are highly dynamic, affording the splicing machinery its accuracy and at the same time flexibility. There are two types of spliceosomes: the U2-dependent spliceosome, which catalyzes the removal of U2-type introns, and the less abundant U12-dependent spliceosome (Hall and Padgett, 1994), which is present in only a subset of eukaryotes and splices the rare U12-type class of introns (Will and Luhrmann, 2011). From now on, I will only discuss the major spliceosome. Spliceosome assembly begins with the recognition of the 5′ ss by the snRNP U1 and the binding of splicing factor 1 (SF1) to the BS and of the U2 auxiliary factor (U2AF) heterodimer to the PPT and 3′ terminal AG. This assembly is ATP independent and results in the formation of the E complex, which is converted into the ATP-dependent, pre-spliceosomal A complex after the replacement of SF1 by the U2 snRNP at the BS. Further recruitment of the U4/U6U5 tri-snRNP complex leads to the formation of the B complex, which is converted into to the catalytically active C complex after extensive conformational changes and remodelling (Chen and Manley, 2009).

Exon and Intron definition

There is a clear relation between the length of the exon and flanking introns and how the splice sites are recognized (Marais et al., 2005). There are two models, exon definition (Robberson et al., 1990), where the exon is recognized as a unit by the splicing machinery before intron excision; and intron definition, when the 5′ss and the 3′ss are recognized directly as a unit of splicing (Berget, 1995). Intron definition is used when the length of the intron is between 200-250 nt. This mechanism works inefficiently when the introns

get longer.; in such cases, the recognition is mediated via the exon definition (Hertel, 2008). In such cases, in higher eukaryotes exons are first defined by the splicing machinery and then the intron is removed in a second step.

Splicing signals

The intron is defined by the short conserved sequences at the 5'ss, 3'ss and BS. The 5'ss is defined by the presence of mostly GT or GC dinucleotide that marks the beginning of the intron (Breathnach et al., 1978; Catterall et al., 1978). It is mostly characterized by 9nt positions, 3 from the end of the upstream exon, and the 6 first positions from the downstream intron. This sequence approximately matches the complementary sequence of U1 snRNA (Seraphin et al., 1988; Siliciano and Guthrie, 1988). The 3'ss delimits the intron-exon boundary at the 3' end of the intron. This signal consists of an AG dinucleotide, preceded by a T or a C (Mount, 1982). Although it only contains 3 nt the 3'ss is crucial in the second step of the splicing process in most of the introns.

The BS is typically located 18-40 nucleotides upstream from the 3'ss and in higher eukaryotes is followed by a PPT. The BS is characterized by the presence of an A, which is fundamental in the first step of the splicing reaction (Padgett et al., 1984). The splicing efficiency is affected by the position at which the BS is located, longer distance between the BS and the 3'ss leads to a lower splicing efficiency (Cellini et al., 1986). Different splice site and branch site sequences are found in U2-type versus U12-type introns. The U2-type consensus sequences found in *Saccharomyces cerevisiae* exhibit a higher level of conservation than those in metazoans, reviewed in (Will and Luhrmann, 2011; Smith and Valcarcel, 2000; Wang and Burge, 2008).

There are more sequence elements that are crucial in the recognition and usage of the splice sites. They are typically short and diverse in sequence and modulate both constitutive and alternative splicing by binding regulatory proteins that either stimulate or repress the assembly of spliceosomal complexes at adjacent splice site. These include the exonic splicing enhancers (ESEs), exonic splicing silencers (ESSs), intronic splicing enhancers (ISEs)

and intronic splicing silencers (ISSs), which are able to promote or inhibit exon recognition by the spliceosome (Matlin et al., 2005). These sequences allow the correct identification of exons, distinguishing them from pseudoexons (Corvelo and Eyras, 2008). Splicing enhancers and silencers are short conserved sequences that can be found isolated or in clusters and widely distributed within the genome. They act mostly through specific binding of regulatory proteins like SR proteins and hnRNPs (Dreyfuss et al., 1993; Graveley, 2000), (Manley and Tacke, 1996). ESEs are usually bound by SR proteins while ISSs and ESSs are commonly bound by hnRNPs. However, there are some splicing silencers that form a particular pre-mRNA secondary structure that may prevent the recognition of the splicing enhancer by SR proteins (Buratti and Baralle, 2004) and (Sirand-Pugnet et al., 1995). ISEs are not as well characterized as the other three types of elements, although recently, several proteins, such as hnRNP F, hnRNP H, NOVA1, NOVA2, FOX1 and FOX2, have been shown to bind ISEs and in that way stimulate splicing (Ule et al., 2006; Hui et al., 2005; Yeo et al., 2009; Mauger et al., 2008). Some of these factors show expression in a tissue specific manner, so they are able to regulate splicing in specific tissues. The combined action of splicing factors, such as SR proteins and hnRNPs, can promote or inhibit spliceosome assembly upon different external clues, leading to alternative splicing. This is particularly common in exons with weak splice-sites, which are more dependent on splicing factors for their inclusion.

1.4 Alternative splicing

Alternative splicing provides a way to explain proteomic complexity from a limited number of genes, as found in mammals. With more than the 90% of the human genes being affected by alternative splicing (Pan et al., 2008; Wang and Burge, 2008), alternative splicing contributes to functional diversity and tissue specificity (Graveley, 2001). Regulation of alternative splicing have long been thought to occur mostly by splicing enhancers and silencers, modulating the use of alternative splice sites through the binding of specific regulatory proteins and at the early spliceosome assembly, but this is not always the case (Smith and Valcarcel, 2000; Chen and Manley, 2009). The use of regulatory sequences, enhancers and silencers (Caceres and Kornblihtt, 2002) and the secondary structure of the nascent mRNA can influence positively or negatively the selection of a specific splice site (Blanchette and Chabot, 1997; Buratti and Baralle, 2004). There are also different proteins that can affect directly binding to the mRNA or interacting with other proteins from the splicing machinery. Which affects the inclusion of an alternative exon in the mature transcript.

The current view poses that alternative splicing regulation does not only depends on the interaction of splicing factors, but also in the coupling with the RNA transcription by RNA polymerase II (RNAPII) (Munoz et al., 2009; Kornblihtt et al., 2004; de la Mata et al., 2003), the role of epigenetics such as chromatin structure and histone modifications (Tilgner et al., 2009; Schwartz et al., 2009; Spies et al., 2009; Kolasinska-Zwierz et al., 2009; Nahkuri et al., 2009; Andersson et al., 2009; Luco et al., 2010) and the emerging role of non-coding RNAs (ncRNAs) as novel regulators of alternative splicing (Allo et al., 2009) reviewed in (Allo and Kornblihtt, 2010; Allo et al., 2010; Luco and Misteli, 2011; Luco et al., 2011). The study of how all these components are integrated, will lead to the better understanding of the regulation of alternative splicing.

Types of alternative splicing

In alternative splicing, different combinations of splice sites can be joined to each other. In a typical multiexonic mRNA, the splicing pattern can be altered in many ways. The are five main types of alternative splicing (Breitbart et al., 1987): Alternative 5'ss usage, alternative 3'ss usage, inclusion or skipping of an entire exon (cassette exon), splicing or retention of an entire intron (intron retention) and the combination of two differentially included exons (mutually excluded exons). There are other mechanisms that can change the mRNA composition. The 5-terminal exons of an mRNA can be switched through the use of alternative promoters and alternative splicing. Similarly, the 3-terminal exons can be switched by combining alternative splicing with alternative polyadenylation sites (Colgan and Manley, 1997). In this thesis I will center specifically in exon skipping events, delimiting most of the results and conclusions to these so-called cassette exons.

Co-transcriptional splicing

Alternative pre-mRNA splicing is closely linked to transcription. Much of splicing occurs co-transcriptionally (Tilgner et al., 2012) and the splicing machinery is physically linked to the transcriptional components via association of splicing factors with the elongating RNAPII (Perales and Bentley, 2009). Early models differentiated spatially the spliceosome compartment and transcription. However, recent findings by different groups have provided new evidence for co-transcriptional splicing: the visualization of Drosophila embryo nascent transcripts by electron microscopy revealed looped RNAs attached to chromatin (Beyer and Osheim, 1988), spliced mRNAs were found associated with mechanically dissected or biochemically fractionated chromatin (Bauren and Wieslander, 1994; Pandya-Jones and Black, 2009; Gornemann et al., 2005; Kotovic et al., 2003; Lacadie and Rosbash, 2005; Listerman et al., 2006), RNA in situ hybridization with splice junction probes detected spliced mRNAs at their gene loci (Zhang et al., 1994) and splicing factors were localized at sites of transcription using inmunofluorescence (Misteli and Spector, 1999). In human, co-transcriptional splicing was first showed in the dystrophin gene (Tennyson et al., 1995). All these findings show the clear evidence of a spatial and functional coupling between transcription and splicing.

From a kinetic point of view, the co-transcriptionality of splicing at the 5′ end of genes makes sense due to the composition of a typical eukaryotic cell gene, short exons and long introns. This structure makes it easier for the splicing machinery to recognize de 5′ss and the 3′ss of a transcribed exon while the elongation complex is proceeding through the downstream intron (Pandya-Jones and Black, 2009). Most introns at the 5′ end are co-transcriptionally removed, but there is a fraction near the 3′ end of genes that is spliced post-transcriptionally (Bauren and Wieslander, 1994; Pandya-Jones and Black, 2009). There is in fact a general gradient related with the direction of the gene transcription, from co-transcriptional to post-transcriptional splicing. However, there are exceptions where internal splicing events occur after transcription (Kessler et al., 1993). Now, with recent imaging experiments of live-cell microscopy, it is possible to track the kinetics of transcrip-

tion and splicing in vivo (Brody et al., 2011). Moreover, high-throughput techniques like RNA-Seq of nascent transcripts showed the predominance of co-transcriptional splicing. For instance, 87% of the analyzed introns by Khodor et al. were co-transcriptionally spliced more than 50% of the time in Drosophila (Khodor et al., 2011) and 84% of exons that were flanked by a large intron showed clear evidence of co-transcriptional splicing using RNA from fetal brain tissue in human (Ameur et al., 2011). The combination of nascent transcription and co-transcriptional splicing leads to a specific pattern over the introns, with high RNA-Seq signal at the 5′ end decreasing towards the 3′ end of the intron. This is explained by the successive rounds of transcription entering the 5′ end of the intron combined with the excision and release of the intron from the nascent transcript, once its synthesis is completed (Ameur et al., 2011).

Transcription and splicing are not always coupled. There is a coupling when there is an interaction between the splicing and transcription machineries or when the kinetics of one process affects the outcome of the other. The optimal coordination between the transcription and the processing seems to be specific of the RNAPII and of the carboxy terminal domain (CTD), its catalytic subunit (Sisodia et al., 1987; Dower and Rosbash, 2002). The CTD consists of tandem YSPTSPS repeats, which vary in number from 26 in yeast to 52 in humans (Phatnani and Greenleaf, 2006). Dynamic phosphorylation of serine residues on the CTD heptad repeats is associated with the stages of RNAPII elongation (Sims et al., 2004) and is required for the stimulatory effect of the CTD on splicing (Hirose et al., 1999; Millhouse and Manley, 2005). The serine residues are referred to as serine-2, serine-5 and serine-7, depending on their position in the repeat. Phosphorylation of serine-5 in the heptad is most prominent at the 5′ end of genes, while serine-2 phosphorylation increases toward the 3′ end and is characteristic of an elongating RNAPII (Komarnitsky et al., 2000; Schroeder et al., 2000). When the CTD is deleted, the three steps of pre-mRNA processing in vertebrates; capping, splicing and polyA site cleavage, are inhibited. CTD deletion is also able to affect the overall levels of transcription but it does not necessarily affect the accuracy of initiation (McCracken et al., 1997a,b). It is now well established that the CTD plays a direct and major role in coupling transcription with processes such as chromatin modification and pre-mRNA processing

(Egloff et al., 2012; Buratowski, 2009). Modifications in CTD undergo dramatic changes during transcription to recruit factors at the appropriate point of the transcription cycle (Egloff et al., 2012). It is known that specific splicing factors recognize specific CTD modifications. As an example, Prp40 and U2AF65 recognize the serine-2 serine-5 double mark, U2AF65 then recruits PRP19C, subsequently activating splicing.

The elongation rate

Eperon et al showed in 1988 that the rate of RNA synthesis affects its secondary structure, thereby affecting splicing (Eperon et al., 1988). In a different experiment, inserting a sequence that leads to the pausing of the RNAPII in the tropomyosin gene, higher inclusion in tropomyosin exon 3 was observed (Roberts et al., 1998). Another evidence for a role of the RNAPII elongation in alternative splicing is that promoter identity and occupation by transcription factors modulates alternative splicing (Cramer et al., 1997; Kornblihtt, 2005), which affects alternative splicing outcome (Cramer et al., 1997, 1999; Auboeuf et al., 2002; Pagani et al., 2003; Robson-Dixon and Garcia-Blanco, 2004).

The first promoter effect in alternative splicing was observed using reporter minigenes for the alternatively spliced cassette exon 33, also known as EDI or EDA, of human fibronectin (FN) (Cramer et al., 1997, 1999). The insertion of pausing elements in reporter minigenes affects the elongation rate of RNAPII, leading to the inclusion of the alternative exon when RNAPII is paused or slowed down (de la Mata et al., 2003). The same mechanism was shown using the fibroblast growth factor receptor 2 (FGFR2) gene (Robson-Dixon and Garcia-Blanco, 2004). The effects on the different inclusions of these exons were not the consequence of the promoter strength, but depended on some qualitative properties conferred by promoters to the transcription/RNA processing machinery. These findings opened a way to consider that factors and elements, classically defined as transcriptional regulators, could also be necessary for splicing regulation. Recent results have shown that inhibition of RNAPII elongation can yield not only the inclusion but also the skipping of alternative exon (Ip et al., 2011), probably due to

splicing silencing mechanisms playing a role at some genes at low RNAPII elongation.

1.5 Chromatin as a regulator of alternative splicing

The control of RNAPII elongation rate and the factors associated with transcription do not seem to be sufficient to completely explain the regulation of alternative splicing. A major recent discovery is that chromatin structure can act as a key regulator of alternative splicing. In 1991, Adamis and Babiss proposed that pre-mRNA splicing might be regulated by chromatin involving changes in transcriptional elongation rates (Adami and Babiss, 1991). Later, the model was confirmed by Kornblitth's lab: they found evidence in mammalian cells that alternative splicing can be modulated using a slow mutant RNAPII (de la Mata et al., 2003) but also applying a treatment with a histone deacetylase inhibitor TSA (Nogues et al., 2002), which confirms that a change in the chromatin state can regulate alternative splicing. Further support came from the results in CD44 using minigenes: the effect on splicing was due to the recruitment of specific hormone receptor coregulators that remodeled the chromatin (Auboeuf et al., 2002). Later results showing that chromatin remodelers of the SWI/SNF family have an effect in alternative splicing (Batsche et al., 2006) would definitively confirm the importance of chromatin structure in splicing. Moreover, results from Groudine's lab provided consistency to this idea, showing that methylation of a DNA sequence within a gene affects locally chromatin leading to RNAPII slow down without affecting transcription rate (Lorincz et al., 2004).

The relation between chromatin and splicing may have an explanation in an structural relation between exons and nucleosomes. Nucleosomes are defined as a extent of 147 bp DNA wrapped around an octamer of four core histone proteins (H3,H4,H2A and H2B), the structural unit that determines the conformation and compaction of chromatin (Luger et al., 1997; Kornberg and Lorch, 1999). Nucleosomes can act as barriers that locally modulate the progression of RNAPII (Hodges et al., 2009), suggesting that nucleosome density and chromatin structure can modulate the RNAPII kinetics and in

that way regulate alternative splicing (Subtil-Rodriguez and Reyes, 2010). Exons and introns have a positional preference within a gene, which was observed by the finding of a periodical pattern for successive 3'ss and 5'ss compatible with nucleosome phasing (Beckmann and Trifonov, 1991). In agreement with these prior evidences, genome wide mapping of nucleosomes allowed to find their enrichment at exons, independently of gene expression (Kolasinska-Zwierz et al., 2009; Schwartz et al., 2009; Tilgner et al., 2009) and their positioning seem to be determined by the very same sequences defining introns. Enrichment of the nucleosome around exons is conserved in evolution and is found in gametes and somatic cells (Nahkuri et al., 2009). The differences observed between alternatively spliced exons, with lower enrichment of nucleosomes, and constitutive exons flanked by long introns, with higher enrichment, suggest a link between chromatin structure, nucleosome positioning and the regulation of alternative splicing. In this model, the chromatin structure would be capable of determining the splicing choices.

Several groups have used chromatin inmunoprecipitation followed by high-throughtput sequencing (ChIP-Seq) or by microarray analysis (ChIP-chip) to examine the relation between histone modifications and exon-intron boundaries (Kolasinska-Zwierz et al., 2009; Tilgner et al., 2009; Schwartz et al., 2009; Spies et al., 2009; Andersson et al., 2009). They have shown that histone modifications and DNA methylation are not randomly distributed through exons and introns, which suggests a probable relation to exon definition and therefore to splicing. Some histone modifications, like H3K36me3, H3K4me3 and H3K27me2, showed enrichment at exons, while others like H3K9me3 showed depletion (Spies et al., 2009); and these enrichments levels were not all consequence of transcriptional activity (Spies et al., 2009). However, H3K36me3 which is a transcription mark, was more highly enriched in constitutive exons than in the alternative ones in active genes (Kolasinska-Zwierz et al., 2009; Andersson et al., 2009).

To explain how alternative splicing is regulated by chromatin, two non-exclusive models have been proposed; the recruitment model, by which different factors associate with the histone modifications and the transcription machinery to regulate splicing choices, and the kinetic model; where the chromatin state controls exon inclusion by modulating RNAPII elongation.

A third mechanism has also been proposed, in which both chromatin state and transcriptional regulation are influenced by spliceosomal factors (Lin et al., 2008; Zhou et al., 2011; de Almeida et al., 2011; Kim et al., 2011).

The kinetic model is based in two assumptions: first, there is a variable elongation rate and second, its modulation results in alternative splicing (see The elongation rate section above). Chromatin states control the RNAPII elongation rate. The evidence of a preferential nucleosome positioning over the exons indicates a possible modulation of the elongation rate in both constitutive and alternative splicing. As we commented previously, the nucleosome can act as a roadblock for the transcription, which gives time to the splicing machinery to ensure the coordinated recruitment and assembly of the spliceosome. These mechanisms have been observed in a number of cases: For instance, membrane depolarization of neuronal cells in the regions around exon 18 in the NCAM gene, triggers H3K9 hyperacetylation leading to the skipping of the exon due to the chromatin relaxation and the enhancement of transcriptional elongation rate (Schor et al., 2009). In the FN gene, where previous reports had shown the effect of the transcription rates in the inclusion of the alternative exon (de la Mata et al., 2003), it was shown that targeting siRNAs to the downstream intron of the alternative exon produces an increment of heterochromatin marks, H3K9me2 and H3K27me3, and the recruitment of HP1α, slowing down the elongation rate and leading to the inclusion of the alternative exon (Allo et al., 2009). Further results support the kinetic model: the CD44 gene was found to be enriched in H3K9me3 and HP1γ, a transcriptional repressor, acting as a barrier for elongation and leading to the inclusion of alternative exons (Saint-Andre et al., 2011). More recently, the CCCTC-binding protein (CTCF) was found to bind the DNA near an alternative exon of the gene CD45, which would locally pause RNAPII, affecting the inclusion of the alternative exon (Shukla et al., 2011b). Since CTCF cannot bind to mehylated DNA, the effect of CTCF in splicing is dependent on the methylation of the target sites. Vezf1, another binding protein, has been found to regulate splicing by modulating transcriptional elongation similarly to CTCF, and there is evidence for the interaction between Vezf1 and Mrg15/Mrgbp, a protein that recognizes H3K36me3 (Gowher et al., 2012).

The recruitment model is based on the assumption that some chromatin features can directly interact with the splicing machinery, thereby affecting the splicing choices. For instance, specific histone modifications can direct the recruitment of splicing factors to the nascent RNA via formation of a chromatin-splicing adaptor complex (Luco et al., 2010), reviewed in (Luco and Misteli, 2011; Luco et al., 2011). Initial results also showed the recruitment of the U2 snRNP to H3K4me3 through the direct interaction with the binding protein CHD1, a chromatin remodeling factor related to transcription elongation and open chromatin maintenance (Selth et al., 2010), and one of the components of U2 snRNP, the SF3a1 (Sims et al., 2007). Similarly, there is evidence of HP1 binding to H3K9me3 and interacting with ASF/SF2 and various hnRNP splicing repressors (Loomis et al., 2009; Piacentini et al., 2009). More recently, the analysis of different PTB-dependent genes showed that the MRG15 protein interacts with H3K36me3 inhibiting the inclusion of the alternative exon (Luco et al., 2010). Several lines of evidence support the idea of a direct interaction between H3K36me3 and proteins involved in pre-mRNA splicing. For instance, the Psip1/Ledgf protein specifically recognizes H3K36me3 in active genes. Psip1/Ledgf co-localizes and interacts with SRSF1 and other members of the SR protein family (Pradeepa et al., 2012). The direct interaction through SRSF1, H3K36me3 and splicing factors could possibly modulate pre-mRNA splicing (Pradeepa et al., 2012). Recent results suggest that the CHD1 protein could preferentially remodel H3K36-methylated nucleosomes, mediating its activity in coding regions in association with RNAPII in *S. cerevisiae* (Smolle et al., 2012). If this mechanism is conserved in higher eukaryotes and is localized near alternative exons, it could represent another protein interacting with H3K36me3. Since H3K36me3 is enriched in exons (Kolasinska-Zwierz et al., 2009) and there is increasing evidence of its role as a major recruitment mark of spliceosome components, the relevance of H3K36me3 in the regulation of splicing could be studied more generally using genome-wide datasets, to extend it beyond the subset of genes for which there is evidence so far.

Results from different studies indicate that the kinetic model and the recruitment are not mutually exclusive. In most cases, the recruitment of splicing factors through the interaction with chromatin leads to changes in the modulation of RNAPII elongation rates, leading to the combination of the

two models in an interconnected system. That is the case of Brahma (Brm), a subunit of the SWI/SNF complex, which binds to the CD44 alternative exons and interacts with spliceosomal components and the RNA-binding protein Sam68. At the same time Brm, triggers the accumulation of RNAPII (Batsche et al., 2006). Similarly, HP1γ binds to H3K9me3, favoring the inclusion of alternative exons, probably by affecting the elongation of the RNAPII (Saint-Andre et al., 2011).

All the facts presented before, show that the relation between chromatin and splicing may be more complex than initially expected. The observation that H3K36me3 can be dependent on splicing (de Almeida et al., 2011), suggests that in actively transcribed genes transcription and pre-mRNA splicing can reach back to chromatin. This leads to the proposal of an interconneted feedback loop between RNAPII, chromatin and splicing (de Almeida and Carmo-Fonseca, 2010).

1.6 Non-coding RNAs in alternative splicing

Overview

Non-coding RNAs (ncRNAs) were first initially proposed to be regulators of gene transcription in mammalian cells (Green and Weinberg, 2011; Gagnon and Corey, 2012). However, they are now known to regulate genes and genomes at different levels, including chromatin structure, transcription, RNA stability and translation (van Wolfswinkel and Ketting, 2010; Buhler and Moazed, 2007; Carthew and Sontheimer, 2009). Furthermore, they can act as activators or inhibitors and their disruption has been linked to disease (Taft et al., 2010). Recent reports show that there is an emerging role for ncRNAs as regulators of alternative splicing by regulating the expression of key splicing factors (Makeyev et al., 2007; Boutz et al., 2007; Kalsotra et al., 2010), by RNAi mediated transcriptional gene silencing (Allo et al., 2009; Ameyar-Zazoua et al., 2012), by direct regulation with ncRNAs binding directly to the target pre-mRNAs (Kishore and Stamm, 2006a,b; Kishore et al., 2010) or even regulating levels of active SR proteins, as is the case of the long ncRNA MALAT1 (Tripathi et al., 2010; Zong et al., 2011) and the more recently described new class of snoRNA like long ncRNA (Yin et al., 2012). Here, I will focus on the nuclear RNAi process, the transcriptional gene silencing and the Argonaute proteins, specially Argonaute-1 (AGO1), which are the integral effectors of the transcriptional and post-transcriptional RNA silencing pathways.

RNA interference

In eukaryotes, sRNA-mediated epigenetic gene silencing pathways are well conserved. sRNAs can regulate gene expression in mammals, *Drosophila*, *Caenorhabditis elegans* and plants by three different conserved pathways; transcriptional gene silencing (TGS), post-transcriptional gene silencing (PTGS) and translational (miRNA) inhibition (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). RNAi consists on small RNAs that act cotranscriptionally targeting the nascent RNA while the effector complexes interact and regulate the transcriptional machinery (Castel and Martienssen, 2013). In eukaryotes the RNAi machinery is one of the most conserved components (Ketting, 2011). RNAi is initiated in the cytoplasm by long dsRNAs or hairpin RNAs which are processed by Dicer (DCR) into short duplex small RNAs (Bernstein et al., 2001). Small RNAs can enter through different ways in RNAi pathways, but they all exert their effects through the RNA-induced silencing complex (RISC), that at its core has an Argonaute protein. After being processed by DCR, duplex RNAs are loaded onto Argonaute, which is known as the component of RISC which directly interacts with the sRNA (Hammond et al., 2001).

The discovery of RNA interference (RNAi) (Fire et al., 1998), showed that there was a class of small RNAs with important regulatory functions (Hamilton and Baulcombe, 1999), collectivelly called sRNAs. They are generally short $\sim$ 18-30 nucleotides long; they do not code for proteins; exert their function as RNA molecules generally combined with protein factors; and represent a substantial portion of the RNA output of cells. Small RNAs are pervasive throughout eukaryotes and they are also present in some archea and eubacteria (Hock and Meister, 2008; Karginov and Hannon, 2010). Transcriptional gene silencing (TGS) was the first function of nuclear RNAi to be described, where RNAi reduces transcription guiding localized heterochromatin at the target region. TGS takes place through DNA methylation or accumulation of histone modifications and has been related to protection of genome integrity by silencing of repetitive regions of the genome (Sabin et al., 2013). Interestingly, there is also evidence of endogenous sRNAs mediating the TGS (Kim et al., 2008). TGS caused by siRNAs was first observed in

plants showing the supressed phenotype of a transgene (Matzke et al., 1989). This process was later explained by the involvement of DNA methylation in the targeted gene as the cause of suppression. Evidence for siRNA mediated TGS in *S. pombe* was correlated with histone 3 lysine 9 (H3K9) methylation (Mette et al., 2000; Jones et al., 2001; Volpe et al., 2002). On the other hand, post-transcriptional gene silencing (PTGS) is related to either the destruction or the translational inhibition of mRNA, where the transcription of the gene is not affected but gene expression is lost due to unstable mRNA molecules. In this RNAi pathway, short double-stranded RNAs can trigger post-transcriptional silencing through sequence specific recognition of endogenous transcripts (Joshua-Tor and Hannon, 2011). PTGS was also first found in plants, where nuclear-run on assays clearly showed that the transcript was present but that it failed to accumulate in the cytoplasm (Ingelbrecht et al., 1994). Although sRNAs typically known to silence gene expression by mRNA degradation, recent results show a broader function, including the regulation of transcription and splicing, leading to the evidence of nuclear AGO-RNA complexes in mammals (Allo et al., 2009; Ameyar-Zazoua et al., 2012).

Argonaute protein family

The Argonaute proteins form an evolutionarily conserved family defined by their role in silencing the expression of genes in various ways (Murphy et al., 2008). Moreover, all RNAi dependent pathways share the presence of an Argonaute protein. Argonaute proteins are highly conserved and are present from Archaea to human (Murphy et al., 2008). The number of AGO genes changes widely from 1 in *S. pombe* to 27 in *C. elegans* (Meister and Tuschl, 2004). The Argonaute protein family was first identified in plants (Bohmert et al., 1998). They are known to be key players identifying RNAi targets through the binding to 21-35 nt long sRNAs, whose sequence identifies the genes to be silenced. These Argonaute-RNA complexes carry out different functions like repression of gene transcription, cleavage or degradation of target mRNAs, or even block mRNA translation. Argonaute proteins are divided in two subfamilies, the AGO subfamily and the PIWI subfamily. In

human there are 8 Argonaute proteins, 4 from the AGO clade (AGO1-4) and 4 from the PIWI clade (PIWIL1-4) (Carmell et al., 2002; Sasaki et al., 2003). Ago1, Ago3 and Ago4 genes are clustered in chromosome 1 while Ago2 is located in chromosome 8 (Hock and Meister, 2008). In humans AGO proteins bind to 21 nt small interfering RNAs (siRNAs) and 21-23 nt microRNAs (miRNAs). AGO proteins are essential for development and differentiation. The PIWI subfamily of genes are located in chromosomes 12, 11, 22 and 8, (Hock and Meister, 2008). While AGO subfamily is known to be expressed ubiquitously in many organisms, the expression of PIWI proteins is restricted to the germ line, where they bind 23-30 nt Piwi interacting RNAs (piRNAs). However, recent findings show evidence of PIWI proteins and their bound piRNAs in the somatic cells (Yin and Lin, 2007; Lin and Yin, 2008), showing interaction with HP1α which suggests a more general epigenetic function of piRNAs.

Argonaute proteins show three key structural characteristic domains: PAZ, PIWI and MID domains. The PAZ domain recognizes and binds the 3' end of sRNAs (Yan et al., 2003; Lingel et al., 2004; Ma et al., 2004) and the PIWI domain is an RNase H-like fold that harbors the slicer activity for cleavage of target RNA substrates (Song et al., 2004; Wang et al., 2008b,c). The MID domain anchors the 5' monophosphate of an sRNA to the Argonaute protein, ensuring the guide by multiple cycles of target cleavage (Parker et al., 2005; Ma et al., 2004). For several years, the analysis of the crystal structure was limited to isolated domains of the proteins and to archaeal full length proteins (Yuan et al., 2005). Recently, the crystal structure of the human Ago2 was solved revealing new features not present in prokaryotes and also providing new mechanistic insights about the interaction between the protein and the target molecules (Schirle and MacRae, 2012). Moreover, Elkayam et al provided the structure of Ago2 loaded with miR-20a, a 20 mer miRNA, and showed that the guide is threaded through the entire protein structure, in interaction with every domain (Elkayam et al., 2012).

The heterochromatic gene silencing

The silent state of chromatin was believed in the past to be only the consequence of the nucleosome packaging into a dense state due to the different histone modifications. This dense state would act as a barrier for the transcriptional machinery (Grewal and Moazed, 2003) preventing access of the RNAPII to the DNA. Currently, it is known that transcription is fundamental for the assembly of heterochromatin. Several classes of histone modifications have been described, but the most studied have been methylation and acetylation (Kouzarides, 2007). While acetylation is related with gene activation, methylation can act as activator or repressor of transcription. Methylation of histone H3 on Lys9 (K9) and Lys27 (K27) and of histone H4 on Lys20 (K20) are strongly correlated with gene silencing. In particular, H3K9me has an important role in heterochromatin formation, it is known to be highly enriched in condensed heterochromatin and it is recognized by the heterochromatin protein 1 (HP1). Experimental evidence in Drosophila shows that apart from being part of the large domains of the pericentromeric regions, HP1α was also found to bind the transcription regions in genes that were actively transcribed. The transcriptional repression mechanism has been well describeb in *S. pombe* were there is evidence of Argonaute and siRNA complexes that associate with H3 Lys9 (H3K9) methyltransferase Clr4 to promote silencing marks at target regions. Then, the methylation of H3K9 leads to the recruitment of HP1 homolog swi6 that promotes silencing (Nakayama et al., 2001; Maison and Almouzni, 2004).

Evidence that siRNAs can mediate TGS in mammalians cells has been reported from different groups (Morris et al., 2004; Castanotto et al., 2005; Suzuki et al., 2005; Ting et al., 2005; Gonzalez et al., 2008). These groups showed that siRNAs targeting the promoter region or nearby can lead to the silencing of the gene (Morris et al., 2004; Castanotto et al., 2005; Suzuki et al., 2005; Ting et al., 2005). The common patterns of the mechanism involve increase of DNA methylation and/or H3K9 methylation within the targeted region. Chromatin modifications seemed to have an important role in the silencing effect, as a first requirement for establishment of DNA methylation. Morever, in plants, before DNA methylation and gene silencing, H3K27me3

histone modification is required (Lindroth et al., 2004). Weinberg et al provided another property of the mechanism, showing that it is the antisense strand of the siRNA and not the sense, the one who directs histone methylation, and that the inhibition of RNAPII prevents the effect (Weinberg et al., 2006). This means that TGS mediated by antisense siRNAs targeted to promoter regions is RNAPII dependent, and hence, it is connected to transcription (Weinberg et al., 2006). The effect of RNAi and cotranscription appeared like a commonplace for understanding the spreading of heterochromatin marks (Zaratiegui et al., 2007). First evidences showed similarities between cotranscriptional silencing and position effect variegations (PEVs). The PEVs occur in cis within the reach of RNAPII transcription and can influence the target genes with silencing elements such as histone modifications and RNAi in a limited distance (Irvine et al., 2006). This explains how the RNAi machinery silences elements far from the origin of the siRNA as it happens in the case of the PEVs.

In plants, AGO4 directs chromatin modifications and directly correlates with H3K9me2 (Zilberman et al., 2003). In *S. pombe*, Ago1 protein correlates with H3K9me2 formation and the mechanism also involves the RNA induced transcriptional silencing (RITS) complex (Verdel et al., 2004). In particular, AGO1 and AGO2 were found to be implicated in sRNA mediated regulation. The Argonaute protein with a more clear role in RNAi is AGO2. AGO2 acts as the catalytic engine that drives mRNA cleavage, while AGO1, which is almost identical to AGO2, cannot cleavage RNA in a efficient way. Both, AGO2 and AGO1 are found to be present in the nucleus (Ohrt et al., 2008; Ahlenstiel et al., 2012), leading to a possible connection with the chromatin. In human cells, AGO1 was found to be involved in the TGS (Kim et al., 2006; Janowski et al., 2006) while AGO2 was only involved in PTGS (Janowski et al., 2006). Moreover, AGO1 associates with RNAPII having an unphosphorilated CTD, probably preventing its elongation through the siRNA targeted promoter region (Janowski et al., 2006; Kim et al., 2006). All the results showed that the two main histone modifications involved in the process were H3K27me3 and H3K9me2 silencing marks. These two silencing marks, are characteristic of facultative heterochromatin and are the histone modifications associated with siRNA mediated TGS (de Wit et al., 2007).

Argonaute proteins and endogenous sRNA-directed epigenetic processes

First evidences showed that AGO1 could act as effector for the TGS with endogenous miRNAs, as it does in siRNA mediated TGS in human (Kim et al., 2008). In this case, AGO1 was enriched at the POLR3D gene promoter followed by an overexpression of miR-320 in HEK-293 cells(Kim et al., 2008). The POLR3D promoter showed an enrichment of the silencing H3K27me3 histone modification. Younger et al, transfecting sequences that mimic miRNAs, showed that miR-423-5p mediated TGS when targeting the PR gene promoter in association with AGO2 and correlating with an enrichment of H3K9me2 (Younger and Corey, 2011). In that case, there was no proof that the mechanism could be endogenous. Besides, they also found evidence of an sRNA similar to a miRNA that can act in the TGS, targeting the promoter region as in the typical siRNA-directed TGS. Later, several reports suggested that epigenetic changes can be related with endogenous miRNA inducing activation of transcription. For instance, Huang et al identified miR-744, miR-1186 and miR-466-3p in mouse with high complementarity to the promoter of the Ccnb1 gene, activating its transcription (Huang et al., 2012), and similarly for miR-373 in the homologous gene in human (Place et al., 2008).

Additionally, there is evidence that miRNA-induced silencing complex (miRISC) components functional in the nucleus as well as in the cytoplasm (Robb et al., 2005) and cross-linking immunoprecipitation (HITS-CLIP) data showed evidence of AGO2-mRNA CLIP tags mapped to introns (Chi et al., 2009). All this suggests that AGO1 may be involved in a miRNA-mediated gene regulation in association with histone modifications and not exclusively with siRNA-mediated silencing pathways (Gonzalez et al., 2008). More recently, TNRC6A protein was found to be the AGO navigator protein into the nucleus (Nishi et al., 2013). TNRC6A belongs to GW182 family proteins which are components of miRISC in animals cells (Ding and Han, 2007). Nishi et al suggested that the AGO2 protein is transported into the nucleus by binding with TNRC6A protein in order to perform RNA-mediated silencing in the nucleus (Nishi et al., 2013).

Splicing regulation through non-coding RNAs

There are only few examples of splicing regulation through sRNAs. There is evidence that expression of the C/D box small nucleolar (snoRNA) HBII-52 regulates the usage of the alternative exon Vb of the serotonin receptor 2C (HTR2C) (Kishore and Stamm, 2006b; Kishore et al., 2010). SnoRNAs, are small nuclear RNAs that can be detected in the nucleolus and generally originate from introns (see Appendices). Kishore et al found that the snoRNA HBII-52 can be shortened by exonuclease trimming, leading to smaller RNA variants, the psnoRNAs. These psnoRNAs were found to bind with sequence complementarity to other RNAs, including pre-mRNAs, influencing splice-site selection by competing with existing splicing regulatory factors directly on the pre-mRNA or through association with hnRNPs (Kishore et al., 2010).

Inspired in mammalian TGS, (Allo et al., 2009) showed the first evidence of an exogenous siRNA targeted to the body of a gene affecting alternative splicing in Fibronectin-1 (FN1) gene. Instead of the above described promoter directed siRNA mediated TGS, siRNAs were targeted at the intron downstream of an alternative exon leading strand basepairing with the nascent pre-mRNA, which generated a closed chromatin structure that was able to slow down RNAPII elongation (Allo et al., 2009). The coupling between alternative splicing and transcription was affected due to the chromatin changes. Where a fast, highly processive polymerase favored exon skipping and a slower, less processive polymerase favored the exon inclusion. More recently, both AGO1 and AGO2 were find to act as regulators of co-transcriptional pre-mRNA processing in the CD44 gene model. Earlier evidence, showed accumulation of H3K9me3 in phorbol-12-myristate-13-acetate (PMA) dependent alternative splicing of the CD44 variant exons, resulting in the recruitment of HP1γ, accumulation of RNAPII and recruitment of the spliceosome (Saint-Andre et al., 2011). Further examination in the same gene concluded that these effects rely in activator AGO1 and AGO2 complexes, which act affecting the deposition of chromatin marks (Ameyar-Zazoua et al., 2012). Depletion of AGO1 and AGO2 showed lower efficiency in inclusion of CD44 alternative exons, meaning that Argonaute

proteins can act as a link between the splicing machinery and chromatin modifiers (Ameyar-Zazoua et al., 2012). Moreover, they proposed a mechanism were AGO1 and AGO2 proteins associated with sense sRNAs and an antisense transcript induce heterochromatin marks leading to RNAPII slow down, promoting spliceosome recruitment and affecting alternative splicing (Ameyar-Zazoua et al., 2012). However, recent results on *D. melanogaster* showed that Drosophila AGO2 chromatin occupancy was not correlated with splicing of target transcripts regulated upon AGO2 depletion, hypothesizing that the observed splicing changes due to AGO2 were not performed by siRNA mediated heterochromatin formation (Taliaferro et al., 2013). Drosophila AGO2 was localized mainly at the promoter regions acting as a transcriptional repressor with Polycomb group (PcG) proteins and could act as the mediator of splicing changes, but probably the observed changes in pre-mRNA were due to AGO2 binding to the pre-mRNA and not to chromatin (Taliaferro et al., 2013). On the other hand, there was evidence of snoRNAs binding to Drosophila AGO2 CLIP targets (Taliaferro et al., 2013), which could indicate a mechanism of Drosophila AGO2 sRNA mediated regulation of splicing changes different from (Allo et al., 2009; Ameyar-Zazoua et al., 2012).

CHAPTER 2

Methods

Contents

This section includes the methods used in the articles: (Allo et al., 2009) and two others that at the moment of writing have been submitted for publication. Since the three articles are connected, and in most of the cases the data is the same, it is more convenient to describe first the methods used and then show the results obtained in a separate section.

2.1 Genomic annotations

In many of the results in the next section we use genomic annotation. We retrieved all the information from UCSC tables (Meyer et al., 2013) for the human genome NCBI36/hg18 assembly. We initially used Refseq (Pruitt et al., 2012) gene annotation. In order to recover more gene and transcript information, we later used Ensembl genes, release 54 (Hubbard et al., 2009). Refseq genes are based on cDNA evidence, meaning that a Refseq gene is not always a unique locus. In contrast, Ensembl annotates gene loci. Thus, it can happen that two refseq genes correspond to the same gene locus, or that the same Refseq gene appears in two different loci. In order to avoid possible gene duplicates, we merged RefSeq genes and transcripts that overlap in the same locus, according to exon overlap. From the annotation we selected different genic regions: exons, introns, TSS, pA, gene tails and promoters.

2.2 Predicting miRNA targets

We used Miranda (John et al., 2004) for miRNA target prediction. We chose Miranda (John et al., 2004) as it generally gives a large amount of predicted targets for each query sequence and we can filter them according to different criteria. Miranda provides optimal sequence complementarity, i.e. mature miRNA:mRNA pairs, and a score. The score is a weighted sum of match and mismatch scores for base pairs. The weights are position dependent and reflect the relative importance of 5′ and 3′regions. We choose as initial threshold a minimum alignment score of 150.

In order to remove possible false positives, we performed an empirical p-value calculation for the target predictions. For this calculation, we selected randomly 30 miRNAs from MiRbase Database (Griffiths-Jones et al., 2008). In order to have a control dataset of miRNAs, we shuffled 30 miRNA sequences from MiRbase Database (Griffiths-Jones et al., 2008). The miRNAs generated from shuffling were thus similar in sequence length and composition to the real miRNAs. We used the dataset of intron sequences extracted from hg18 RefSeq genes at UCSC, in total 58692 introns. We predicted miRNA like targets, with minimum score 150, in the 58692 introns dataset from the 30 miRNAs and the 30 shuffled miRNAs. For each of the predicted targets we recovered an score from Miranda (John et al., 2004). We found that for a score of 165, the empirical p-value was 0.03160667 (Table 2.1), i.e. 165 was the minimum score that showed significant differences between the targets predicted with shuffled and real miRNAs. Real miRNAs distribute slightly more widely than random miRNAs (Figure 2.1) (Kolmogorov-Smirnov p-value = 0.005235).

Table 2.1: Empirical p-values for different alignment scores

For each score greater than 150, we calculated an empirical p-value from the distribution of the scores of the targets predicted with the shuffled miRNAs. We found that at a score of 165 or greater this p-value was significant.

score >=	p-value
160	0.06588081
165	0.03160667
170	0.01266513

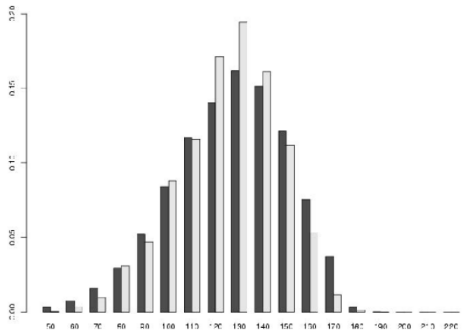


Figure 2.1: Distribution of all posible scores of microRNA targets in a set of 58692 introns with 30 real microRNAs (black) and 30 shuffled microRNAs (grey). We calculated empirical p-values for different scores.

2.3 Alternative event definition

For many of the analysis in this thesis we used alternative splicing events defined using different methods. Initially, we used EST data to define if an exon was always included or not. Subsequently, we add information from RNA-Seq data and for some analysis we also used evidence from Exon Junction arrays. For all these cases, we define as alternative an exon that is not 100% included considering all the evidences, or any combination of them. The inclusion levels are measured in a scale from 0 (totally excluded) to 1 (totally included). We focused in cassette exon events for our calculations.

Alternative exons from EST evidence

We extracted all 185,500 introns from the 19,339 hg18 RefSeq genes from UCSC. We used the available EST data from dbEST (Boguski et al., 1993) to classify the exons from the annotation and considered only exons sup-

ported by more than 10 ESTs. 77507 exons had 100% inclusion level and were classified as constitutive, and the rest, 67802 exons, were classified as alternative. We obtained the two introns flanking each exon that has been classified as alternative or constitutive. Since some of the exons could share the same flanking introns, to avoid repeated signals we selected the 110047 different introns flanking constitutive exons and the 101205 introns flanking alternative exons. Moreover, we used 14843 introns next to exons defined as cassette, which referred to exons that were entirely exon skipped. Finally, the introns were classified according to the splicing properties of the flanked exons.

Afterwards, we added information from Ensembl54 (Hubbard et al., 2009), in order to recover more exons. Using the same procedure described before, we extracted the exons from the annotation and built all possible cassette exon triplets. We retrieved only those that had evidence of more than 10 ESTs and we recalculated the alternative and constitutive events. From now on, we refer as alternative exon a cassette exon event. The (Figure 2.2) shows the distribution of the inclusion values for alternative (cassette exons) events and constitutive events with more than 10 ESTs obtained from Ensembl54 annotation.

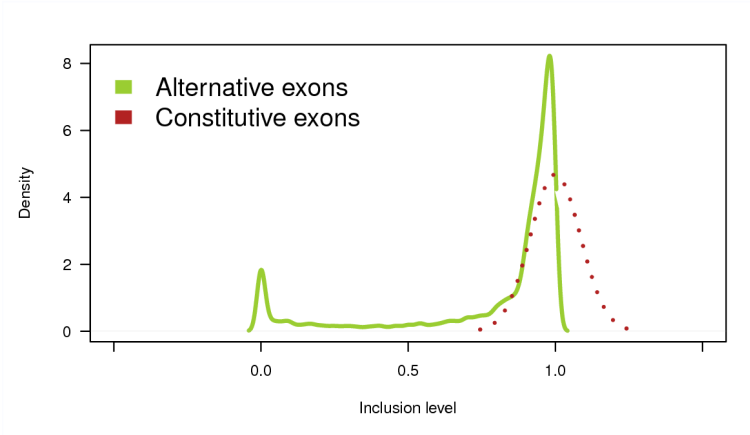


Figure 2.2: Distribution of inclusion levels values for the exons with evidence from more than 10 ESTs, classified as alternative (green) or constitutive (100% inclusion) (red). In this case alternative exons are cassette exons.

2.4 Study of splicing from high-throughput sequencing

Alternative splicing events inference from RNA-Seq data

RNA-seq can be used to quantitatively examine splicing diversity using reads that span splice junctions. RNA-seq reads can span exon-exon junctions, which make them undetectable when using an unspliced alignment method. The first approaches to study splicing with RNA-Seq data were based on custom reference databases of all possible exon-exon junctions, reconstructing event sequences from each exon start and end, on which the reads were mapped (Figure 2.2). Based on (Wang et al., 2008a; Pan et al., 2008) and using different RNA-seq datasets, we developed a method for calculating significant changes in inclusion for cassette exons.

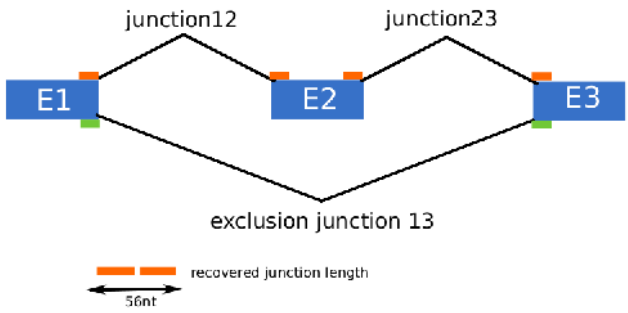


Figure 2.3: We mapped the reads to the set of pre-calculated junction sequences (exclusion junction 13 and the inclusion junctions 12 and 23) of 56nt length, containing the last 28 bases of the upstream exon the first 28 bases from the downstream exon.

Candidate cassette events were calculated considering all combinations

of three adjacent exons (E1, E2, E3) in the Ensembl54 annotation. All the possible exon-exon boundaries (splice junctions) of 56nt were generated from each exon pair, containing the last 28 bases of the upstream exon and the first 28 bases from the downstream exon (Figure 2.3). We used this requirement based on the read length, to ensure that junctions had a minimum of 4 bases in each side of the exon-exon boundary. Reads from the RNA-Seq samples were uniquely mapped to the junctions using Bowtie (Langmead et al., 2009), allowing up to 2 mismatches. The inclusion level (I) of the middle exon was calculated as the fraction of reads that include the exon over the total number of reads that include and skip the exon;

$$I = \frac{\frac{(n_{12}+n_{23})}{2}}{n_{13} + (\frac{n_{12}+n_{23}}{2})} \tag{2.1}$$

Where n_{12}, n_{23} and n_{13} are the number of reads that span the junctions E1E2, E2E3 and E1E3, respectively. In this way, we can measure the inclusion changes between two different conditions (Cond) as the log2-rate of the inclusion levels.

$$M = log_2 \left(\frac{I_{Cond1}}{I_{Cond2}} \right) \tag{2.2}$$

In order to estimate the significance of the inclusion changes, we used Pyicos (Althammer et al., 2011) to calculate a significance score as a function of the total number of reads in the junctions, described by the A value:

$$A = log_2 \left(n_{12Cond1} + n_{23Cond1} + n_{13Cond1} + n_{12Cond2} + n_{23Cond2} + n_{13Cond2} \right) \tag{2.3}$$

The significance of the enrichments was estimated by comparing the enrichment values between the two conditions (Cond) to those obtained by comparing the replicas. When there were no replicas, the relative changes were compared to a null distribution built from the two conditions. The null distribution is inferred from the random rearrangement of the reads from both conditions, taking into account the relative sizes of the original samples. Finally, we selected those events with a significant change in all comparisons,

Benjamini-Hochberg corrected p-value < 0.01. The (Figure 2.4) shows the comparison MCF7 over MCF10A, the upper part in red shows the exons that are significantly included in MCF7 compared to MCF10A. The bottom part in red shows the exons that are significantly skipped in MCF7 compared to MCF10A.

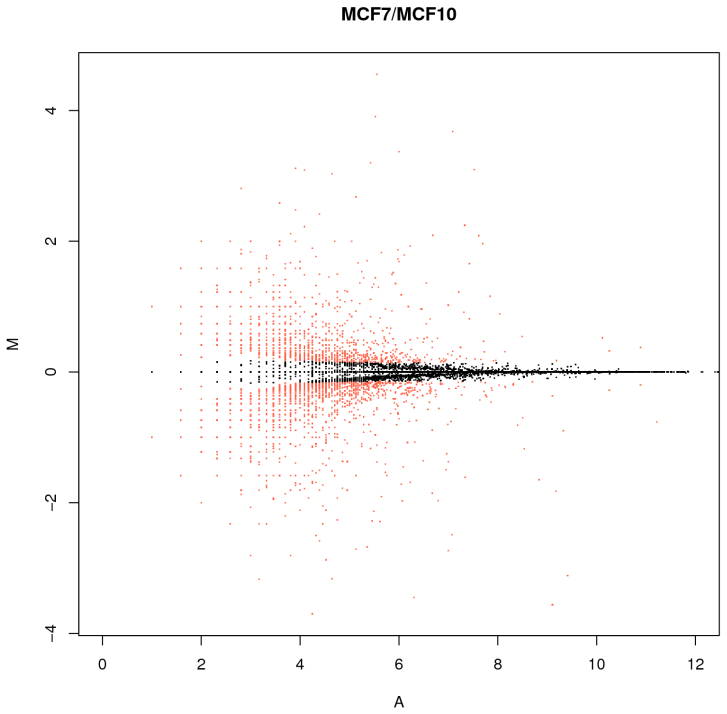


Figure 2.4: Example showing the events changing significantly between MCF7 and MCF10A using our method, RNA-Seq data from (Sun et al., 2011). The figure shows significantly included exons in MCF7 compared to MCF10A, red points in the upper part. In the bottom part, the red points show significantly skipped exons in MCF7 compared to MCF10A. Adjusted p-value < 0.01. In black, the exons without significant inclusion level change.

As a test of our method, we used RNA-seq data from 6 different tissues (brain, cerebral cortex, heart, liver, lung and skeletal muscle) and a set of 5739 ASEs from microarray evidence in the same 6 tissues (Pan et al., 2008). Using the RNA-Seq data, we were able to recover 3515 ASEs, using our exon triplets built from Ensembl54 annotation, which overlapped 60% of the 5739 ASEs from the microarray data (Pan et al., 2008). For each ASE we com-

pared our inclusion values with inclusion values and from (Pan et al., 2008) obtained from RNA-Seq and the microarray experiment (Pan et al., 2008). We found a Pearson correlation coefficient of 0.6 with the microarray (Figure 2.5), similar to what they found (Pan et al., 2008) between the RNA-Seq and the microarray and a Pearson correlation coefficient of 0.97 with the values of the RNA-Seq from (Pan et al., 2008).

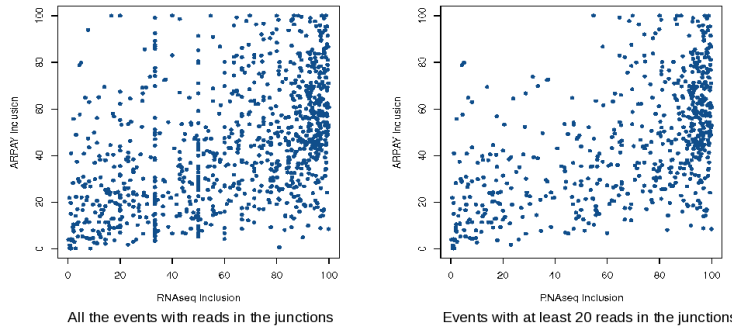


Figure 2.5: Comparison between our inclusion values in the 6 studied tissues and the inclusion from microarray evidence in the same tissues from (Pan et al., 2008). The left plot shows all the cases for which we have evidence of RNA-Seq data and the right plot shows these events with at least 20 reads

Splicing score efficiency

We defined a new splicing score measure, the Splicing Efficiency Score (SES). For each intron, we define s as the number of spliced reads defining the intron and u as the number of reads over exon-intron boundaries (Figure 2.6). To avoid confounding effects from alternative splice sites in the calculation of u, all reads that fall entirely inside any of the annotated exons were first discarded.

SES is thus defined as:

$$SES = \frac{s}{s+u} \tag{2.4}$$

Significant changes in the SES score were calculated analogously to the AS change calculation described above. We kept the introns with an adjusted p-value less than 0.05. The SES is given in a scale from 0 to 1, where 0 is when

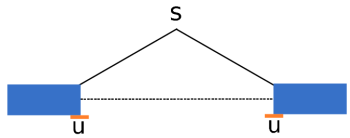


Figure 2.6: Schema showing the reads used for the calculation of the SES. We calculate the fraction of reads that define the intron over the total reads that cross the intron-exon boundaries. We avoid reads that fall entirely in the exons.

the intron is not spliced and 1 when it is totally spliced. The SES represents the efficiency of intron excision from the pre-mRNA.

2.5 ChIP-seq data processing and normalization

We obtained ChIP-seq data in MCF7 and MCF10A cells for AGO1, total H3, H3K36me3, H3K9me2, H3K27me3, HP1α, 5 methylated Cytosine (5metC) and RNAPII. Additionally, we included RNAPII in MCF7 (Welboren et al., 2009) and CTCF in MCF7 and MCF10 (Ross-Innes et al., 2011). As control samples we used ChIP-seq data with a non-specific antibody, for MCF7 and MCF10A. Additionally, we used a specific control sample for HP1α and 5metC. Reads were mapped to the reference genome hg18 using Bowtie-0.12.7 (Langmead et al., 2009) keeping the best unique matches with at most 2 mismatches to the reference (with these parameters -v 2 –best –strata -m 1). All the samples reads were extended to 200nt in the 5′ to 3′ direction using Pyicos (Althammer et al., 2011), apart from AGO1 reads that were extended to 350nt, based on the mean size of the fragments obtained after sonication for each sample. Using BedTools (Quinlan and Hall, 2010) we removed the reads overlapping centromeres, gaps, satellites, pericentromeric regions and the “Duke excluded“ regions, which are regions of low mappability defined by ENCODE (Myers et al., 2011). For each sample, we built clusters with the reads that were overlapping each other on the genomic coordenates using Pyicos (Althammer et al., 2011). We finally discarded singletons, defined as

clusters with only one read.

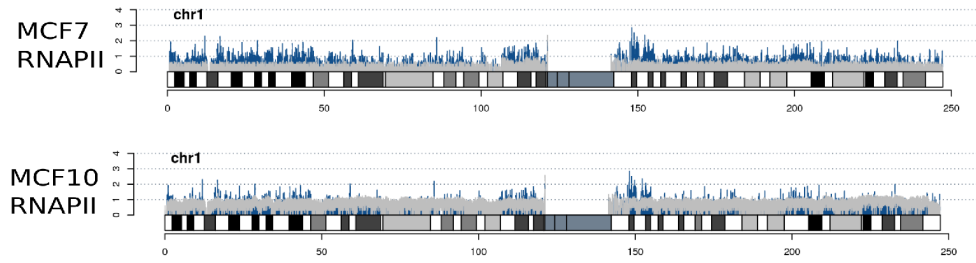


Figure 2.7: The differences in coverage between sample (RNAPII) and control for the two cell lines, RNAPII (blue) over control (grey) in chromosome 1. Y axis shows the read coverage values.

In order to avoid the usage of clusters that are possibly part of the background noise and not of the real ChIP-seq signal, we used the control ChIP-seq samples to identify significant clusters. We observed that the coverage of reads between samples and control can vary in a high degree in each sample (Figure 2.7). In order to perform an accurate normalization, we needed to estimate the level of background reads (Liang and Keles, 2012). We assumed that each ChIP-seq sample is composed of a number of regions with high coverage, which are significant, and a number of non significant regions with low coverage, likely corresponding to the majority of regions. We considered these low coverage regions to be equivalent to the background, so that we expected to find them with similar densities in the control samples (Rozowsky et al., 2009). We selected the overlapping clusters for each sample with control and measured the log_{10} of the number of reads for each cluster, log_{10}(n), and calculated the log-rate as a function of the average read count, similarly to an MA-plot (M log ratios and A mean average) (Figure 2.8 right plot).

Some regions showed clearly that the number of reads was higher in the ChIP-seq samples than in the control (Figure 2.8 right plot). To measure how significant these regions were compared to the background coverage of the control, we considered the cases with few reads in both the sample and the control. In that way, we recovered the differences of the background signal

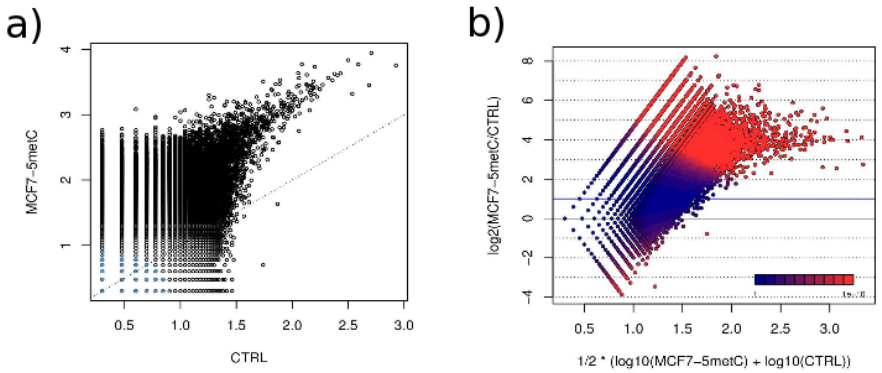


Figure 2.8: ChIP-Seq analysis example given for 5metC in MCF7. a) For the overlapping clusters between control and 5metC we plot the log of the number of reads for each cluster. There is an enrichment of cases in 5metC with more reads than in the control (the points above the diagonal). For some regions 5metC is more enriched, and probably significant compared to the coverage in the same region for the control sample. Blue points correspond to the overlapping regions where sample and control have fewer reads, which we consider to be genomic background regions. b) The blue straight line indicates the mean of the log_2 ratio between sample 5metC and control, calculated from the background regions in the left plot, which provides the ChIP/control normalization factor (CNF). From the CNF we obtain a normalized Bayesian p-value, as the significance of the ratio of the number of reads in the sample 5metC vs the control. p-values follow a color gradient.

between sample and control, and used this relative difference to normalize the signals. We used the mean of the log_2 ratios between overlapping sample and control selected clusters to calculate the CNF (ChIP/control normalization factor) estimated from the overlapped regions with fewer read-tags, assuming that these regions were part of the background noise:

$$CNF = \frac{1}{N}\sum_{i=1}^{N} log_2\left(\frac{Sample_i}{Control_i}\right) \tag{2.5}$$

where N is the total number of overlapping sample and control cases. *Sample* and *Control* are the selected cases of overlapped regions with few reads in both sample and control.

We used the CNF to obtain a normalized Bayesian p-value, as the significance of the ratio of the number of reads in the sample vs the control based

on (Audic and Claverie, 1997):

$$p(y|x) = (CNF)^y \frac{(x+y)!}{x!y!\,(1+CNF)^{(x+y+1)}} \tag{2.6}$$

Where p is the conditional probability inferred from control number of reads (y) in a given region given a number of sample reads (x) in the same region.

For each sample, following the described procedure we chose a threshold to select the significant clusters. This threshold could be based on the number of resulting clusters and the enrichment values relative to the CNF. In order to get the maximum number of significant clusters que used the minimum accepted p-value 0.05 (see results).

2.6 ChIP-seq motif and overlap analysis

To study the significance of the co-occurrence of the different ChIP-seq clusters in specific regions we used the block bootstrap and segmentation method developed in the Encode project (http://encodestatistics.org/). Using a list of genomic regions and two lists of features mapped to them, this method provides a z-score corresponding to the number of standard deviations of the observed overlap compared to the random expected overlap. We ran version 0.8.1 of the script Block Bootstrap and Segmentation (GSC) available at http://www.encodestatistics.org with parameters -r 0.1 -n 10000, where *r* is the fraction the fraction of each region in each sample and *n* is the number of bootstrap samples used. The estimates of significance are conservative because we assumed homogeneity to the regions, thus no segmentation was used. As input for this method, various regions (genome, promoters or intragenic regions) and the blocks of ChiP-Seq clusters that fall within those regions were used. As a positive control, the block bootstrap method was applied to calculate the overlap between significant clusters for HNF4A and CEPBA ChIP-seq datasets (Schmidt et al., 2010) in promoter regions, defined to be 500 upstream of the TSS.

The following motif analysis was carried out independently for CTCF, AGO1 and HP1α clusters: Given a sample set *S* of *N* sequences and a control set $S^{(0)}$ of $N^{(0)}$ sequences, the number of times $n_{a,i}$ that each 7-mer *a* appeared in each sequence i was calculated. Likewise, for the control set, the number of occurrences $n_{a,i}^{(0)}$ of each 7-mer *a* per sequence *i*, was also calculated. The expected density $d_a^{(0)}$ of each 7-mer was calculated as the ratio between the total number of occurrences in the control set over the total sequence length of the control set:

$$d_a^{(0)} = \frac{\sum_{i \in S^{(0)}} n_{a,i}^{(0)}}{\sum_{i \in S^{(0)}} l_i} \tag{2.7}$$

where l_i is the length of each sequence in the control set. For each sequence i in the sample set, and each 7-mer a, it was recorded whether the observed 7-mer count ($n_{a,i}$) was greater than the calculated expected count ($d_a^{(0)} l_i$):

$$\delta_{i,a} = 1 \quad \text{if} \quad n_{a,i} > d_a^{(0)} l_i \tag{2.8}$$

$$\delta_{i,a} = 0 \quad \text{otherwise} \tag{2.9}$$

Similarly, for the counts in the control set:

$$\delta_{i,a}^{(0)} = 1 \quad \text{if} \quad n_{a,i}^{(0)} > d_a^{(0)} l_i \tag{2.10}$$

$$\delta_{i,a}^{(0)} = 0 \quad \text{otherwise} \tag{2.11}$$

The sum of the $\delta_{i,a}$ values over the sequences i represent the number of sequences for which the 7-mer a has an observed count greater than expected. Thus, for each 7-mer, the odds-ratio (7-mer score) and corresponding p-value were obtained by performing a Fisher test (one-tailed) with these sums for the sample set and the control set:

7-mer a	more than expected	less than expected
S	$\sum_{i \in S} \delta_{i,a}$	$N - \sum_{i \in S} \delta_{i,a}$
$S^{(0)}$	$\sum_{i \in S^{(0)}} \delta_{a,i}^{(0)}$	$N^{(0)} - \sum_{i \in S^{(0)}} \delta_{a,i}^{(0)}$

In order to build the consensus motifs, a procedure similar to (Schmidt et al., 2010) was carried out. First, the top 2500 genome-wide clusters according to mean cluster height from each set were considered, selecting only the clusters with significant 7-mers. For each of the clusters, the sequence of 200 bp centered on the middle position of the cluster was extracted to run MEME (Bailey and Elkan, 1994) with options dna revcomp zoops.

2.7 Predictive models: Accuracy Testing and Attribute Selection

In the last part of the results section we apply a Machine Learning approach to build a model to classify skipping and inclusion events. In order to select the most important attributes for the classification, we used a combination of attribute selection methods.

First, to build the predictive model 15 regions around each exon triplet (E1, E2, E3) were considered (see results). For each region and sample, the read density was calculated using the RPKM. The relative difference of ChIP-seq signal for region *a* between the two cell-lines was calculated as:

$$M = log_2 \left(\frac{d_{(a)_{MCF7}}}{d_{(a)_{MCF10A}}} \right) \tag{2.12}$$

Where $d_{(a)}$ is the density in RPKM. The significance of the changes was calculated using Pyicos (Althammer et al., 2011), as a function of the average densities in each region:

$$A = \frac{1}{2} \left(log_2 \left(d_{(a)_{MCF7}} \right) + log_2 \left(d_{(a)_{MCF10A}} \right) \right) \tag{2.13}$$

An attribute for the classification corresponds to the vector of z-scores for the relative enrichment for each sample-region pair.

Three attribute selection methods were applied: Wrapper Subset Evaluator (WSE), Correlation Feature Selection (CFS) and Information Gain (IG). IG is defined as the expected reduction in entropy caused by partitioning the examples according to one attribute, thus the higher the IG value, the better the attribute can separate skipping and inclusion classes. On the other hand, CFS works by iteratively testing subsets of attributes, retaining those that best correlate with the class values (inclusion and skipping), removing those that have high redundancy between them. In WSE, subsets of attributes are tested iteratively using a 10-fold cross validation and the space of all possible subsets is explored heuristically, such that only those subsets that perform above an optimal threshold are scored as informative. The WSE

method gives thus the frequency at which each attribute is selected in the optimization procedure. For WSE we used a Genetic Search algorithm to explore different combinations of attributes and an ADTree to evaluate the attributes. Repeated runs of WSE did not change the resulting top attributes. We selected attributes that showed WSE and CFS frequencies $>= 50\%$, and a position in the IG ranking in the top 50%. Some of these attributes corresponded to overlapping regions. From these we selected the ones with highest IG. In this way, a minimal set of non-redundant attributes with optimal performance is selected.

For each of the models the accuracy was calculated using 10-fold cross validation. In this procedure, the datasets are split into training and testing sets in 10 different ways. Testing sets were chosen such that each event is predicted just once. The accuracy was measured as the average value of the sensitivity and specificity over all 10 splits. We also reported the number of events, either skipped or included, that were correctly predicted by the model. Each model was built with a given number of attributes, for our initial model we used 135 attributes. Since our attributes were expected to be dependent, we applied two different classifiers that were based on dependencies to build the model and test the predictive power of our attributes: a Bayesian Network (BN) and an alternate decision tree (ADTree). A BN consists in the combination of conditional probabilities between attributes to define a network, where each attribute has a probability distribution given by the conditional probability on one or more parent attributes. ADTree is a classification method based on binary decision trees, using a voting system to combine the output of individual tree models. Each individual model has a tree structure, where each node of the tree represents a binary partition. At every partition a test is performed for every attribute and the test set that maximizes the entropy based gain ratio is selected, leading to a tree where every leaf contains instances from one class when there is no over-fitting. Individual trees are combined into a single tree using a voting system to weight the contribution from the multiple binary tests into a final classification, which is represented in the leaves. The ADTree has been shown before to be a good learning algorithm for genetic regulatory response (Middendorf et al., 2004).

CHAPTER 3

Results

Contents

This section includes the results from three articles: (Allo et al., 2009) and two other that at the moment of writing have been submitted for publication. Each of the subsections refers to the results from each of the articles. Together the three articles, conform a global framework for alternative splicing regulation by chromatin and with a special focus on AGO1 and RNA mediated transcriptional gene silencing.

3.1 siRNA mediated transcriptional gene silencing affects alternative splicing

In this section I describe the first evidence for a siRNA mediated transcriptional gene silencing (TGS) alter the inclusion of an alternative exon. These results were published in (Allo et al., 2009), where I performed the bioinformatics analysis. Experiments showed that transfection of mammalian cells with siRNAs targeting a gene region located near the alternative FN EDI exon can affect alternative splicing of that alternative exon. Previously, the FN1 EDI exon was found to be regulated by transcriptional elongation (de la Mata et al., 2003). Several experimental procedures support a transcriptional gene silencing (TGS) mechanism leading to the slow down of RNAPII elongation rate. In this example, postranscriptional gene silencing (PTGS) was not responsible for the alternative splicing effect and the transfection of the siRNAs caused H3K9me2 and H3K27me3 marks. Moreover, the experiments showed that the effect depends on the presence of AGO1, the Argonaute protein previously shown to be necessary for TGS. Furthermore, the effect is reduced or abolished by treatments that promote chromatin relaxation or increase RNAPII elongation, and siRNA transfection causes H3K9 dimethylation and H3K27 trimethylation at the target gene. It is notable that the silencing marks are in the vicinity of the siRNA target sites, inside the FN1 gene, and not on its promoter. This was the first evidence of siRNAs mediating TGS targeting in genebody regions. Based on these evidences, the transcriptional gene silenced alternative splicing pathway (TGS-AS) was proposed (Allo et al., 2009).

Experimental evidence

The experimental results performed by Allo et al (Allo et al., 2009) showed that there was a significant increase of the inclusion over exclusion ratio of the EDI exon in HeLa and Hep3B cells when siRNAs were directed to the exon vicinity regions but the total FN1 mature mRNA levels were not affected in both, HeLa and Hep3B cells. They found that the antisense strand of the transfected siRNAs directed TGS and heterochromatin formation. Two histone marks, H3K9me2 and H3K27me3, appeared as a consequence of the transfection and colocalized in the vicinity of siRNA targets sites. Finally, to validate the importance of AGO proteins in the process, they performed AGO1 and AGO2 knockdown, which abolished the effect of the siRNA on splicing. This result confirmed the involvement of AGO proteins in TGS pathway.

siRNAs directed to intronic regions

First evidences of siRNA mediated TGS indicated that siRNAs are directed to the promoter regions with high complementarity (Morris et al., 2004; Kim et al., 2006; Janowski et al., 2006). In order to find the similar effect of a siRNA but in the gene regions near alternative exons, we focused in the intronic regions flanking alternative cassette exons. First, measures of alternative exon inclusion were done with EST data from dbEST. Although we were able to select exons with enough evidence, in some cases these data is not enough to detect all possible splicing changes. Splicing inference methods rely currently on HTS (RNAseq) data, which provides a higher coverage in alternative splicing changes.

We used our dataset of alternative exons classified from EST evidence (see Methods for alternative exon definition). We downloaded the dataset of 294,058 transfrags from HeLa cells from (Kapranov et al., 2007) and calculated the overlap of the genomic coordinates with our intron data set using fjoin (Richardson, 2006). We found that 43% of the transfrags were completely included in introns.

Using different sRNA databases we checked the overlap between different endogenous sRNAs and intronic sequences. We found that 137,108 of 294,058 endogenous short transcribed RNAs previously identified in a HeLa cell database (Kapranov et al., 2007) overlap intronic regions, of which 37,727 flank ASEs and 9,713 flank cassette exons, such as our model EDI. Moreover, using the UCSC database (http://genome.ucsc.edu; wgRNA table from sno/miRNA track) we found that 289 miRNA precursors, from a total of 689 miRNAs are located in introns, of which 110 are flanking ASEs. This indicates that endogenous miRNAs can be located in introns flanking alternative exons, such the case of the EDI with exogenous siRNAs.

TGS affects ASEs and needs AGO1

AGO1 and AGO2 are required for PTGS but the TGS pathway only needs AGO1 (Janowski et al., 2006). Additionally, some sRNAs acting in the TGS depend on DCR and not on AGO1. Based on these two assumptions, we selected a list of 96 cancer related ASEs from an automated high-throughtput RT-PCR pannel for Hep3B and HeLa cells (Klinck et al., 2008), (Figure 6.1) and (Figure 6.2), which were transfected with siLuc, siAGO1 and with siDCR. 91 of the 96 events tested had evidence of skipping from EST data. We found that 35% of the ASEs in Hep3B and 38% of the ASEs in HeLa cells were affected by AGO1 depletion, providing evidence for TGS-AS.

Since we were interested in an AGO1 dependent mechanism, we continued to analize in more detail these ASEs (Figure 3.1). We found that 53% in Hep3B and 51% in Hela of the ASEs showed no change upon DCR or AGO1 knockdowns, meaning that AGO1 dependent changes on alternative splicing may be exclusive of a certain number of cases. On the other hand, 18% of the ASEs in Hep3B and 27% in HeLa were affected only by AGO1 depletion. These ASEs that are DCR independent, can be involved in pathways exclusive to TGS. Finally, we found that 12% of the ASEs in both HeLa and Hep3B were affected only by DCR depletion (Figure 3.1). We hypothesized that these ASEs may be involved in regulation of alternative splicing by PTGS pathway, in a process independent of AGO1.

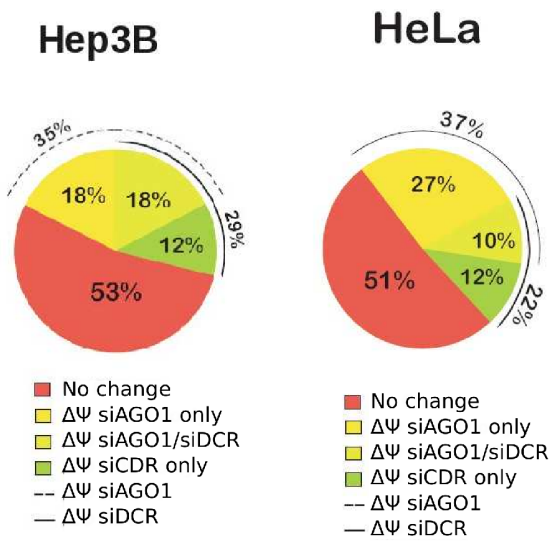


Figure 3.1: Automated high-throughput RT-PCR platform to screen 96 cancer-related ASEs. Distribution of DC (percent of genes from the panel) significantly affected (green, dark and light yellow) or unaffected (red) by AGO1 and/or DCR knockdown are shown for Hep3B and HeLa. $\Delta\psi$ Inclusion change.

In order to add more information and to look for possible dependencies or specifities due to the intron exon structure, we compared the lengths of the flanking introns for each of the ASEs. First, for all length distributions we performed a normality test (Shapiro-Wilk test) and calculated a p-value for each comparison using the Wilcoxon rank-sum test. We found statistically significant differences for the upstream introns (Figure 3.2), when we compared ASEs affected by siAGO1 knockdown and ASEs that were not affected. The AGO1 dependent ASEs showed shorter upstream introns (HeLa p-value=0.0268 and HepB3 p-value=0.0201). If we considered both, HeLa and Hep3B ASEs that were affected upon AGO1 depletion together, upstream introns were also shorter than downstream ones, p-value= 0.0064 (first boxplot in (Figure 3.2). However, we didn't find significant differences in the downstream intron lengths for any of the comparisons, second boxplot in (Figure 3.2). When comparing the ratios between the lengths of upstream and downstream introns we found significant differences between

ASEs affected by AGO1 depletion and the ones that were not affected, but in this case only for Hep3B ASEs (p-value = 0.0290), third boxplot (Figure 3.2). These results showed that ASEs dependent of AGO1 share common structural characteristics.

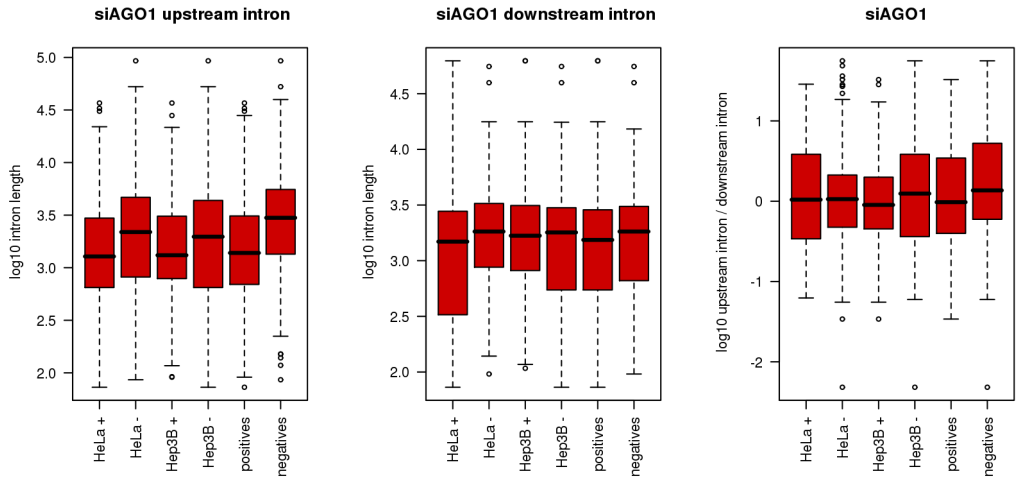


Figure 3.2: Left panel, upstream intron lengths. Middle panel, downstream intron lengths. Right panel, log ratio between upstream and downstream intron lengths. Upstream introns are shorter than downstream ones, p-value=0.0064, when ASEs are affected by siAGO1 knockdown. We did not find significant differences in downstream introns.

Then, we followed the same approach with DCR dependent ASEs for both Hep3B and HeLa. We found significant differences in HepB3 ASEs in downstream introns when we compared the introns from ASEs that were affected with siDCR knockdown and introns from ASEs that were not affected by siDCR knockdown, p-value = 0.00241. ASEs affected by DCR depletion showed similar characteristics to AGO1 dependent introns (Figure 3.3).

In general we found that ASEs affected by AGO1 depletion have shorter upstream introns. We hypothesized that the downstream intron may favor excision of the upstream intron by epigenetic marks that stall the RNAPII and/or by recruitment of splicing factors. For example, C11orf17 was affected by AGO1 depletion but not DCR depletion in Hep3B cells (Figure 3.4).

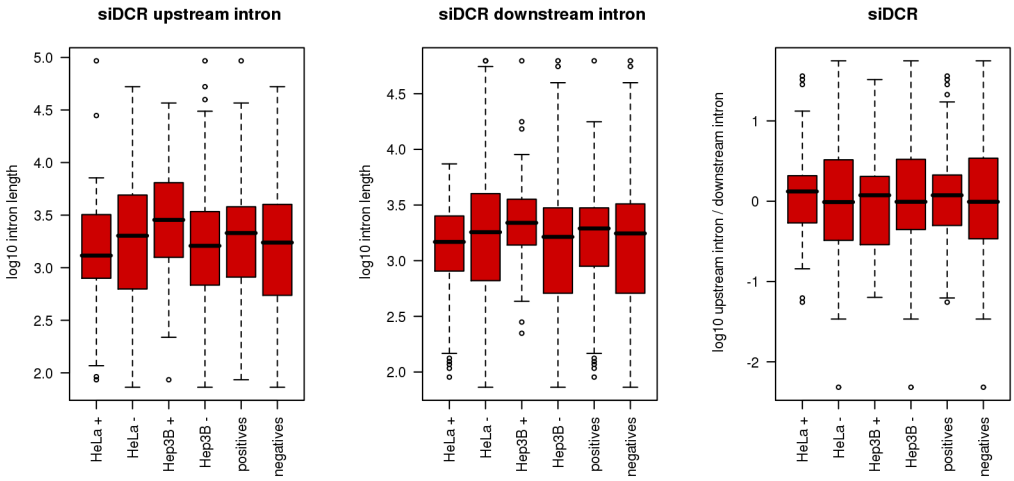


Figure 3.3: Left panel, upstream intron lengths. Middle panel, downstream intron lengths. Right panel, log ratio between upstream and downstream intron lengths. Only downstream introns from ASEs in Hep3B showed significant differences.

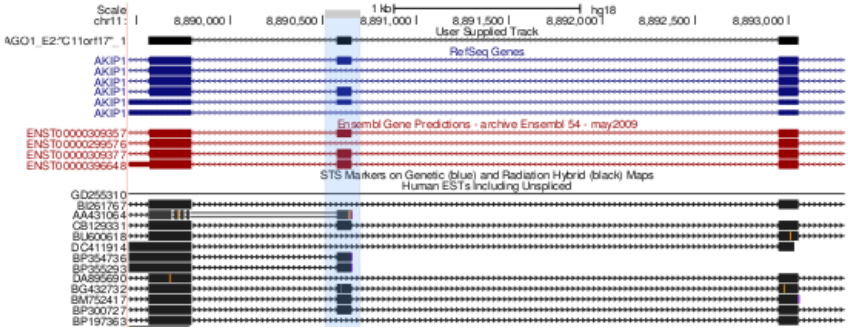


Figure 3.4: UCSC Genome Browser image, hg18. In the figure, Refseq gene track, ENSEMBL genes track and human ESTs track from UCSC are shown for a fragment of the gene C11orf17, which is in the minus strand and corresponds to a cancer-related event tested in the PCR platform from primers and that we defined as alternative exon based on EST data. C11orf17 ASE changed under AGO1 depletion in Hep3B cells and not under DCR depletion. We considered it as AGO1 dependet ASE.

On the other hand, ASEs affected by DCR depletion have longer downstream introns. For example, FANCA exon was affected by DCR depletion but not by AGO1 depletion in Hep3B cells (Figure 3.5).

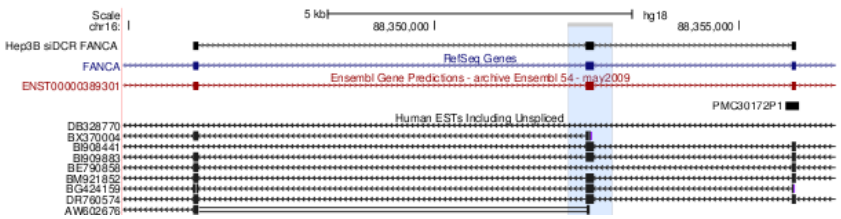


Figure 3.5: UCSC Genome Browser image, hg18. In the figure, Refseq gene track, ENSEMBL genes and human ESTs tracks from UCSC are shown for a fragment of the gene FANCA, which is in the minus strand and corresponds to a cancer-related event tested in the PCR platform. The exon is defined as alternative exon based on EST data. FANCA ASE changed under DCR depletion in Hep3B cells and not under AGO1 depletion.

Endogenous sRNA target prediction near Alternative exons

We focused on the events defined from EST data and the ASEs dependent on AGO1 from the previous section for the prediction of small RNA (sRNA) targets. We hypothesized that other endogenous sRNAs, like miRNAs or endogenous siRNAs, can regulate alternative splicing when they are directed to regions in the vicinity of an ASE. If an endogenous sRNA is affecting the regulation of an alternative exon we expect to find sRNA targets in the intron flanking the alternative exon. We used different sRNA databases. For more information on sRNAs databases resources see appendices (Agirre and Eyras, 2011) and methods chapter. Additionally, we used Miranda (MicroRNA Target Detection Software) (John et al., 2004) (details in methods chapter) with all the mature miRNA entries from MirBase 13.0 (Griffiths-Jones et al., 2008) and the Argonaute database (Shahi et al., 2006) and the events as target regions.

Initially, we selected as target regions the introns flanking 67802 alternative exons and 77507 constitutive exons from hg18 RefSeq genes from UCSC. We performed an exhaustive search of miRNA targets in the introns using Miranda (John et al., 2004) (see Methods). We extracted the 2323 mature miRNAS from MiRbase Database (Griffiths-Jones et al., 2008) and 1834 mature microRNA from Argonaute Database (Shahi et al., 2006) to perform the target search. For this set, we found higher number of targets in introns

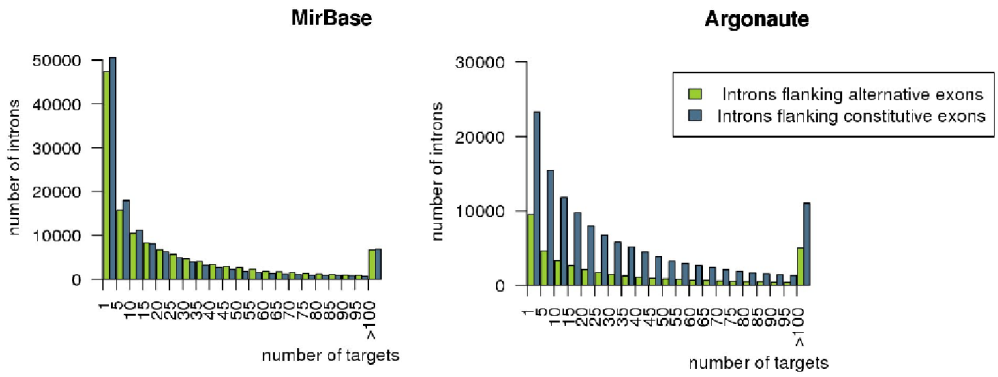


Figure 3.6: The cumulative distribution of microRNAs with different targets, minimun score 165, MirBase database (left) and Argonaute database (rigth), respectively.

flanking constitutive exons. However, the number of miRNAs from MiRbase with targets was similar (Figure 3.6). The results showed that miRNAs can target introns flanking alternative exons but also introns flanking constitutive exons.

The differences between constitutive and alternative exons can lie in the structural characteristics of the intron sequences or in others elements, like the presence of AGO1. Considering the possibility that the TGS-AS involves AGO1, as described previously (Allo et al., 2009), we performed the same target prediction approach using the cancer related ASEs (Klinck et al., 2008) that were tested upon depletion of AGO1 in HeLa and Hep3B. We extracted the flanking introns sequences for all the ASEs and run MiRanda with the same parameters, and the same two miRNA databases as before. We found that the mean proportion of events with targets was higher in the ASEs dependent on AGO1 specifically in Hep3B (Figure 3.7). miRNAs that only target ASEs affected by AGO1 depletion are candidates for TGS-AS, and miRNAs that only had targets in ASEs that were not affected upon AGO1 depletion are candidates for a post-transcriptionally regulated alternative splicing.

Tables with all the predicted targets in ASEs dependent on AGO1 in HeLa and Hep3B; Tables (6.1), (6.2) ,(6.3) and (6.4).

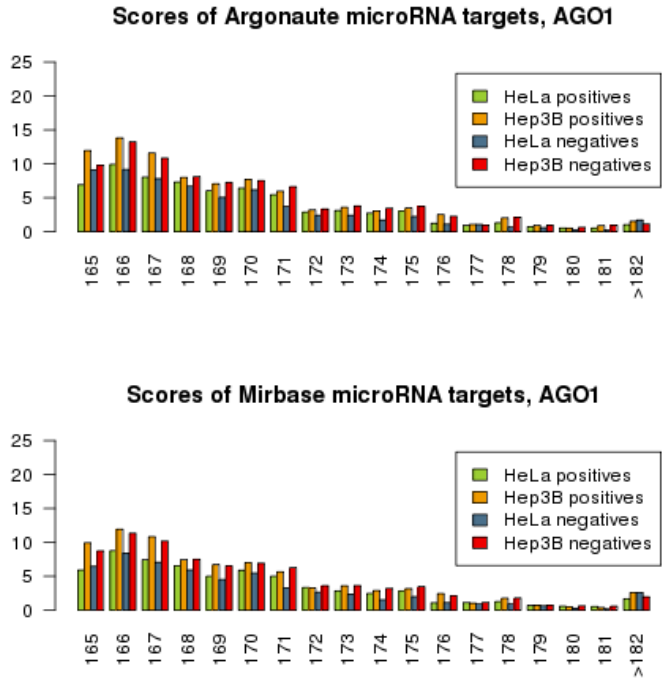


Figure 3.7: Proportion of each score value in the two cell lines, HeLa and Hep3B, for the Argonaute database microRNA targets and MirBase database microRNA targets. ASEs affected by AGO1 knockdown (positives) showed higher number of targets than ASEs that were not dependent on AGO1 in Hep3B (negatives).

3.2 Genome-wide analysis of AGO1 and its role in alternative splicing

The results of this section are part of the analysis performed for a second publication, that was under submission at the time this thesis was being written . Here I'll show the results of the analyses I performed about AGO1 distribution genome-wide and its possible role in alternative splicing.

AGO1 distribution in relation to sRNAs

Apart from miRNAs and siRNAs in the last years a myriad of new classes of sRNAs had been described . It is very reasonable to think that the endogenous sRNA classes that can act in the TGS-AS are not restricted to be miRNAs or siRNAs. Other small RNAS, like piRNAs, affect chromatin structure independent from DCR (Klattenhoff and Theurkauf, 2008). If a piRNA triggers heterochromatin formation close to an ASE, it will regulate alternative splicing in a DCR independent pathway. piRNAs are known to act mainly in germ line and databases of human piRNAs are still not very common. Taking into account these restrictions, we focused our analysis on other classes of sRNAs. New classes of sRNAs have been recently found, like transcription initiation RNAs (tiRNAs) (Taft et al., 2009) and promoter associated RNAs (pasRNAs) (Fejes-Toth et al., 2009). For a detailed description of these sRNA classes, see appendices (Agirre and Eyras, 2011). The biogenesis of these sRNAs is mainly related with TSS and promoter regions, but their targeting mechanism is still unknown.

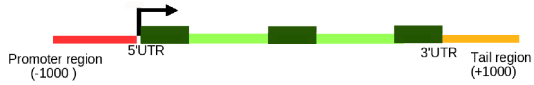


Figure 3.8: Defined regions: exons, introns, tails (defined as the region 1kb downstream of the pA) and promoters (defined as the 1kb region upstream of TSS).

First we looked at the distribution of sRNAs along genes using these sRNAs and sRNA-Seq data from MCF7 (Mayr and Bartel, 2009). We used hg18 Ensembl54 annotation for the different regions: introns, exons, gene tails (defined as the region 1kb downstream the polyA-site) and promoters (defined as the 1kb region upstream the TSS) (Figure 3.8). We overlapped the sRNAs, tiRNAs and pasRNAs using fjoin (Richardson, 2006) to the different genic regions. We normalized the proportions by the region length and removed the first exons, to avoid overlapping signals from the TSS and the exons.

We found, as expected, that both tiRNAs and pasRNAs have an enrichment in promoters compared to the rest of regions (Figure 3.9). These sRNAs

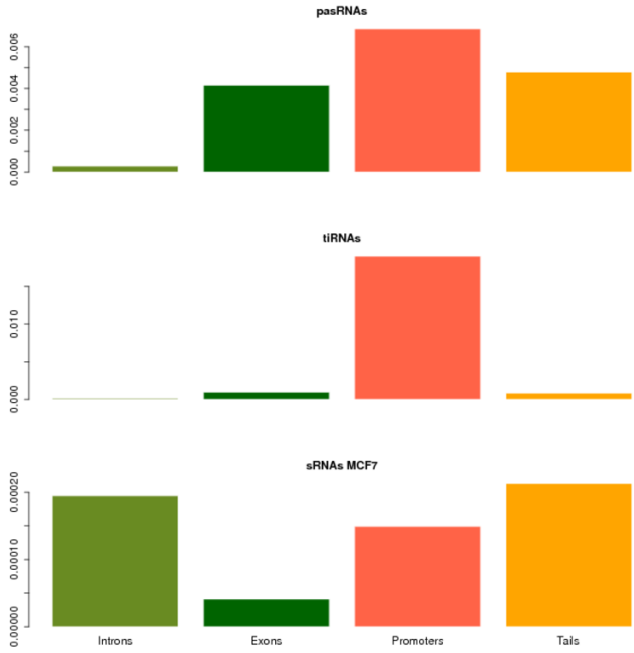


Figure 3.9: pasRNAs (top), tiRNAs (middle), MCF7 sRNAs (bottom). We found higher proportions of tiRNAs and pasRNAs in promoters regions while sRNAs from MCF7 are more related to intron and tail. Proportions were normalized by region length.

are known to be generated from the backtracking of the RNAPII in the TSS in the case of tiRNAs and from promoter regions in the case of pasRNAs. It is known that Argonaute proteins load sRNAs, which are then transported to their target regions (Meister et al., 2004). Since we found evidence of AGO1 dependent ASEs (Allo et al., 2009) and on the involvement of Argonautes in all the sRNA pathways (Hock and Meister, 2008), we looked for the distribution of AGO1 clusters in general and in relation to the origin site of the snRNA. We normalized the proportion of clusters by the gene regions, in order to remove biases on the different regions lengths. We also removed first exons, to avoid overlapping signals. We found that AGO1 co-localize with sRNA enrichment in intragenic regions, specially introns (Figure 3.10). When we calculated the proportions of AGO1 clusters that overlap with sRNAs, we found that the highest enrichment was between pasRNAs and AGO1 and specially in intronic and tail regions (Figure 3.10).

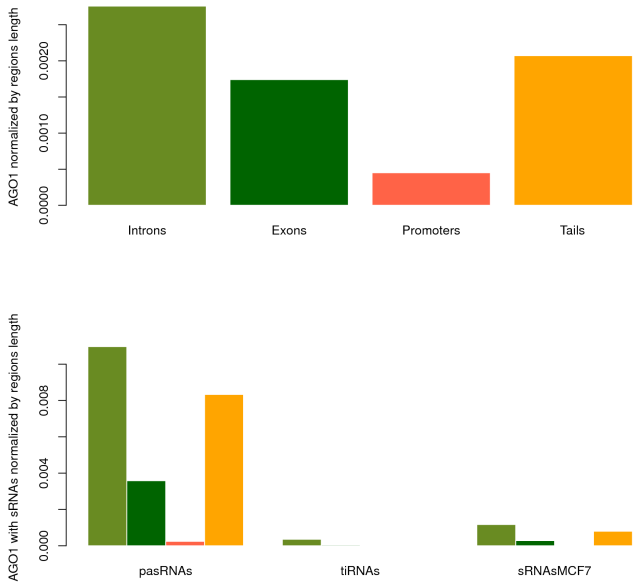


Figure 3.10: Top barplot: Proportion of AGO1 clusters that overlap the different genic regions. Bottom barplot: Proportion of tiRNAs, pasRNAs and MCF7 sRNAs overlapping the AGO1 clusters. Proportions were normalized by region length.

AGO1 genome-wide

In view that AGO1 can be related to sRNAs in intragenic regions, we decided to perform a more exhaustive analysis of the distribution of AGO1.

First, we divided genes into 3 groups according to the expression level in MCF7 cells (high, medium and low), using published RNA-seq data (Sun et al., 2011). We found an enrichment of AGO1 in the TSS in MCF7 for highly expressed genes, whereas there was no enrichment for lowly expressed genes or for the same gene using a randomized set of AGO1 reads (Figure 3.11). Additionally, we looked at RNAPII densities on the TSS in relation to AGO1 using data from the same publication (Sun et al., 2011) (Figure 3.12). We found a peak of RNAPII where there was AGO1. In order to assure this possible enrichment on the TSS of AGO1 related to RNAPII, we analyzed the distribution of AGO1 in all the genic regions comparing observed vs ex-

pected. We found that there is a positive enrichment (red) in CpG islands (CGI) and 5'UTRs, while there is a depletion in exons, gene tail, introns and intergenic regions (blue) (Figure 3.13 a). We performed the same analysis with AGO1 in MCF10A cells and found the same enrichments except for CGIs (Figure 3.13 *a*). We hypothesized that AGO1 may be enriched in 5'UTR or near the first 5'ss rather than on the TSSs. In fact, we found that there is a significant enrichment of AGO1 primarilly in the first 300nt of first introns (Figure 3.13 a). This enrichment was also present in first introns for AGO1 in MCF10A (Figure 3.13 *b*). We compared the densities of AGO1, calculated as RPKM, on the first introns according to expression levels on MCF7 RNA-Seq and we found that highly expressed genes showed higher densities of AGO1 in the first intron compared to lowly expressed genes (Figure 3.15 *a*).

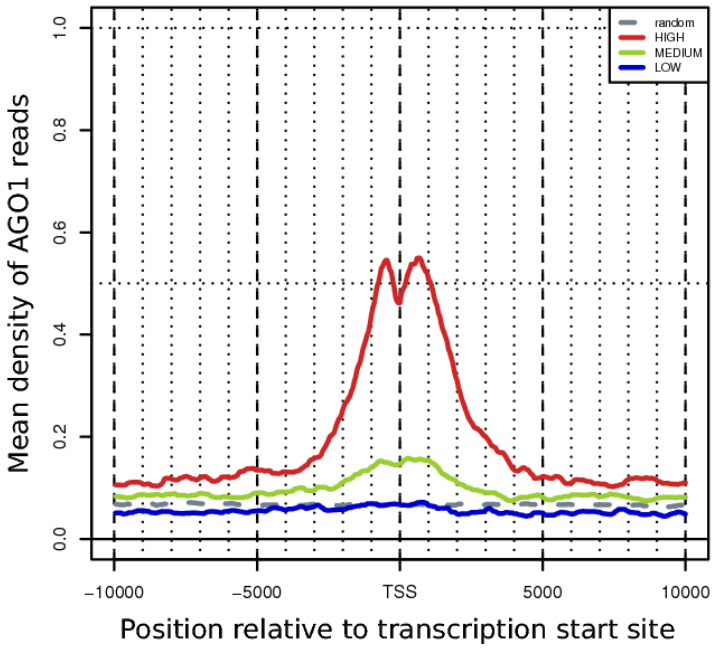


Figure 3.11: Mean density of AGO1 MCF7 reads on the TSS, according to high, medium and low gene expression in MCF7. In grey the densities of random clusters on the TSS. Random clusters are calculated as regions from the same length, not overlapping real clusters and excluding conflictive regions (gaps, centromeres, satellites...)

The MCF7 sRNA library (Persson et al., 2009) showed an overlap of AGO1

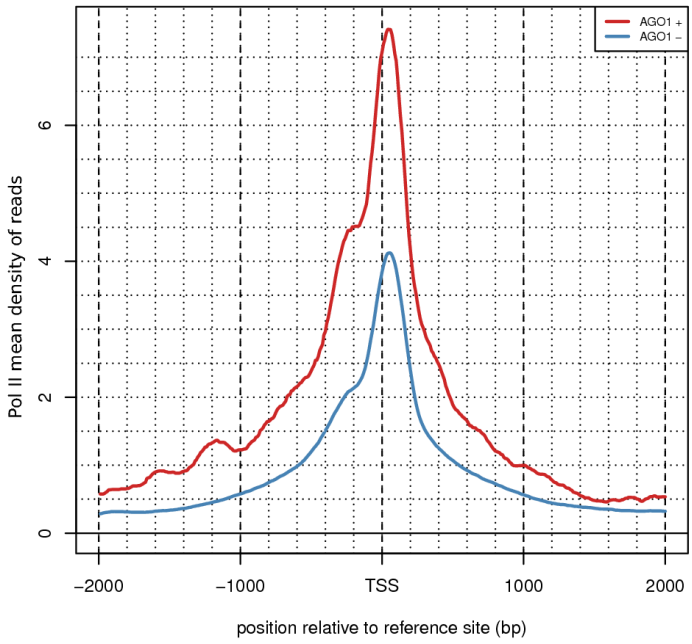


Figure 3.12: PolII (RNAPII) mean density of reads on the TSS when there is AGO1 (red) and without AGO1 (blue) in the same region.

clusters with significantly higher than that of randomized clusters (Figure 3.15 *b*). Similar results were found using sRNA data collected from HeLa, HepG2 and Gm12878 cells from the ENCODE project (Fejes-Toth et al., 2009; Myers et al., 2011), see (Figure 3.15 *b* and *c*). Moreover, the percentage of canonical promoters with sRNAs containing AGO1 clusters is also higher than that of promoters with small RNAs but without AGO1 clusters in the four cell lines (Figure 3.15 *c*). There is evidence of a genome-wide relation between alternative splicing and antisense transcription (Enerly et al., 2005) and Allo et al found an antisense EST that covered the affected region of the EDI exon (Allo et al., 2009). In this sense, we found that the sRNAs from MCF7 library that overlapped with AGO1, apart from overlapping in a higher proportion in introns (Figure 3.10), the number of antisense sRNAs was also higher in introns compared to other regions. Moreover, most of the antisense sRNAs from this library in introns were overlapping the first intron, there were also notable number of antisense sRNAs in 5'UTRs and

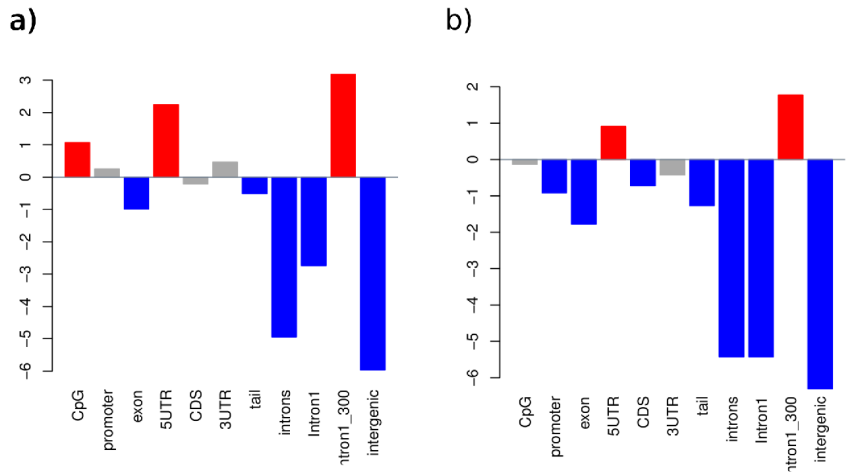


Figure 3.13: AGO1 enrichment in different genic regions. a) AGO1 MCF7, b) AGO1 MCF10A. The barplots show regions that are positively enriched (red), regions negatively enriched (blue) and regions without enrichment (grey)

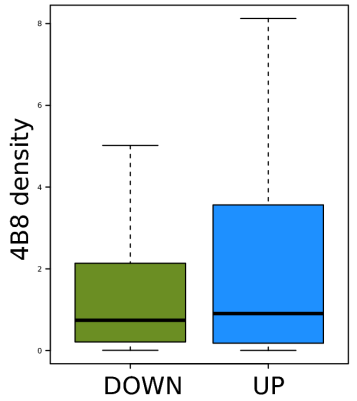


Figure 3.14: AGO1 density in RPKM (4B8 density) over the gene bodies of genes downregulated (DOWN) and upregulated (UP) upon AGO1 depletion.

promoter regions.

We next focused on the overlap between AGO1 and histone marks. At genome wide level (Figure 3.15 *d* left) the overlap between pairs of histone marks was lower than the overlap at AGO1 target sites. In fact there is an increase in the enrichment of H3K9me2 and H3K27me3 in AGO1 targets

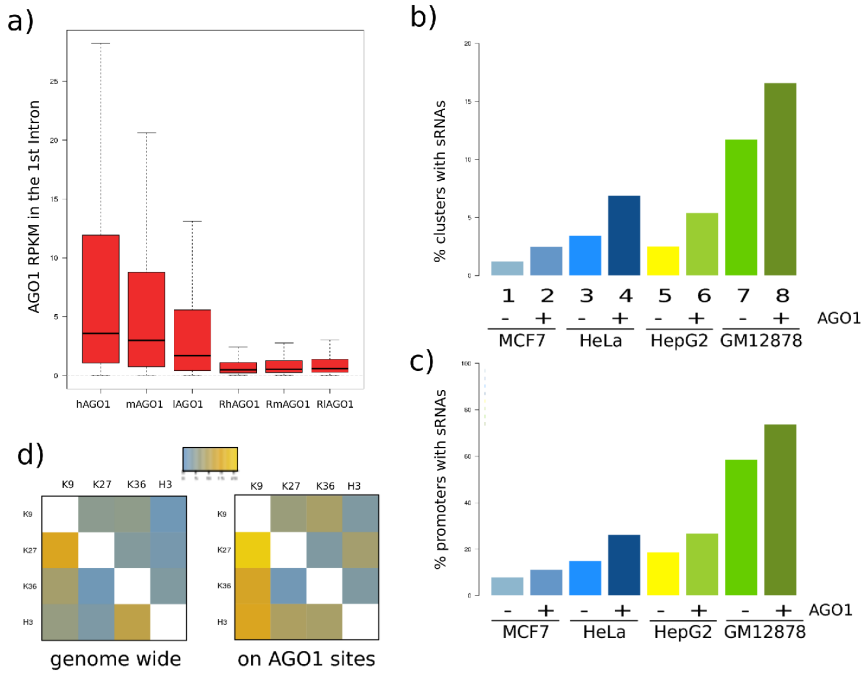


Figure 3.15: a) AGO1 MCF7 densities in first introns for genes of high (hAGO1), medium (mAGO1) and low (lAGO1) expression according to MCF7 RNA-Seq data. We compare them to the densities of random AGO1 reads in the same genes (RhAGO1, RmAGO1, RlAGO1). b) Proportions of AGO1 (+) clusters overlapping with sRNAs and proportions of random AGO1 clusters (-) overlapping with sRNAs, using the four different sRNAs datasets. c) Proportions of promoters (2kb upstream from the TSS), overlapping with sRNAs and AGO1 clusters (+) and overlapping with sRNAs but not with AGO1 clusters (-). d) Left heatmap: Genome-wide enrichment values of the overlaps between histone modifications clusters. Right heatmap: Enrichments of the overlap between histone modifications on AGO1 clusters sites. Values correspond to the proportion of sites for each row that overlaps with the signal in each column.

sites (Figure 3.15 *d* right), but also in total H3 and H3K36me3, suggesting a possibly role between AGO1 and histone modifications, primarily silencing marks.

The association of AGO1 to TSS of highly expressed genes, see (Figure 3.11), may imply a role for AGO1 in transcriptional regulation beyond, and perhaps opposite to, TGS. To explore this hypothesis, we performed analyses

with poly(A)+ RNAs (RNA-seq) in MCF7 cells treated with siAGO1, using siLuc as control. Upon AGO1 depletion, expression of 813 genes was down-regulated (DOWN) while upregulation was observed in 461 genes (UP). However, only 5.42% (UP) and 7.5% (DOWN) of genes with AGO1 clusters in the promoter region were affected by the AGO1 knockdown, indicating that AGO1 presence at the TSS vicinity is not generally related to the activation or silencing of the corresponding gene transcription. However, we found that genes upregulated by AGO1 depletion had higher AGO1 densities in the gene body than down-regulated ones, see (Figure 3.14).

AGO1 and alternative splicing

We used the siAGO1 vs siLuc RNA-Seq data to calculate events that changed significantly inclusion levels between siAGO1 and siLuc conditions (see Methods). We found that AGO1 knockdown promoted skipping of 334 and inclusion of 401 cassette exon events. Figures (3.18) and (3.18) show examples of these skipping and inclusion events. We found that higher changes in inclusion were related to higher densities of AGO1 (Figure 3.16). We calculated skipping and inclusion events upon AGO1 depletion using RNA-Seq (see Methods). Additionally we used as non-regulated events exons that did not change significantly between siAGO1 and siLuc. AGO1 correlated with exons skipped upon AGO1 depletion suggesting that AGO1 may prevent exon skipping of the regulated exon, probably co-localizing with other factors. Moreover, we recovered some of the ASEs regulated upon AGO1 depletion from (Allo et al., 2009). For instance, we found that BCL2L11, which was regulated upon AGO1 depletion in HeLa and FANCL, which was regulated upon AGO1 depletion in Hep3B (Allo et al., 2009), were classified as inclusion events in siAGO1. C11ORF4 and C11orf17, which were regulated upon AGO1 depletion in HeLa and HeLa and Hep3B, respectively (Allo et al., 2009), were classified as skipping events in siAGO1.

In order to find AGO1 dependent alternative splicing events candidates and a relation with sRNAs, we selected the events that changed inclusion upon AGO1 depletion that had an overlap with AGO1 ChIP-Seq significant clusters. Then, we looked for MCF7 sRNAs with exact sequence comple-

mentarity in the event region. We found that 8 exons that were skipped upon AGO1 depletion had sRNA targets with exact match, for the 8 cases AGO1 clusters overlapped in the alternative exon and the downstream region (Figure 3.20), and 7 exons that were included upon AGO1 depletion had exact complementary sRNA targets (Figure 3.21). We used as a control the shuffled sequences of the sRNAs and we did not find any exact matches within the AGO1 dependent alternative splicing events. The results suggested that these events could be regulated by an sRNA mediated regulation of alternative splicing as proposed in (Allo et al., 2009; Ameyar-Zazoua et al., 2012).

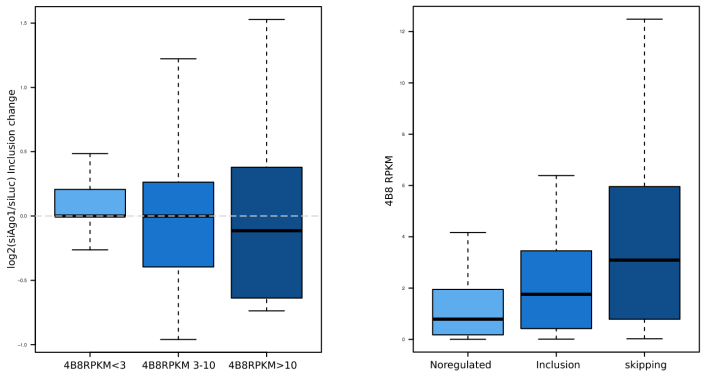


Figure 3.16: Left boxplot: AGO1 density in RPKM (x-axis) compared to the log_2 ratio of the inclusions (siAGO1 over siLuc) (y-axis). Right boxplot: AGO1 MCF7 density in RPKM (y-axis) compared to the non-regulated exons, included exons ($log_2 \frac{siAGO1}{siLuc} > 0$) and skipped exons ($log_2 \frac{siAGO1}{siLuc} < 0$) (x-axis) (see Methods). Skipped exons in siAGO1 showed higher densities.

AGO1 and constitutive splicing

We also looked for a possible association between AGO1 and constitutive splicing. We found that the splicing efficiency score (SES, see Methods) of 150 first introns and 1,279 internal introns was affected by AGO1 knockdown, with a predominance for upregulation of splicing efficiency observed in 119 (87%) first introns and in 943 (74%) internal introns, see (Figure 3.17) top boxplot. This indicates that the presence of AGO1 somewhat inhibits intron

excision. Although we cannot rule out an indirect effect due to changes in expression levels of constitutive splicing factors upon AGO1 knockdown, the fact that the mechanism is global could indicate a direct mechanism.

If AGO1 was involved in a mechanism of constitutive splicing for these events in the first and internal introns, we would expect to find AGO1 clusters near these introns. We looked for the overlap between upregulated first and internal introns and AGO1 clusters, allowing the overlap of AGO1 only in the upstream exon, in the intron or in the downstream exon, which would allow us to find the exact binding region were AGO1 could act preventing the excision of the tested intron. We found for these three regions in first and internal introns that splicing efficiency was higher when AGO1 was depleted, specially in the downstream region (Figure 3.17 middle boxplots). In (Figure 3.17) we show two examples for first and internal intron overlapping with AGO1 cluster.

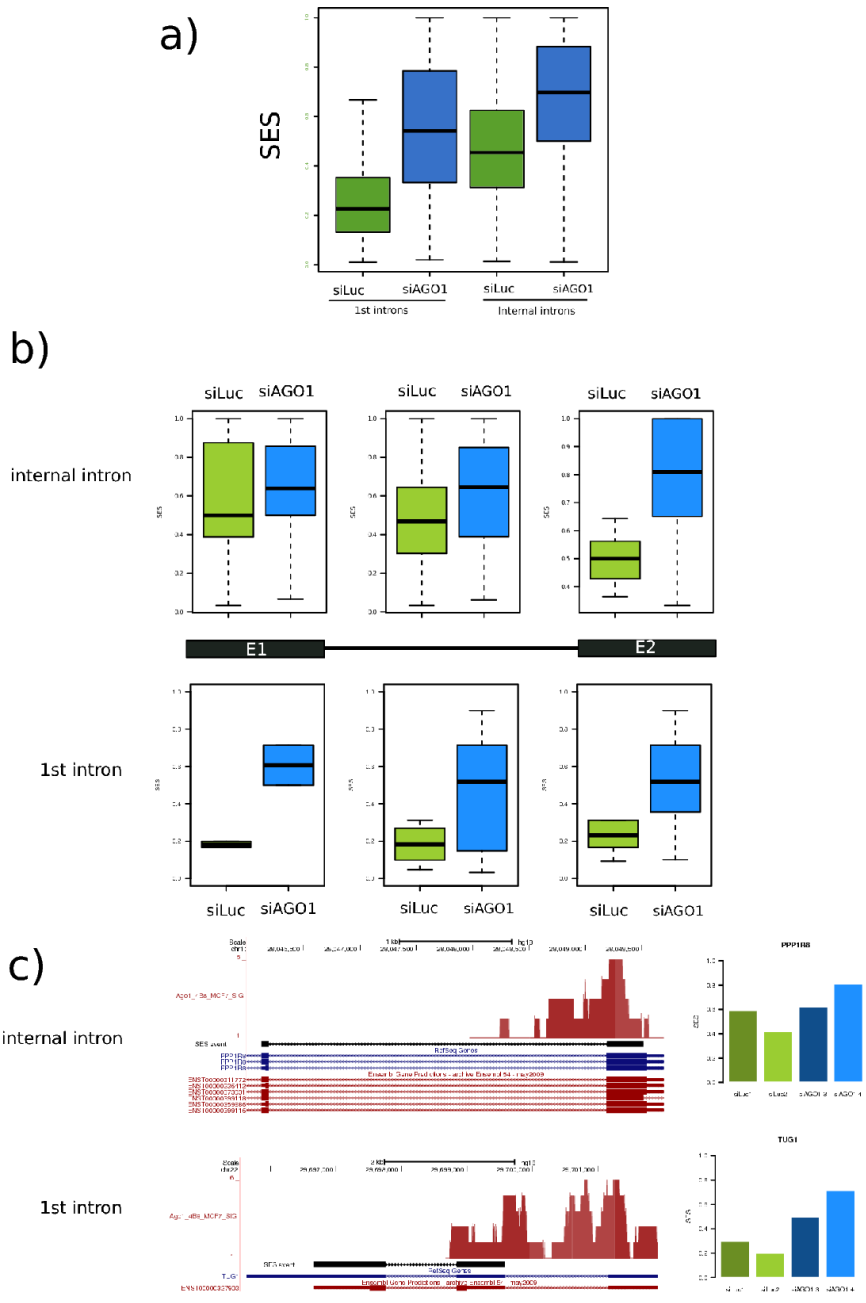


Figure 3.17: a) SES for first and internal intron in siLuc and siAGO1 conditions. b) SES for internal and first introns when there is an overlap with an AGO1 cluster in either the upstream exon, the intron or the downstream exon, for siLuc and siAGO1 conditions. c) Two examples of events overlapping with AGO1 clusters where SES changes significantly between siAGO1 and siLuc.

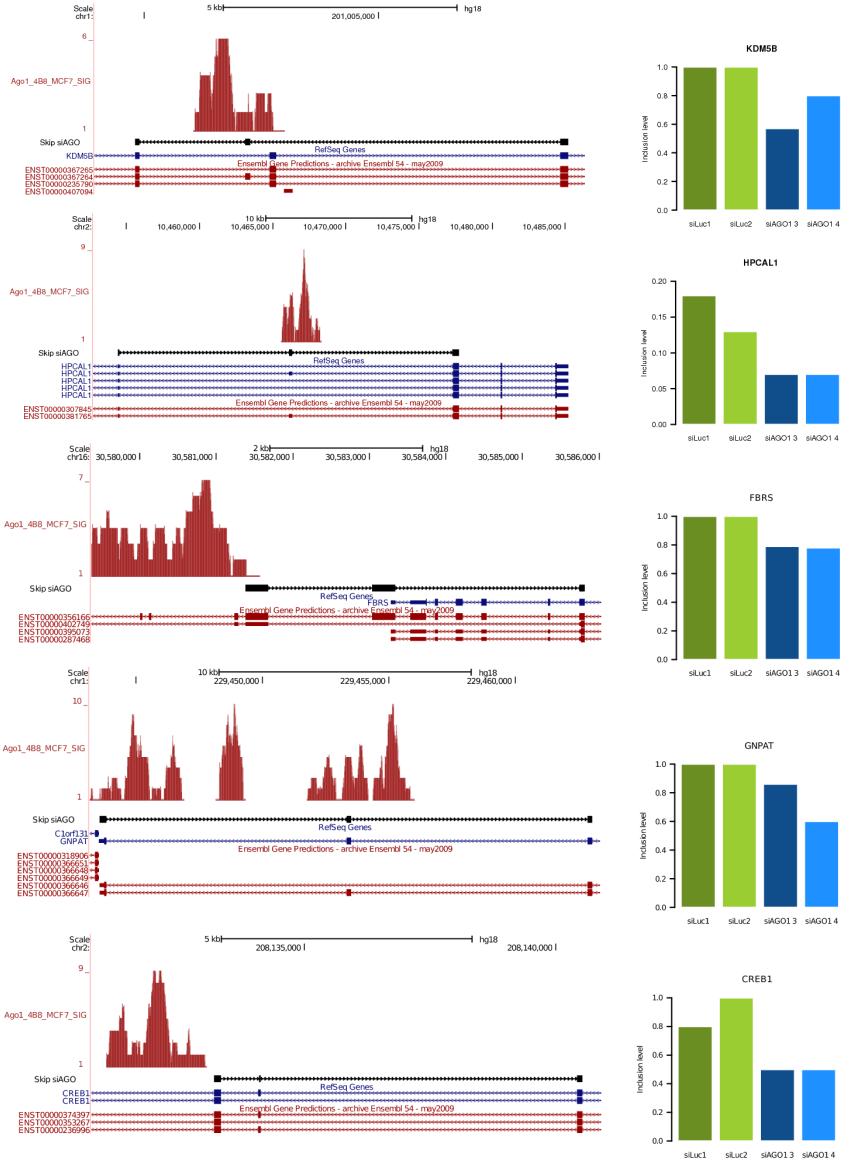


Figure 3.18: Examples of alternative splicing events regulated upon AGO1 knockdown. Inclusion levels are shown for each replicates in siLuc (green) and siAGO1 (blue). All these events are skipped upon AGO1 depletion and overlap with AGO1 clusters in MCF7.

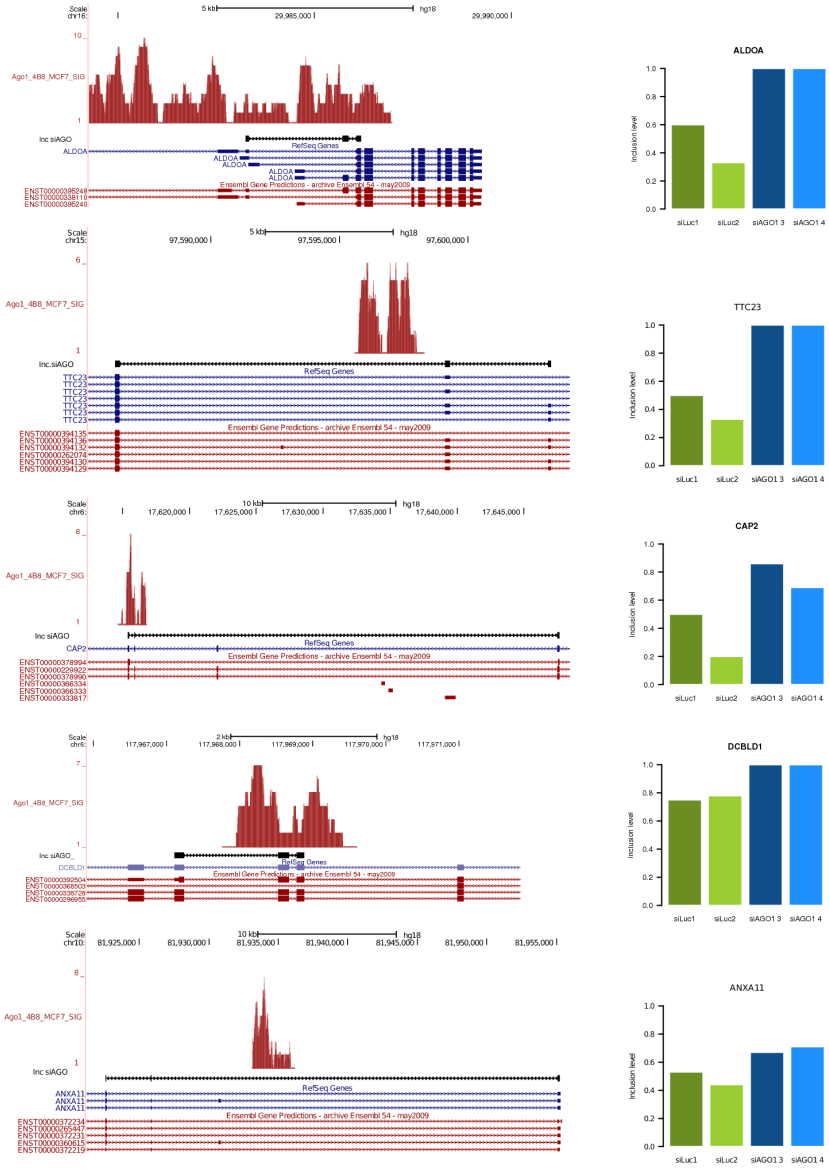


Figure 3.19: Examples of AGO1 alternative splicing events regulated upon AGO1 knockdown. Inclusion levels are shown for each replicates in siLuc (green) and siAGO1 (blue). All these events are included upon AGO1 depletion and overlap with AGO1 clusters in MCF7.

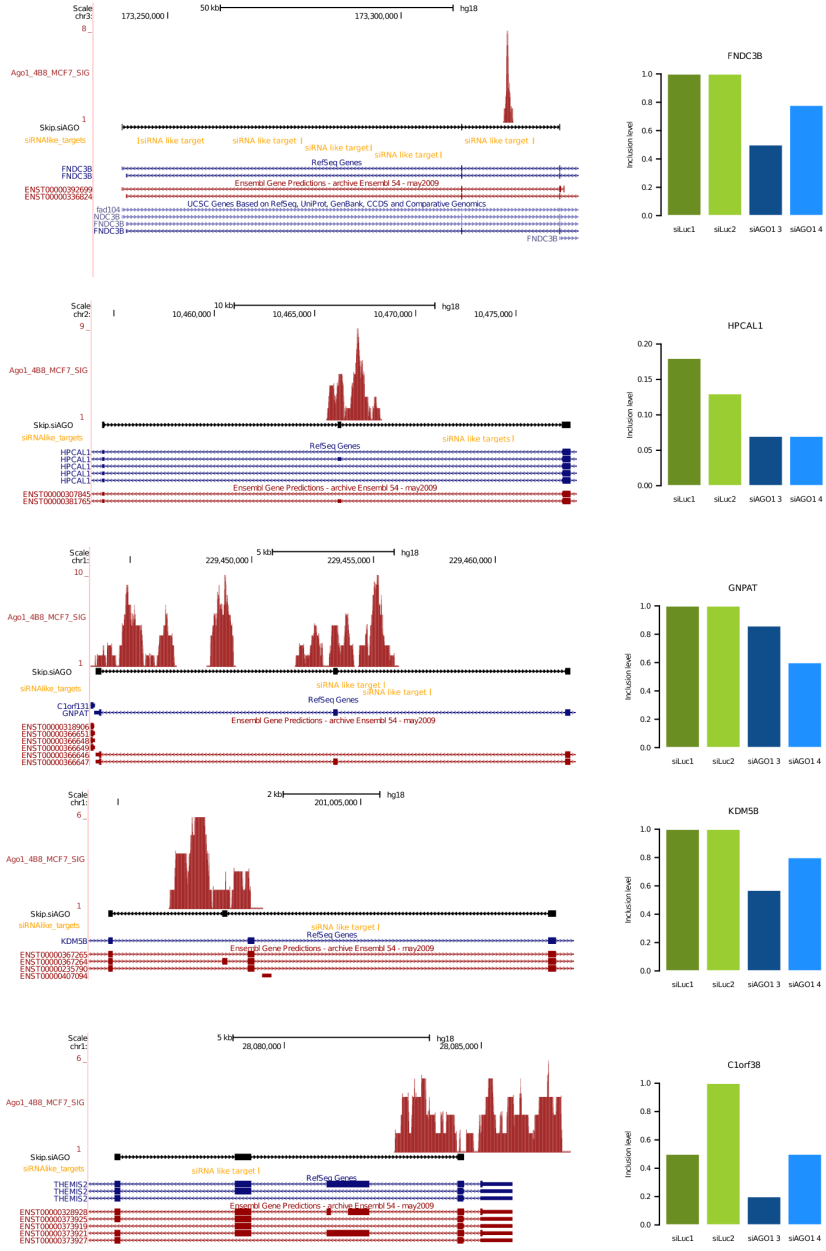


Figure 3.20: Examples of AGO1 alternative splicing events regulated upon AGO1 knockdown. Inclusion levels are shown for each replicates in siLuc (green) and siAGO1 (blue). All these events are skipped upon AGO1 depletion, overlap with AGO1 clusters in MCF7 and had exact complementarity MCF7 sRNA targets. The figure shows 5 events from the 8 that had evisende of sRNA targets, where the sRNAs were originating from a different locus. In orange, sRNA targets.

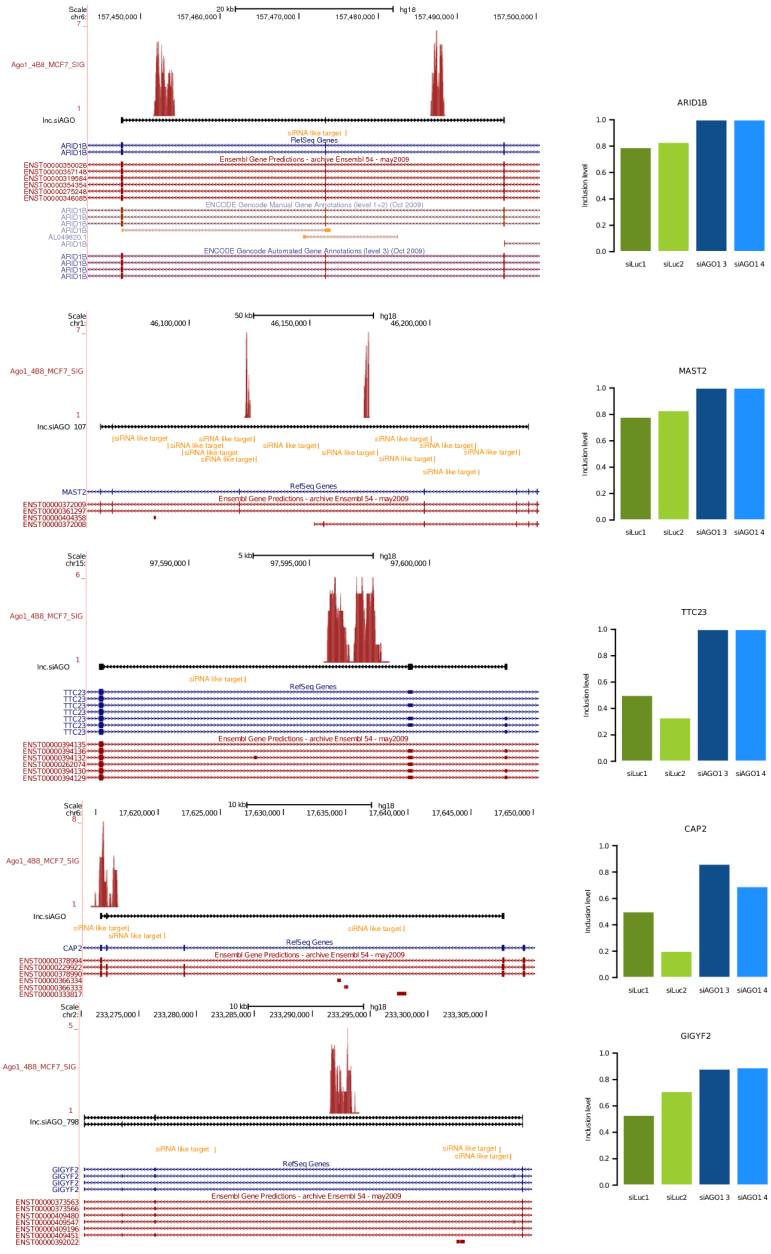


Figure 3.21: Examples of AGO1 alternative splicing events regulated upon AGO1 knockdown. Inclusion levels are shown for each replicates in siLuc (green) and siAGO1 (blue). All these events are included upon AGO1 depletion, overlap with AGO1 clusters in MCF7 and had exact complementarity MCF7 sRNA targets. The figure shows 5 events from the 7 that had evidence of sRNA targets, where sRNAs were originating from a different locus. In orange, the sRNA targets.

3.3 A chromatin code for cell specific alternative splicing

The results of this section belong to a third publication, that was under submission at the time this thesis was being written. In view of the results from previous section, we decided to combine ChIP-Seq data with alternative splicing arrays to investigate to what extent AGO1, CTCF and other chromatin signals can participate in cell-specific alternative splicing regulation.

We analyzed the combinatorial code of AGO1, CTCF, H3K27me3, H3K9me2, H3K36me3, RNAPII, HP1α, total H3 and 5metC in relation with the alternative splicing regulation between two cell lines, MCF7 and MCF10A. Using Machine Learning (ML) techniques, we obtained the relevant changes of chromatin-associated signals associated to splicing regulation between these two cell lines to build a chromatin RNA-map that can explain 602 (68.0995%) of the regulated events. Moreover, we found a possible intragenic association between HP1α, CTCF and AGO1 in association with these alternative splicing changes.

A chromatin description of splicing changes

We considered 1694 regulated cassette events between MCF7 and MCF10A cells obtained from a splicing junction array (Affymetrix HJAY). We selected 442 events of each type (inclusion in MCF7 and skipping in MCF7 compared to MCF10A) to have a balanced set for training and testing. All events were selected in genes that do not change expression significantly (logfold change p-value > 0.01) and had a significant change in splicing. We selected non-regulated events as control, obtained from exon triplets from the same host genes as the regulated events, non-overlapping with regulated events, and that were negative for splicing change in the same array, which resulted in 1970 non-regulated events. We analyzed ChIP-Seq data in MCF7 and MCF10A cells, for which we applied a processing and normalization

method, described in the Methods section. We recovered datasets of significant clusters for each ChIP-Seq sample applying a threshold of a p-value $<$ 0.05 for all the samples, in order to recover more signal from each of the data samples but being significant respect to the controls. For the analysis below we used the reads overlapping these significant clusters.

We focused on characterizing the cell-specific splicing changes in terms of chromatin signals. To this end, we defined 15 different regions around the cassette events (Figure 3.22). For each window, we calculated the relative change of read density, for each ChIP-Seq experiment between MCF7 and MCF10A, where the densities were calculated as RPKM. The significance of the changes was calculated using Pyicos (Althammer et al., 2011), in terms of the average densities in that region. Each attribute corresponds to the z-scores for relative enrichment for each sample-region pair. We thus generated 9x15=135 values per event, corresponding to the enrichment z-scores for each experiment-region pair and we therefore had 135 attributes.

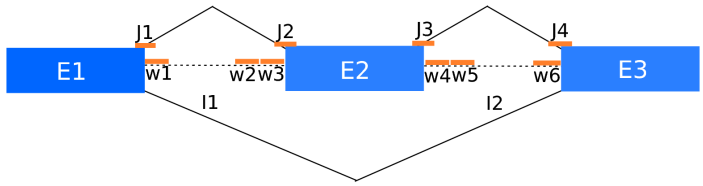


Figure 3.22: Diagram of the 15 defined regions on exon triplets. 300nt length windows flanking exons (w1, w2, w3, w4, w5 and w6), 200nt length junction regions, covering 100nt on either side of the exon boundaries (J1, J2, J3 and J4), the extent of the three exons (E1, E2 and E3) and the extent of the flanking introns (I1 and I2).

We first performed a pairwise correlation analysis of these 135 attributes for skipping, inclusion and non-regulated events. We obtained heatmaps by calculating Pearson correlation coefficients for every pair of attributes. The Figure (3.23) shows the heatmaps of the Pearson correlation coefficients for the 135 attributes. We found more dependencies between attributes in regulated events than in non-regulated events, which mostly showed correlations between attributes corresponding to histone modifications in the same regions. In non-regulated events, the average correlation coefficient was lower and significant anticorrelations were not found Table (3.1). High

correlations between different histone marks on large regions (specially the flanking upstream intron) in general were common for skipping, inclusion and non-regulated events. For instance, (H3K27me3-I1 vs. H3K9me2-I1) and (H3K36me3-I1 vs. H3K27me3-I1) showed a high correlation coefficient in the three groups of events. Meaning that the histone modifications enrichments in these regions were not informative enough to differentiate regulated from non-regulated events. However, differences between inclusion and skipping events were apparent: inclusion events showed high correlations between histone marks on the exon-intron junctions (Pearson correlation r^2 = 0.76 for H3K36me3-J1 vs. H3K9me2-J1, r^2 = 0.76 for H3K36me3-J3 vs. H3K9me2-J1), between histone marks and methylation on the downstream flanking intron and the downstream intron-exon junction (Pearson correlation r^2 = 0.66 for H3K9me2-I2 vs. metC5-J4) and on the first exon-intron junction and downstream windows (Pearson correlation r^2 = 0.7 for 5metC-J1 vs. CTCF-w5). In contrast, in skipping we found high correlations between histone marks and RNAPII (Pearson correlation r^2 = 0.8 for H3K36me3-I1 vs. RNAPII-w5, r^2 = 0.7 for H3K36me3-I1 vs. RNAPII-w2), CTCF and histone modifications (Pearson correlation r^2 = 0.66 for CTCF-J4 vs. H3K9me2-w2) and between AGO1 and H3K27me3 downstream of the alternative exon (Pearson correlation r^2 = 0.65 for AGO1-J4 vs. H3K27me3-w6). Tables (3.2) and (3.3) show the correlations with $r^2 > 0.6$ in inclusion and skipping events. We also found anti-correlating attributes for skipping and inclusion Tables (3.4) and (3.5) show all the anti-correlations with $r^2 <$ -0.6 for inclusion and skipping events. Inclusion events showed mainly AGO1 anti-correlating with RNAPII and H3K36me3 (Pearson correlation factor r^2 = -0.72 for AGO1-J2 vs. H3K36me3-J4, r^2 = -0.69 for AGO1-J2 vs. RNAPII-J4) and H3K36me3, while skipping events also showed strong anti-correlations between AGO1 and H3K27me3. These correlations and anti-correlations suggested a cooperative role of the different chromatin signals with different patterns in skipping and inclusion events and no changes in non-regulated events. Figure (3.24) shows the heatmaps for high correlations and anti-correlations between attributes for skipping, inclusion and non-regulated events.

Table 3.1: Pairwise correlation of attributes, non-regulated events

Pearson r^2 pairwise correlation values between attributes for non-regulated events with $r^2 > 0.5$, used as control. All correlations corresponded to histone modifications in the same region.

Non-regulated		
H3K9me2-I1	H3K27me3-I1	0.76
H3K27me3-I2	H3K36me3-I2	0.67
H3K27me3-I2	H3K9me2-I2	0.66
H3K36me3-I1	H3K27me3-I1	0.64
H3K36me3-I1	H3K9me2-I1	0.64
H3K36me3-I2	H3K9me2-I2	0.54
H3K27me3-w2	H3K9me2-w2	0.53
H3K27me3-J1	H3K9me2-J1	0.50

Table 3.2: Pairwise correlation of attributes, inclusion events

Pearson r^2 pairwise correlation values between attributes for inclusion events with $r^2 > 0.6$, in bold sample-regions with high correlation coefficient common in non-regulated events.

Inclusion		
H3K27me3-I1	H3K9me2-I1	0.77
H3K36me3-J1	H3K9me2-J1	0.76
H3K36me3-J3	H3K9me2-J1	0.76
H3K36me3-w1	H3K9me2-J1	0.74
H3K27me3-I1	H3K36me3-I1	0.71
5metC-J1	CTCF-w5	0.7
H3K36me3-E1	H3K9me2-J1	0.7
H3K36me3-J4	RNAPII-J4	0.69
H3K9me2-I2	5metC-J4	0.66
H3K36-w4	RNAPII-J4	0.64
H3K36-J1	H3K9me2-w1	0.64
H3K9-w1	H3K36me3-J1	0.64
H3K27-w5	H3K9me2-w5	0.63
H3K27-w3	H3K36me3-w3	0.62
H3K27-I1	H3K36me3-w3	0.61
H3K27-J1	H3K9me2-J1	0.61

Table 3.3: Pairwise correlation of attributes, skipping events

Pearson r^2 pairwise correlation values between attributes for skipping events with $r^2 > 0.6$, in bold sample-regions with high correlation coefficient common in non-regulated events.

Skipping		
H3K27me3-I1	H3K36me3-I1	0.87
H3K9me2-I1	H3K27me3-I1	0.81
H3K36me3-I1	RNAPII-w5	0.8
H3K9me2-I1	H3K36me3-I1	0.77
H3K36me3-I1	RNAPII-w2	0.7
RNAPII-J4	H3K36me3-E1	0.68
H3K36me3-I1	RNAPII-I1	0.68
RNAPII-J4	H3K36me3-w1	0.67
H3K36me3-I1	RNAPII-I2	0.66
CTCF-J3	H3K9me2-w2	0.66
H3K27me3-I1	RNAPII-w2	0.65
H3K36me3-J3	RNAPII-I1	0.65
RNAPII-I1	H3K36me3-J1	0.65
AGO1-J4	H3K27me3-w6	0.64
H3K27me3-I1	RNAPII-w5	0.64
H3K36me3-J3	RNAPII-J1	0.64
RNAPII-J1	H3K36me3-J1	0.64
H3K36me3-J3	RNAPII-J4	0.63
RNAPII-J4	H3K36me3-J1	0.63
metC5-J4	H3K9me2-I2	0.63
HP1-w2	RNAPII-w1	0.63
RNAPII-E1	HP1-w2	0.63
H3K27me3-w5	CTCF-E1	0.62
AGO1-w2	RNAPII-J1	0.62
HP1-E2	AGO1-w2	0.62
CTCF-w1	H3K27me3-w5	0.62
H3K27me3-E1	RNAPII-w2	0.62

A chromatin RNA map of splicing regulation

In order to select the most important attributes for the classification, we used a combination of attribute selection methods (see Methods). Table (3.6) shows the scores for the selected attributes.

Table 3.4: Pairwise anticorrelation of attributes, inclusion events

Pearson r^2 pairwise correlation values between attributes for inclusion events with $r^2 <$ -0.6

Inclusion		
AGO1-J2	H3K36me3-J4	-0.72
AGO1-J2	RNAPII-J4	-0.69
AGO1-E2	H3K36me3-J4	-0.68
RNAPII-I2	AGO1-w1	-0.66
RNAPII-I2	AGO1-E1	-0.66
H3K36me3-J4	AGO1-J3	-0.65
AGO1-J2	H3K27me3-w2	-0.65
H3K27me3-w5	HP1-E3	-0.63
HP1-w6	H3K27me3-w5	-0.63
AGO1-I1	RNAPII-J4	-0.62
RNAPII-J4	AGO1-E2	-0.61

Feature selection resulted in 16 attributes, involving 6 out 9 of the experiments considered. Interestingly, among the most relevant attributes we found AGO1, HP1α and CTCF downstream of the alternative exon related to inclusion events plus H3K36me3 and 5metC highly related to skipping events. Using the most informative attributes, we built a chromatin RNA-map that represents how the relative changes in signal densities at the chromatin level correlate with inclusion or skipping of exons figure(3.25).

Building a classifier with these 16 attributes, we were able to obtain 602 (68.09%) events correctly classified (312 inclusion, 290 skipping). This indicates that a considerable number of regulated events between MCF7 and MCF10A can be explained by the interplay of histone marks with AGO1, CTCF and HP1α. Figure (3.26) shows the distributions of the enrichment values of the correctly classifed events for the 16 selected attributes compared to the non-regulated events distribution. We found that AGO1 in the downstream window (w5) associates strongly with the direction of the splicing change: splicing events with an increase of AGO1 signal between MCF7 and MCF10A downstream of the alternative exon were more frequently associated to inclusion. This correlated with our previous results from the siAGO1 knockdown RNA-Seq, where we found that AGO1 correlates with

exons skipped upon AGO1 depletion.

The chromatin RNA-map also showed an increase of HP1α in the downstream intron, associated to inclusion. Similarly, an increase of CTCF on the downstream window (w6) was related to inclusion, leading to a possible association between CTCF, AGO1 and HP1α related to the inclusion of the alternative exon. Interestingly, we found that an increase of RNAPII downstream of the regulated exon had association with inclusion. On the other hand, for H3K36me3 and 5metC we found the opposite pattern: an increase on the flanking regions of the regulated exon correlated with inclusion. These results indicated a possible mechanistic association of these marks. As a comparison, when considering the relative enrichment of these signals in a set of non-regulated exons, we did not observe a significant bias (figure(3.26 shows the non-regulated events in gray).

Association between different signals and its role in splicing regulation

We decided to investigate whether our data showed any evidence of a possible association between AGO1 and the signals that appeared as strong predictors. First, we looked for positive or negative associations between the different samples. For this, we calculated a zscore from the overlap between the clusters of the different samples, overlapping the clusters from one sample vs. the rest of the samples, and we looked for reciprocal associations between them, (see methods). Interestingly, we found that there was a positive reciprocal association between CTCF & HP1α, CTCF & RNAPII and 5metC & HP1α. On the other hand, signals that are known to be antagonistic did not show any significant association, which was the case of 5metC & CTCF. Moreover, AGO1 was not reciprocally associated with any of the samples, we found that 30% of AGO1 clusters were overlapping with HP1α, 26% with 5metC and 20% with CTCF, meaning that AGO1 could be associated specifically to these signals but not the other way around (Table 3.7 shows the proportions of clusters from one sample overlapping with other sample and their zscores). Based on recent evidence that a *Drosophila* homolog of AGO1 associates to CTCF at the level of chromatin (Moshkovich et al., 2011) and that HP1α is one of the components of TGS-AS (Allo et al., 2009), we decided to investigate whether these associations were localized in specific genic regions. We found that there was a significant overlap between ChIP-Seq clusters of AGO1 & CTCF and HP1α & CTCF in promoters and intragenic regions (Table 3.7). Additionally, we looked for the densities of CTCF around HP1α clusters and of HP1α around CTCF clusters compared to random clusters. The densities showed that there was colocalization between HP1α and CTCF, while in the random clusters was absent (Figure 3.27).

In order to further explore the relation between AGO1, CTCF and HP1α activity, we calculated enriched DNA motifs in the three cluster sets independently (see Methods). Although there was a high overlap between the three samples clusters, we found different 7mers for each of them (see Methods for the 7mers enrichment calculation). We built consensus motifs using all significant clusters genome-wide (see Methods) and we found a logo for AGO1,

HP1α and CTCF (Figure 3.28), CTCF logo coincided with the known logo previously reported (Essien et al., 2009). The HP1 family is a non-histone chromosomal protein family involved in the establishment and maintenance of higher-order chromatin structures that repress gene expression. The HP1 proteins consist of two highly conserved domains, one for chromatin binding and the other for protein protein interaction. The two domains are separated by a hinge region of variable length that has been related to DNA and RNA binding (Hiragami-Hamada et al., 2011; Kwon and Workman, 2011). Apart from that, there is no evidence of DNA binding domains in HP1 proteins. However, we found a consistent logo for the binding sites of HP1α.

To investigate the association of AGO1, CTCF and HP1α in relation to splicing regulation, we analyzed whether any particular combination of changes between them was enriched in either inclusion or skipping. To this end, we encoded the changes of AGO1, CTCF, HP1α and 5metC in terms of discrete values: increase (+) or decrease (-) (see Methods). Using this discretization, we found that inclusion events were associated with increase of HP1α and CTCF attributes (Figure 3.29 *a* shows a higher proportion of inclusion events (blue) when both, CTCF and HP1α increase, whereas skipping events were associated with the opposite pattern Figure(3.29 *a*). Increase of CTCF in the exons (E1, E2 and E3) and HP1α in the flanking introns (I1 and I2) corresponded with inclusion events. This correlated with the selected attributes, where we found HP1α in the downstream intron (I2) and CTCF near exons as relevant attributes to explain inclusion events. On the other hand, when CTCF and HP1α decrease, specially in the downstream intron (I2), we found a high proportion of skipping events. This suggested that HP1α and CTCF may had a cooperative role with respect to splicing regulation in the direction of inclusion. When we compared CTCF and 5metC, we found an antagonistic association, where an increase of 5metC and a decrease of CTCF were more associated to skipping, while a decrease of 5metC and an increase of CTCF were more associated to inclusion. Figure 3.29 *b*, shows that 5metC increase, specially in the flanking introns (I1 and I2), and decrease of CTCF in the exons (E1, E2 and E3) is related to skipping (red), in agreement with the model proposed by (Shukla et al., 2011a).

Since we found AGO1 downstream of the alternative exon (w5) as a

relevant attribute for our model related with the inclusion of the alternative exon, we compared AGO1 with CTCF and HP1α. In both cases, we found that an AGO1 decrease was related to skipping events, which corroborated our results where AGO1 presence prevented skipping. AGO1 increase showed a slight relation to inclusion events (Figure 3.29 *c* and *d*). However, we mostly found higher proportions of inclusion events when HP1α and CTCF increase (Figure 3.29 *c* and *d*). When comparing AGO1 with HP1α (Figure 3.29 *c*), we found HP1α to be dominant over AGO1, whereas an increase of AGO1 with an increase of HP1α showed no change in skipping or inclusion direction. When comparing AGO1 with CTCF (Figure 3.29 *d*), CTCF increase was more related to inclusion than AGO1, only when there was a decrease of AGO1 we found high proportions of skipping events. In summary, our data indicated an association between CTCF and HP1α related to inclusion of the regulated exon and decrease of AGO1 related to skipping of the regulated exon.

Then, we selected the ChIP-Seq samples that provided significant informative features in our splicing chromatin code and looked for the read densities around the 3'ss of the three exons that composed the events triplets, with exons classified as skipped and included in the model and the non-regulated events. In general we saw accumulation of RNAPII at the 3'ss (Kornblihtt, 2006), in the upstream exon (E1) and downstream exon (E3) of the inclusion and skipping events but only in the upstream exon in the non-regulated events. We found a peak of enrichment for RNAPII near the 3'ss of the upstream exon (E1), which could be related to the signal of the TSS. In order to remove possible TSS biases, we removed the events in which the upstream exon (E1) was the first, second or third exon of the gene. We corroborated the relevance of CTCF, with high densities present in the three exon bodies in inclusion events, colocalized with RNAPII (Figure 3.30 *c*). While in skipping events, CTCF was absent in exon bodies but present upstream of the three exons (Figure 3.30 *a*). Every CTCF peak in the exon bodies of the inclusion events was related with a higher density of HP1α, which increased the hypothesis of a collaborative role between CTCF and HP1α in the regulation of the alternative exon (Figure 3.30 *c*). It was prominent the 5metC depletion in the exonic regions while the flanking introns showed high levels of 5metC, in inclusion and non-regulated events (Figure 3.30 *c* and *b*).

On the other hand, skipping events showed higher signal of H3K36me3 in the exon regions and a lack of interplay between CTCF, RNAPII and HP1α. Suggesting, that the modulation between CTCF, RNAPII and other histone modifications maybe be influenced by HP1α and 5metC (Figure 3.30 *b*). Although AGO1 was selected as a relevant attribute in the downstream region (w5). We did not find high differences in the AGO1 signal between skipping, inclusion and non-regulated events (Figure 3.30), which may suggest an AGO1 specific regulation of splicing only for some specific cases.

We found clear differences between alternative exons, skipped and included, and the exons from non-regulated events when we looked at the mean read densities of H3K36me3, HP1α, 5metC and CTCF. The regulated exons showed an upstream peak of CTCF and a peak of CTCF in the exons, when the exon was included (Figure 3.31 *c*). While skipping events showed H3K36me3 higher than HP1α (Figure 3.31 *a*) and non-regulated events did not show high densities of any of the samples (Figure 3.31 *b*). When we looked at the normalized densities (Figure 3.31 *d*), we corroborated the high densities of HP1α and CTCF in inclusion events and high densities of 5metC and H3K36me3 in skipping compared to non-regulated and inclusion events. The results showed that the samples selected by the classifiers were enough to describe the different patterns of inclusion, skipping and non-regulated events for a considerable fraction of the events. Splice junctions and the interplay between different chromatin signals appeared to be strong attributes for alternative splicing regulation choices when comparing MCF7 with MCF10A.

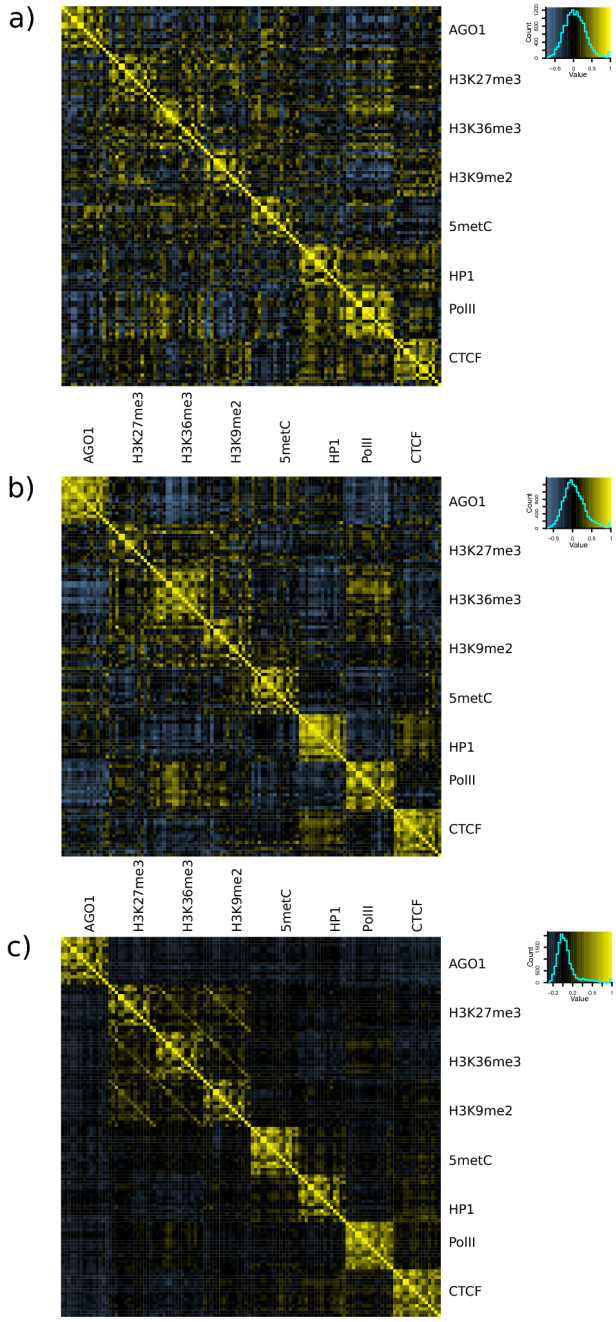


Figure 3.23: Pairwise correlations between attributes for skipping (a), inclusion (b) and non-regulated (c) events. Each attribute is defined by a sample-region pair.

Table 3.5: Pairwise anticorrelation of attributes, skipping events

Pearson r^2 pairwise correlation values between attributes for skipping events with $r^2 <$ -0.6

Skipping		
AGO1-E3	RNAPII-I2	-0.76
AGO1-w6	RNAPII-I2	-0.75
RNAPII-J2	AGO1-I2	-0.75
H3K27me3-w2	AGO1-J2	-0.74
H3K9me2-J3	RNAPII-w5	-0.72
H3K9me2-J4	HP1-w4	-0.72
H3K27me3-w2	AGO1-w3	-0.71
RNAPII-E2	AGO1-I2	-0.7
AGO1-J4	RNAPII-I2	-0.7
H3K9me2-w4	RNAPII-w4	-0.69
RNAPII-w3	AGO1-I2	-0.68
AGO1-w6	RNAPII-J2	-0.68
AGO1-E3	RNAPII-J2	-0.68
H3K9me2-w4	RNAPII-w5	-0.67
AGO1-J4	RNAPII-J2	-0.67
H3K9me2-J3	RNAPII-I2	-0.67
RNAPII-J3	AGO1-I2	-0.66
RNAPII-J3	AGO1-J4	-0.66
AGO1-J4	RNAPII-E2	-0.65
RNAPII-w3	AGO1-J4	-0.65
H3K27me3-I1	AGO1-E3	-0.64
AGO1-w6	H3K27me3-w2	-0.64
H3K27me3-w2	AGO1-E2	-0.64
RNAPII-I2	AGO1-I2	-0.64
H3K27me3-w2	AGO1-E3	-0.64
AGO1-E3	RNAPII-E2	-0.63
AGO1-w6	RNAPII-E2	-0.63
5metC-J4	RNAPII-I1	-0.63
H3K27me3-I1	AGO1-w6	-0.63
H3K36me3-I1	AGO1-E3	-0.63
H3K27me3-w6	RNAPII-I2	-0.63
AGO1-w6	H3K36me3-I1	-0.62
RNAPII-w5	CTCF-J1	-0.61
H3K27me3-w6	RNAPII-w3	-0.61
AGO1-E3	RNAPII-I1	-0.61
AGO1-w6	RNAPII-w3	-0.61

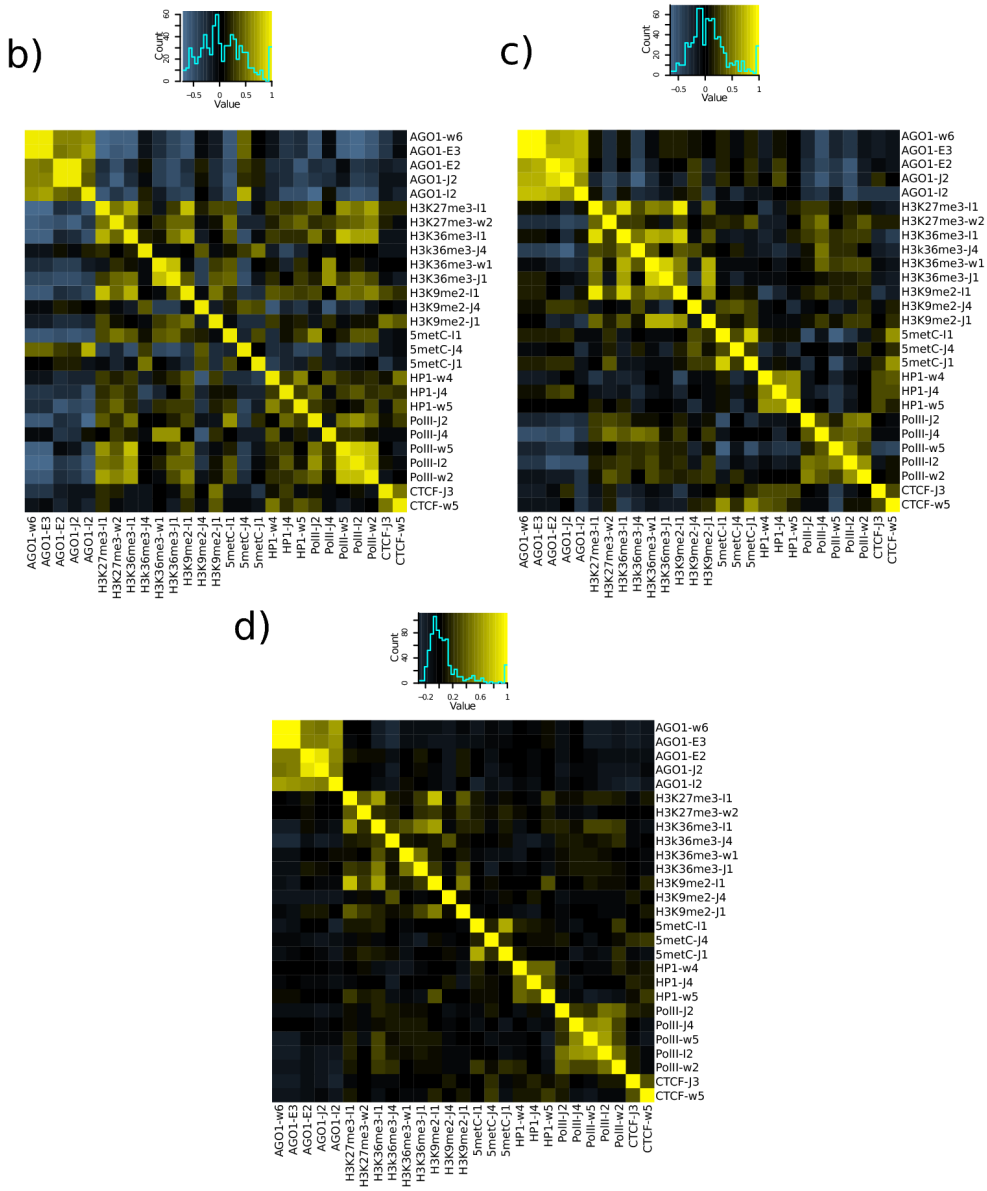


Figure 3.24: Pairwise correlations between attributes for skipping (a), inclusion (b) and non-regulated (c) events. Each attribute is defined by a sample-region pair.

Table 3.6: Attribute selection

We applied three independent attribute selection methods: Wrapper Subset Evaluator (WSE), Correlation Feature Selection (CFS) and Information Gain (IG)(See Methods). The table shows the attributes that have WSE and CFS $>=$ 50%, and a position in the IG ranking on the top 50%. Then we selected the attributes that belong to non overlapping regions based on the IG ranking. The last column shows the 16 attributes selected to describe the chromatin RNA-map.

Attribute	WSE(%)	CFS (%)	IG (Rank)	Selected attribute
AGO1-w5	70	70	43	*
H3K36me3-w3	100	100	1	*
H3K36me3-J3	60	100	2	*
H3K36me3-I1	100	100	6	
H3K36me3-E1	60	70	7	*
H3K36me3-E2	60	80	8	
H3K36me3-J4	90	90	19	*
HP1-I2	80	70	25	*
HP1-w1	60	60	53	*
RNAPII-J4	50	60	40	*
5metC-w4	50	90	11	*
5metC-w1	80	90	12	*
5metC-w6	60	60	18	*
5metC-E1	50	70	20	*
5metC-J3	60	80	22	
5metC-I2	50	50	28	
5metC-w2	50	70	30	*
CTCF-w3	60	70	36	*
CTCF-w6	80	70	52	*
CTCF-J1	50	60	64	*

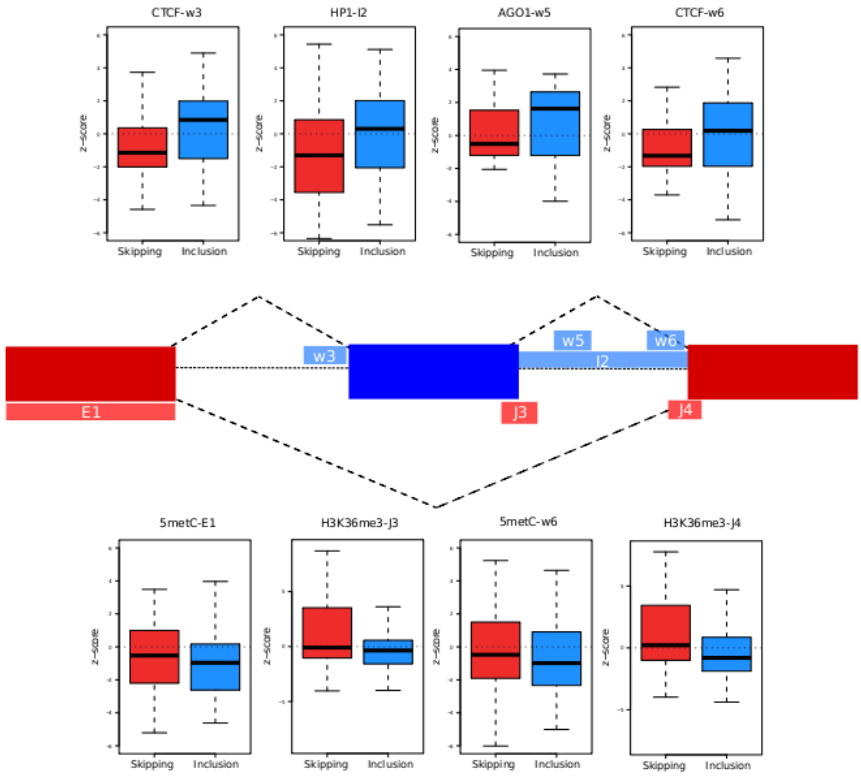


Figure 3.25: Chromatin RNA-map with some of the selected attributes. Each boxplot represents the relative change in signal densities as zscore values correlated with inclusion or skipping exons. Attributes that show enrichment in skipping exons (red) and attributes that show enrichment in inclusion exons (blue). The exon triplet diagram in the middle shows the regions of the selected attributes.

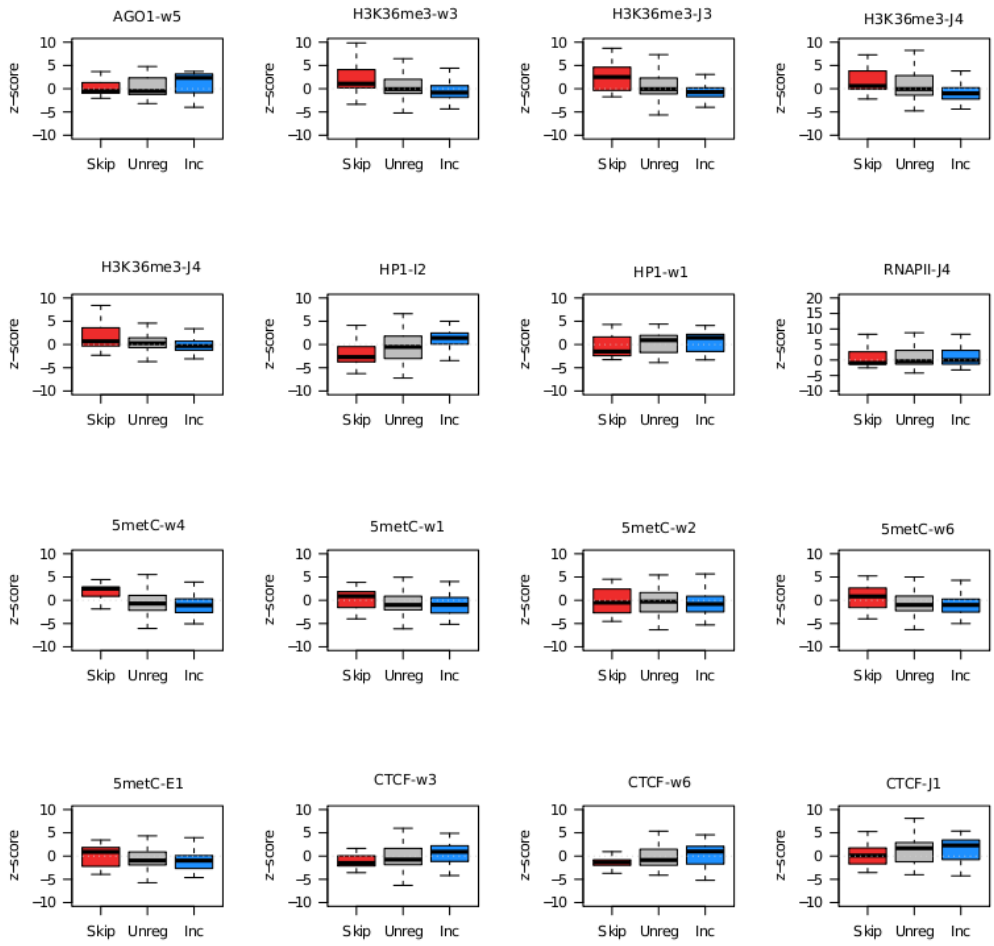


Figure 3.26: Boxplots of the 16 selected attributes from the classifier. Comparison between events that were correctly classified; skipping (red), inclusion (blue) with the distribution of the enrichment zscores for the non-regulated events (grey).

Table 3.7: Genome wide overlaps

Genome wide associations for the ChIP-Seq significant clusters. The table shows the highest % overlap between two samples clusters and their zscore. First column, samples which clusters were overlapped. Second column, the % of clusters from sample A that overlap with sample B clusters.Third column, the zscore from the overlap. Fourth column, the cases where there is a positive reciprocal overlap between sample A and sample B.

SampleA vs SampleB	% of A overlapping B	zscore	Positive association
AGO1 vs CTCF	19.2	3.21	
AGO1 vs 5metC	26.4	24.29	
AGO1 vs HP1α	29.4	8.41	
CTCF vs 5metC	20.4	168.85	
CTCF vs HP1α	18	111.52	*
CTCF vs RNAPII	11	42.94	*
RNAPII vs CTCF	11.2	113.38	*
RNAPII vs 5metC	21.2	173.55	
RNAPII vs HP1α	16.4	110.68	
5metC vs HP1α	30.7	3.05	*
HP1α vs CTCF	11.4	17.12	*
HP1α vs 5metC	29.4	120.7	*

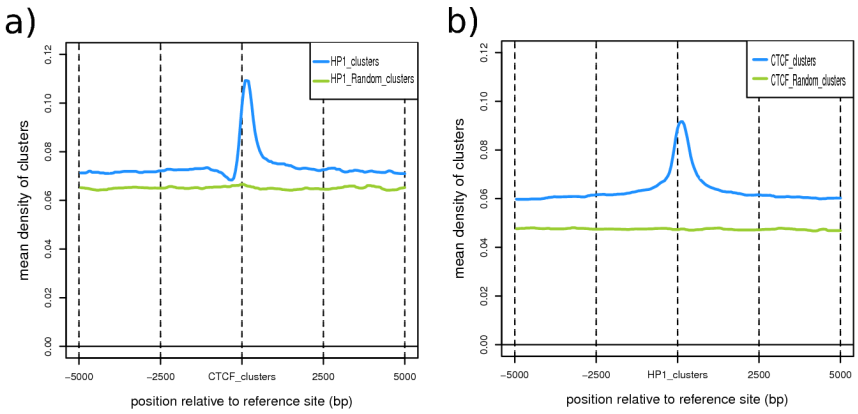


Figure 3.27: a) HP1α clusters mean densities centered in the CTCF clusters, compared to random HP1α clusters. b) CTCF clusters mean densities, centered in the HP1α clusters, compared to random CTCF clusters. There is colocalization between HP1α and CTCF and random clusters do not show any enrichment.

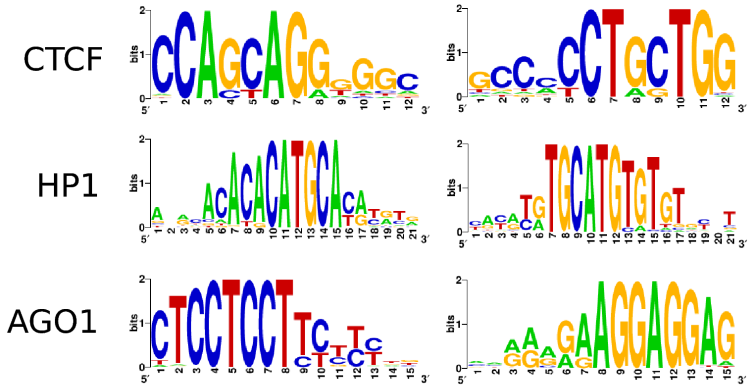


Figure 3.28: Consensus motifs for CTCF, HP1α and AGO1. Motifs were built using all the ChiP-Seq significant clusters (see Methods). CTCF logo corresponds to the logo known previously reported.

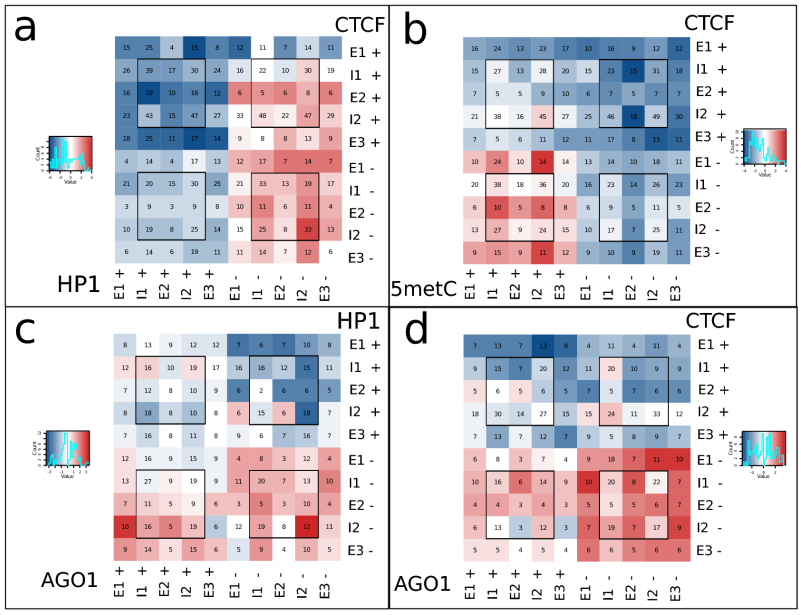


Figure 3.29: Association of inclusion and skipping to the combined relative changes of a) HP1α and CTCF, b) 5metC and CTCF, c) AGO1 and HP1α, and d) AGO1 and CTCF. Density changes as discretized as increase (+), decrease (-) or no change (not shown) from the distribution of all the regulated events correctly classified. We considered every possible combination of the changes for each comparison, HP1α vs. CTCF, 5metC vs. CTCF, AGO1 vs. HP1α and AGO1 vs. CTCF, for the region of the regulated exon (E2), the flanking exons (E1 and E3) and the flanking introns (I1 and I2). The central square shows the regulated exon and the flanking introns (I1,E2,I2). For each pair of discretized attributes, the log_2-rate of the proportions of skipping over inclusion is indicated in color. The color indicates whether the proportion is higher for skipping (red) or inclusion (blue). Inside the box we indicate how many events are involved in skipping (red) or inclusion (blue).

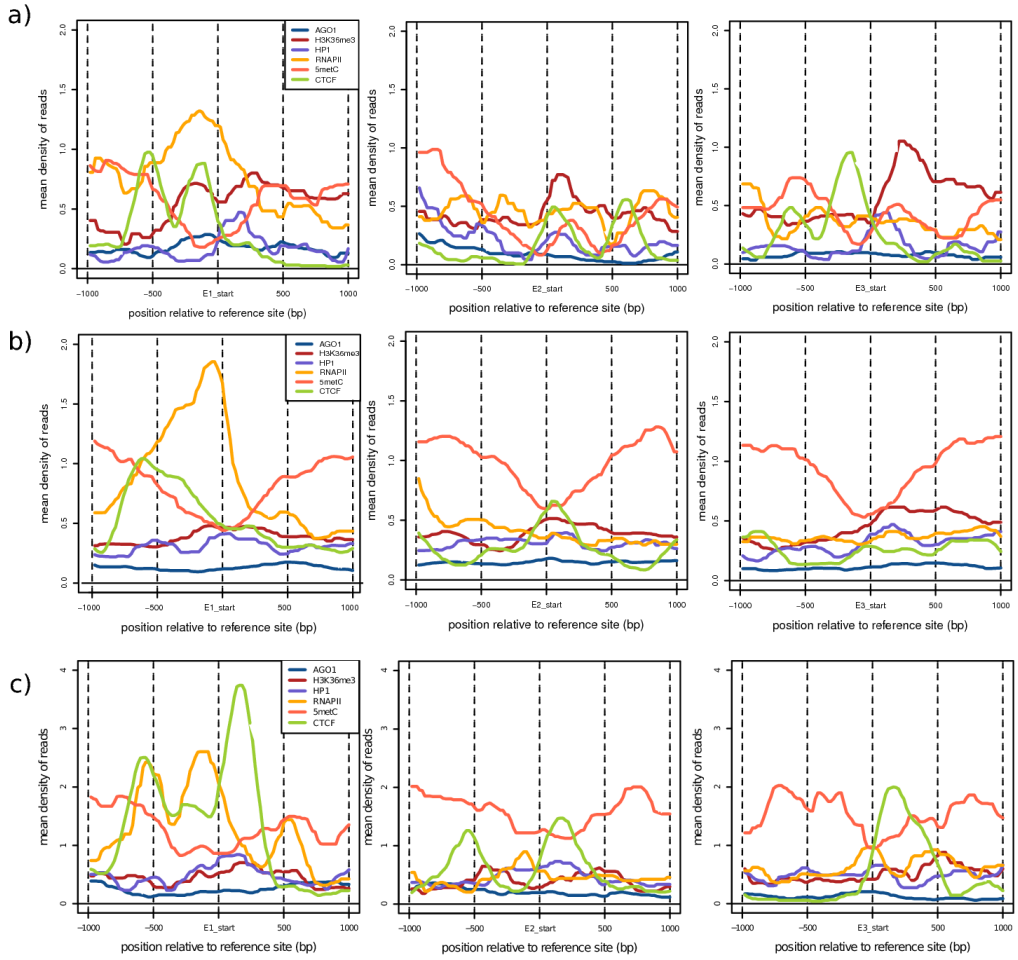


Figure 3.30: Mean reads densities around 3'ss for the three exons that compose the events triplets, for AGO1, H3K36me3, CTCF, HP1α, RNAPII and 5metC. Left profiles show the mean read densities from -1000 bp to 1000 bp centered in the 3'ss of the upstream exon (E1) for the six ChIP-Seq samples, second and third profiles show densities centered in the 3'ss of the internal (E2) and downstream exon (E3), respectively. a) Skipping events that were correctly classified in the model. b) Non-regulated events, that do not show changes in splicing. c) Inclusion events, that were correctly classified in the model.

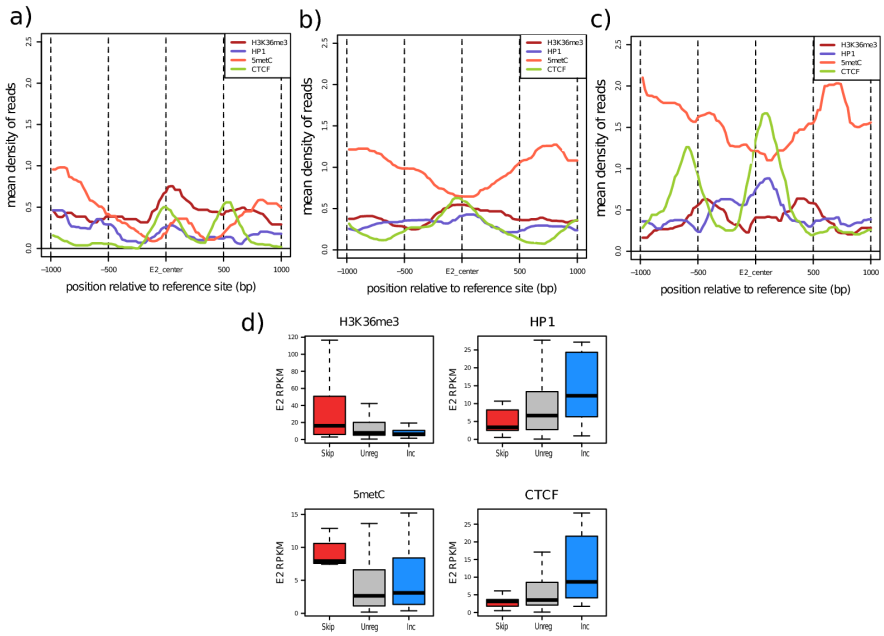


Figure 3.31: Mean reads densities as reference centered in the exon 2 of H3K36me3, HP1α, 5metC and CTCF in MCF7, for a) skipping events, b) non-regulated events and c) inclusion events.d) Distribution of the RPKMs in the exon 2 for the same samples in skipping (skip), non-regulated (unreg) and inclusion (inc) events.

CHAPTER 4

Discussion

Contents

This section is divided accordingly to three articles: (Allo et al., 2009) and two other that at the moment of writing have been submitted for publication. Even though the three articles are connected, the discussion is divided based on the results of each one to make more understandable the context at which they were performed. The second article can be understood as a continuation of (Allo et al., 2009), were we performed genome wide analysis in order to validate our results in (Allo et al., 2009) and to study the genomic distribution of AGO1. The third article describes a chromatin code for splicing regulation in the two studied cell lines MCF7 and MCF10A.

4.1 siRNA mediated transcriptional gene silencing affects alternative splicing

The experimental results from this work provided the first evidence of regulation of alternative splicing by siRNA mediated transcriptional gene silencing. We proposed the transcriptional gene silencing alternative splicing (TGS-AS) model. Previous evidences showed a similar mechanism that was exclusive to promoters, siRNAs directed to promoter regions mediated gene silencing (Morris et al., 2004; Castanotto et al., 2005; Suzuki et al., 2005; Ting et al., 2005). As proposed by (de la Mata et al., 2003), the slow down of the RNAPII elongation rate leads to the inclusion of the alternative exon. In the TGS-AS, silencing marks act as roadblocks slowing down RNAPII subsequently affecting splice site selection, which supports the involvement of an RNAPII elongation mechanism. Recently reports showed further support of a role of Argonaute i mechanism (Guang et al., 2010; Cernilogar et al., 2011). It seems that the role of nuclear RNAi in co-transcriptional silencing and later inhibition of RNAPII elongation and transcription can be generalize to other eukaryotes. Guang et al found that *C. elegans* Argonaute protein NRDE-2, that has the slicer activity as mammalian AGO2, is recruited by siRNAs to mediate transcriptional gene silencing in the nucleus leading to inhibition of RNAPII (Guang et al., 2010).

In order to find a general mechanism, we looked for evidence of alterna-

tive exons that shared the same characteristics that the FN1 model. We found that alternative exons that were affected by AGO1 depletion had shorter upstream introns. This lead us to hypothesize that RNAPII elongation changes may be a response on different genic structural features. Shorter introns could be more sensitive to RNAPII elongation changes originated by chromatin structure. Later evidences showed that histone modifications can promote recruitment of splicing factors affecting the regulation of an alternative exon (Luco et al., 2011). This mechanism can also be explained as cotranscriptional, involving both the RNAPII kinetics and the recruitment of splicing factors as proposed by (Cramer et al., 1999).

Since we found evidence of AGO1 and DCR dependent alternative splicing events (ASE) from a cancer related alternative splicing platform in HeLa and Hep3B, it was possible that endogenous sRNAs could also be acting as promoters of the TGS-AS. We actually found a high number of miRNAs overlapping intronic regions. Moreover, we found introns that were flanking alternative exons with overlapping miRNAs. Our results also showed that miRNAs could have targets in introns flanking alternative exons, as proposed by (Kim et al., 2008). These results suggested that the regulation of alternative splicing by TGS could not be only restricted to siRNAs. Previous reports showed that AGO1 can be localized in the nucleus (Kim et al., 2006; Janowski et al., 2006) and in eukaryotes, different classes of sRNAs associate with Argonaute family proteins (Hock and Meister, 2008). Argonaute sRNA (AGO-sRNA) complexes induce transcriptional silencing attaching chromatin remodelers to target loci (Sugiyama et al., 2005). In *S. pombe* this mechanism is well known, where the AGO-sRNA complex associates with H3K9 to repress the target region. Then, methylation of H3K9 leads to recruitment of swi6, HP1 *S. pombe* homolog, which produces the silencing and the interaction with the AGO-sRNA complex (Nakayama et al., 2001; Maison and Almouzni, 2004). Interestingly, we also found ASEs regulated upon DCR depletion. DCR is also known to affect chromatin structure, Haussecker et al found upregulation of β-globin intergenic transcripts in cells that were regulated upon DCR depletion leading to chromatin relaxation and histone tail modifications (Haussecker and Proudfoot, 2005). This suggested that DCR can also act as one of the effectors of the TGS-AS. These evidences out forward AGO1 as a strong candidate to mediate TGS-AS, and

DCR as a key enzyme to generate some of the physiological effectors of TGS-AS.

Experimental evidence showed that an antisense siRNA lead to formation of local heterochromatin marks near the alternative EDI exon (Allo et al., 2009), probably due to the hybridization with the nascent sense FN1 pre-mRNA. These results suggested a similar mechanism as proposed by Morris et al, where the sRNA guide strand loaded into the silencing complex interacted with a nascent transcript leading to an RNA-RNA hybridization (Morris et al., 2004). We also found an antisense EST that covered the affected region, consistent with (Schwartz et al., 2008; Morris et al., 2008). However, when we looked for further evidence of antisense transcription, we only found three antisense transcripts that fall in the flanking introns of the ASEs affected upon AGO1 depletion. Due to limited available data at that moment, it is easy to think that a more complete annotation and analysis of RNA-Seq data would probably provide more evidence of antisense transcripts overlapping candidate ASEs. In this sense, recently Ameyar-Zazoua et al proposed a model in which they suggested that long antisense non-coding RNAs could have a role in the recruitment of AGO1, AGO2 and splicing factors. Indeed, they proposed that AGO2 sense short RNA complexes bind to antisense transcripts, leading to H3K9 methylation in the CD44 gene model (Ameyar-Zazoua et al., 2012). The genome-wide relation between alternative splicing and antisense transcription has been known before (Enerly et al., 2005). Furthermore, Ameyar et al found a physical association between AGO1 and AGO2 with some spliceosome components. Thus, Ameyar-Zazoua et al proposed a similar mechanism to *S. pombe* Argonaute-containing RNA-induced transcriptional gene silencing (RITS) that generates local heterochromatin remodeling (Nakayama et al., 2001; Maison and Almouzni, 2004). These results showed the link between Argonaute proteins, chromatin remodelers and the splicing machinery (Ameyar-Zazoua et al., 2012), in agreement with previous results (Allo et al., 2009).

It is not yet clear the origin of the sRNAs involved in the mechanism. We found that there was evidence of miRNAs overlapping the ASEs and that the introns flanking these alternative exons could be candidate target regions for miRNAs. On the other hand, our results showed ASEs that were

only regulated upon AGO1 depletion and not DCR which suggested that TGS-AS may be not limited to double-stranded sRNAs. Interestingly, there are multiple classes of sRNAs and some have a DCR independent processing. Indeed, RNAs from different genomic sources have been reported to mediate transcriptional gene silencing and some of them are able to affect chromatin structure independently of DCR. Among these, PIWI small RNAs (piRNAs) are good candidates for TGS-AS. piRNAs can direct a specialized sub-class of Argonaute proteins, the PIWI proteins, to silence complementary target regions. However, results from Ameyar-Zazoua clearly showed that the recruitment of AGO1 and AGO2 to CD44 required DCR (Ameyar-Zazoua et al., 2012). Thus, the sRNAs involved in their model should be restricted to miRNAs or siRNAs. We proposed that the sRNAs involved in our model could be originated from different pathways, DCR dependent and independent. Recently, Le Thomas et al suggested based on their results that *Drosophila* PIWI could act similarly to Argonautes establishing repressive H3K9me3 on its targets by recruitment of an enzymatic machinery, as it is known in *S. pombe* (Le Thomas et al., 2013).

In summary, our results defined a mechanism for the regulation of alternative splicing showing the link between alternative splicing and the RNAi pathway, the TGS-AS. We found that AGO1 was is involved in the pathway, we predicted ASEs canditates that could be regulated by the TGS-AS and we suggested that endogenous sRNAs from different classes and sources could be involved. Later reports confirmed further evidence in that direction. However, there are still differences between the mechanisms described by different groups. For that reason, we continued to study the AGO1 genome-wide distribution and its effect in the regulation of alternative splicing in order to find genome-wide evidence of the TGS-AS.

4.2 Genome-wide analysis of AGO1 and its role in alternative splicing

In view of our results in (Allo et al., 2009) and the more recent results from other groups (Guang et al., 2010; Cernilogar et al., 2011; Ameyar-Zazoua et al., 2012; Taliaferro et al., 2013) we proceeded to define the genomic distribution of AGO1. Recent reports showed further evidence for a role of Argonaute in alternative splicing (Ameyar-Zazoua et al., 2012; Taliaferro et al., 2013). In this work we found that AGO1 regulates constitutive and alternative splicing genome-wide in MCF7 cells. Our results suggested that AGO1 presence can inhibit intron excision through a direct mechanism. Moreover, we found exon skipping events regulated upon AGO1 knockdown that were correlated with AGO1 presence. In addition, AGO1 was enriched genome-wide with silencing heterochromatin marks (H3K9me2 and H3K27me3). Our results showed AGO1 regulated alternative splicing in MCF7 cells and several evidences suggested that the mechanism could act through sRNA directed TGS.

Recently, Ameyar-Zazoua et al found that Argonaute proteins localize in the nucleus with spliceosome components and that AGO1 and AGO2 recruitment leads to local H3K9 methylation affecting the alternative splicing regulation of the CD44 gene model. However, the origin of the sRNAs implicated in the process is still unclear, Argonaute recruitment is RNAi dependent. On the other hand, recent results showed that *Drosophila* AGO2 regulates alternative splicing but by a direct association through chromatin and not through sRNA mediated transcriptional gene silencing (Taliaferro et al., 2013).

We analyzed the distribution of AGO1 in MCF7 and MCF10A cells and we found a non-random distribution with significant enrichment in specific regions. AGO1 was significantly enriched in CpG islands (CGI) and the 5'UTRs. Recently, Taliaferro et al showed the distribution of *Drosophila* AGO2, where they found enrichment at promoter regions (Taliaferro et al.,

2013). However, when we looked to promoter regions we did not find a significant enrichment. The AGO enrichment at promoters (Taliaferro et al., 2013) may be explained by an accumulation on the 5'UTRs that in our results was also correlated with highly expressed genes in MCF7. In addition, the significant enrichment found in the first part of the first intron was consistent with a regulation through the first 5'ss. There is evidence of alternative splicing regulation associated with the position of the first 5'ss (Bieberstein et al., 2012), AGO1 localized in this region could suggest a downstream effect on transcription dependent alternative splicing. The distribution of AGO1 showed the same specific regions enriched in MCF7 and MCF10A. However, only MCF7 (tumoral breast cancer cell line) showed AGO1 enrichment on CGIs, which may suggest cell specific differences due to the tumoral origin of MCF7. On the other hand, we found a higher signal of AGO1 around the TSS in highly expressed genes in MCF7, suggesting a relation of AGO1 with activation of transcription. The overlap between pairs of histone modifications was higher in AGO1 binding sites specially for the repressive marks H3K9me2 and H3K27me3, in agreement with our results in (Allo et al., 2009), which supported the role of AGO1 in gene silencing. However, we found that there was a weaker but detectable enrichment of H3K36me3 mark in AGO1 binding sites when compared to whole genome. This suggests relation to both activation and repression of transcription. Moreover, AGO1 presence in promoter regions was higher when it was localized with sRNAs from different sources, which supported a possible sRNA mediated regulation of transcription. In this sense, it has been found that active promoters could be sites of sense and antisense sRNA generation at canonical promoters (Core et al., 2008; He et al., 2008; Preker et al., 2008; Seila et al., 2009) or at intragenic promoters and enhancers (Kim et al., 2010; Kowalczyk et al., 2012), which may indicate that AGO1 presence around TSS with RNAPII in highly expressed genes might be associated to the loading of short RNAs to act elsewhere in trans.

Our previous results showed that there was evidence of endogenous sRNAs overlapping ASEs intronic regions and also predicted targets. However, we also found experimentally siRNAs targeting both introns and exons, which could suggest different action mechanisms for the sRNAs in the intragenic regions (Allo et al., 2009). Different reports found that there is evidence

of sRNAs involved with AGO complexes, suggesting their involvement in the regulation of the alternative exons (Ameyar-Zazoua et al., 2012). However, as we discussed before, the nature of these sRNAs in still unclear. It has been described that the origin of many sRNAs comes from the backtracking of the RNAPII near the transcription starts sites (TSS) (Fejes-Toth et al., 2009) and derived from sequences on the same strand as the TSS associated with promoters (Taft et al., 2009). Our comparison between tiRNAs (Taft et al., 2009) and pasRNAs with sRNAs from MCF7 (Mayr and Bartel, 2009), showed clearly their different nature. While tiRNAs and pasRNAs where localized in promoters, sRNAs in MCF7 were mainly found on introns. When we considered the sRNAs localized with AGO1 ChIP-Seq signal, the three classes showed higher enrichment in intronic regions, which suggested that AGO1 sRNAs complexes may preferentially localize on introns where they can regulate in some way intragenic processes, like RNAPII elongation. In this sense, Dumesic et al found that intron-containing genes showed siRNAs mapping to intronic and exon-intron junctions, which suggested that they could act as a template for siRNA synthesis (Dumesic et al., 2013). Moreover, they proposed an inverse correspondence between splicing efficiency and siRNA synthesis due to a kinetic mechanism based on the intronic structural features in *C. neoformans* (Dumesic et al., 2013). Which agreed with our results from (Allo et al., 2009), where we found different intronic structure on AGO1 regulated ASEs.

To address this and other questions we performed RNA-seq of MCF7 cells where AGO1 was knocked down by RNAi using an siRNA for luciferase as control. Even though, AGO1 depletion upregulates and downregulates expression of a considerable number of genes (1274), the proportion of genes with AGO1 clusters in their promoters that change their expression levels upon AGO1 knockdown is not enough to attribute to AGO1 a role in the control of transcriptional initiation, neither negative nor positive. Our results suggest that the presence of AGO1 in association to RNAPII around the TSS may reflect an important step for downstream effects, not related to transcriptional regulation at the initiation level. In that sense, recently reported evidence of Argonaute in alternative splicing from different reports and the localization with splicing factors, lead us to think that AGO1 function is alternative splicing is more general than expected. Our results showed

that alternative and constitutive ASEs events could be genome-wide. Furthermore, RNA-Seq revealed that AGO1 depletion promotes skipping of 334 and inclusion of 401 alternative cassette exons. RNA-seq data indicated that AGO1 regulated constitutive splicing negatively. If AGO1 was loading sRNAs, AGO1 depletion promoted intron excision may be related to the results reported from Dumesic et al, where splicing efficiency was affected inversely to the siRNA production (Dumesic et al., 2013).

Interestingly, we found that AGO1 depletion clearly correlated with splicing efficiency, where AGO1 density and alternative exons inclusion changes were related in the sense of exon skipping. Our previous report (Allo et al., 2009) suggested that one of the nuclear roles of AGO1 was the local inhibition of RNAPII elongation with the subsequent upregulation of fibronectin E33 inclusion according to the kinetic coupling model (Caceres and Kornblihtt, 2002). Ameyar-Zazoua et al provided more evidences in that direction (Ameyar-Zazoua et al., 2012) and Taliaferro et al found that AGO regulated splicing but not in an RNAi mediated transcriptional gene silencing (Taliaferro et al., 2013). Thus, we proposed that AGO1 regulation of both constitutive and alternative splicing may be mediated through two different ways, one through AGO1 localization to chromatin by its interaction to other factors (Moshkovich et al., 2011; Taliaferro et al., 2013; Cuddapah et al., 2009) and the other one through sRNA mediated chromatin formation (Allo et al., 2009; Guang et al., 2010; Ameyar-Zazoua et al., 2012). Our results suggest that for some of the regulated events endogenous sRNAs may be involved, and in some cases antisense to the gene direction. In summary, we provide here genome-wide evidence for a direct role of AGO1 in splicing efficiency and alternative splicing.

4.3 A chromatin code for cell specific alternative splicing

The results for the previous work provided some evidence for the role of AGO1 in splicing and alternative splicing control. We found that AGO1 correlated with different histone modifications and RNAPII. Additionally, we found ASEs that supported AGO1 dependent regulation of alternative splicing. The work discussed in this section, we have provided evidence for a chromatin code for splicing differences between MCF7 and MCF10A cell lines.

Even though the core histones are mainly the same throughout the chromatin structure, histone modifications act as markers for different chromatin states. Strahl and Allis proposed that some proteins, like heterochromatin protein 1 (HP1) or polycomb group proteins (PcG), could act as readers of the combination of different histone modifications to translate them into different states, which they named the histone code (Strahl and Allis, 2000). Subsequent genome-wide analysis of different chromatin proteins in *Drosophila* has shown that the genome can be defined by domains characterized by the different chromatin types (Filion et al., 2010). In addition, there is clear evidence of the histone modifications and its relation to alternative splicing regulation (Luco et al., 2010). In the case of splicing, while several evidences suggested a mechanism in which splicing would influence histone marking (Kolasinska-Zwierz et al., 2009; Tilgner et al., 2009; Schwartz et al., 2009; Spies et al., 2009), some reports showed equal exonic methylation in skipping and inclusion of specific exons (Huff et al., 2010). These in principle contradictory results, suggested that splicing changes based on differential combination of histone marks may be related to cell specific patterns and also to the association between histone marks with other factors.

Using ChIP-Seq data combined with alternative splicing arrays from the two cell lines, we have derived a chromatin RNA-map that shows a strong association between HP1α and CTCF activity around regulated exons, as

well as AGO1, RNAPII and histone marks. Our model shows that AGO1 activity downstream of exons, near the 3'ss, correlates with splicing changes in MCF7 compared to MCF10A patterns, providing further indication that AGO1 association to chromatin could be implicated in splicing regulation (Allo et al., 2009; Ameyar-Zazoua et al., 2012; Taliaferro et al., 2013), at least for some specific cases. Additionally, the strong association found between HP1α and CTCF appeared to be strongly directed in the sense of exon inclusion in MCF7, whereas the relation between CTCF and 5 methylated Cytosine (5metC) appeared to be antagonistic, as reported by (Shukla et al., 2011b). The results support the evidence of a genome-wide chromatin regulation of alternative splicing mediated by chromatin, in our case specific of splicing changes between MCF7 and MCF10A cell lines. The model also involve RNAPII and different factors that act as intermediates of different chromatin patterns that are associated to inclusion and skipping of the studied cassette exons.

Since it is still unclear how different histone marks are established and maintained at specific splicing events, we expanded our analysis about AGO1 role on alternative splicing to study a genome-wide splicing chromatin code. We thus, analyzed the code with the combination of the TGS-AS potential partners (AGO1,H3K27me3,H3K9me2,RNAPII and HP1α), markers associated with elongation (Li et al., 2007) or nucleosome occupancy at exons (Kolasinska-Zwierz et al., 2009; Tilgner et al., 2009; Schwartz et al., 2009; Spies et al., 2009) (H3K36me3), factors that were found recently to have a role in alternative splicing (CTCF) (Shukla et al., 2011b) and DNA modifications that appear to be enriched in heterochromatin (5metC). In this sense, we used a Machine Learning approach combining chromatin signals and ASEs, that lead us to be able to predict splicing and inclusion patterns for the 68% of the selected regulated cassette exons. The relative changes of chromatin-associated signals between MCF7 and MCF10A cells lines recapitulated regulatory modes previously described, as the indication that an increase of CTCF binding downstream of the regulated exons correlated with inclusion, and its decrease in combination with 5metC increase correlated with skipping (Shukla et al., 2011b). Moreover, we found association between CTCF and HP1α that was strongly related with inclusion of the cassette exons.

For RNAPII we found that the splicing change correlated better with the densities near the exon boundaries. Such is the case that the accumulation of RNAPII at junctions can reflect the interaction between transcription rates and alternative splicing efficiency (Kornblihtt, 2006). RNAPII was selected as one of the attributes in the model covering the downstream intron-exon junction (J4). There is evidence supporting a regulation of alternative splicing mRNA transcripts directly associated with RNAPII elongation rates (Nogues et al., 2002). In this sense, it has been found that higher RNAPII occupancy nearby exons is associated of alternative exons (Brodsky et al., 2005). Accordingly, our analysis indicated that a change in the region flanking the exon was indicative of the change in exon inclusion, in agreement with (Allo et al., 2009). Moreover, the change was also correlated with higher basal levels in included exons in MCF7.

H3K36me3 appeared as a very informative mark in our model. Several reports showed that H3K36me3 could be considered as an exon marker (Kolasinska-Zwierz et al., 2009; Tilgner et al., 2009; Schwartz et al., 2009; Spies et al., 2009). There is also evidence of higher densities of H3K36me3 at constitutive exons compared to alternative exons (Kolasinska-Zwierz et al., 2009; Hon et al., 2009), however the opposite pattern has been also described (de Almeida et al., 2011). For specific genes, an increased density of H3K36me3 has been related to exon skipping (Schor et al., 2009; Luco et al., 2010). Our model indicates that H3K36me3 is a crucial mark for splicing decisions. In our results the densities of H3K36me3 were not a consequence of gene expression. Thus, the H3K36me3 pattern observed near exon boundaries could correspond to a direct effect of splicing. Interestingly, skipping events showed higher H3K36me3 densities that also correlated with a higher change towards skipping, while RNAPII showed the contrary effect, even though the RNAPII signal found in inclusion events in the downstream junction (J4) was not very evident. The mean densities over the three exons boundaries showed accumulations of RNAPII, as mentioned before, when the exon was included, while in skipping events we found accumulation of H3K36me3. This agrees with recent reports showing that inhibition of RNAPII elongation can yield some kind of dependencie over other marks, particularly inclusion and skipping (Ip et al., 2011). As H3K36me3 affects RNAPII elongation, both effects are likely to be related. Moreover, H3K36me3 was found to

be dependent on splicing (de Almeida et al., 2011; Kim et al., 2011).

Interestingly, we found a strong association between CTCF and HP1α that also correlated with our chromatin map results. Both, CTCF and HP1α, were related to inclusion and the densities around the three exons of the events showed a colocalization of both signals, specially in inclusion events. HP1 recognizes methylated H3K9 (Bannister et al., 2001), is responsible for the spreading and maintaining of heterochromatin formation (Ayyanathan et al., 2003) and is a key player in the transcriptional gene silencing (TGS) pathway (Moazed, 2009). Moreover, HP1 also participates in the TGS-mediated regulation of alternative splicing (Allo et al., 2009; Ameyar-Zazoua et al., 2012), and its accumulation has been observed to correlate with the inclusion of alternative exons (Allo et al., 2009; Saint-Andre et al., 2011; Ameyar-Zazoua et al., 2012). Even though we did not find significant signal of heterochromatin marks correlated with HP1α, possibly because in general H3K27me3 and H3K9me2 did not seem to have enough informative signal in our model, HP1 has been found to interplay directly or indirectly with other non-histone proteins (Kwon and Workman, 2008). There is also evidence that *Drosophila* HP1a binds more strongly to methylated H3 while it shows a weak binding to methylated chromatin, suggesting that HP1a may have another binding to chromatin besides H3K9 (Kwon and Workman, 2011). On the other hand, CTCF has been implicated in diverse functions related to the global organization of chromatin (Phillips and Corces, 2009). Besides acting as insulator, it also works as a barrier for spreading of heterochromatin (Cuddapah et al., 2009) and shukla et al showed CTCF involvement in alternative splicing regulation (Shukla et al., 2011b) as antagonistic to methylation.

Our results showed that apart from the association between CTCF and HP1α, both were reciprocally associated to methylation. We proposed based on our chromatin RNA map, that HP1α signals downstream and upstream of the cassette exon were localized with CTCF and could inhibit in some way RNAPII activity, since non-regulated cassette exons showed high levels of methylation but absence of CTCF, HP1α and RNAPII signals. The reciprocal association with methylation suggests that intragenic CTCF HP1α interaction and possibly binding could be affected by variations in basal DNA methylation. It has been shown that changes in DNA methylation can lead

to alternative splicing regulation changes in tissue or cell specific manner (David and Manley, 2010).

AGO1 was selected as an informative attribute in our model, which suggested that a fraction of the events could be explained by AGO1 presence. Moreover, we found association of AGO1 with CTCF and AGO1 with HP1α, but not the other way around. AGO1 showed correlation in the change of inclusion. However, comparison of the changes with CTCF and HP1α showed that there was not an increase in the fraction of inclusion events when AGO1 was upregulated. Thus, the results suggest that more than leading to inclusion, AGO1 inhibits exon skipping, in agreement with our previous results. We propose that while CTCF and HP1α are directly associated together with inclusion of the exon, AGO1 absence leads to a higher skipping of the exon. This combination of factors would explain the association of AGO1 with CTCF and HP1α in inclusion events.

Interestingly, we found consistent DNA binding motifs for HP1α and AGO1 and recovered the known CTCF motif (Essien et al., 2009). Although, there is no clear evidence of AGO1 binding to DNA, Taliaferro et al found *Drosophila* AGO2 binding sites that were consistent with a previously reported motif for AGO2 (Moshkovich et al., 2011). However, they suggested that AGO would mainly colocalize with other DNA binding factors (Taliaferro et al., 2013). Our AGO1 motif is also consistent throughout our data and is similar to the motif of (Moshkovich et al., 2011). In addition, the G-rich pattern of the motif, as suggested by Ameyar-Zazoua et al, could refer to sequences similar to those that allow AGO binding to mRNA in the cytoplasm. We hypothesize, that enrichment of sequences with the AGO1 motif localized in inclusion events would show a non RNAi mediated regulation of alternative splicing mechanism without a binding between AGO1 and chromatin. However, contradictory results showed that AGO regulated alternative splicing could require the presence of sRNAs (Allo et al., 2009; Ameyar-Zazoua et al., 2012; Dumesic et al., 2013). Additionally, HP1α motif could also show controversy since there is no known DNA binding described. HP1 proteins consist of two main domains that are separated by a hinge region, while one of the domains is known to bind H3K9me the other acts with interaction of other proteins (Kwon and Workman, 2011). Although there is

no defined DNA binding region, the hinge linker domain was found to be involved in length dependent DNA and RNA binding (Zhao et al., 2000).

There have been attempts to systematically establish a relation between histone marks and splicing regulation (Dhami et al., 2010; Hon et al., 2009; Enroth et al., 2012; Zhou et al., 2012). However, only in one case a predictive model has been proposed (Enroth et al., 2012). Our model, which is based on continuous relative changes, can explain about 68% of the splicing changes. Earlier approaches have analyzed the relation between chromatin and splicing looking at one single condition at the time (Dhami et al., 2010; Hon et al., 2009; Enroth et al., 2012; Zhou et al., 2012), rather than comparing between two conditions. In previous methods exons were generally classified as constitutive or alternative based on expression data from one single condition. Our method showed the advantage that, by comparing two conditions, besides circumventing the caveats of comparing genomic regions with different local biases, we could relate changes of the chromatin signal between two conditions to the actual splicing change of exons between the same two conditions. We thus propose that the relative change in the histone mark or any other chromatin-related signal provides a better descriptor of the association between chromatin and splicing regulation. Our analysis also shows an strong association of HP1α and CTCF in the chromatin dependent regulation of alternative splicing as further evidence of AGO1 regulated alternative splicing events. We showed that the chromatin code is defined by the combination of different marks and factors that differentiate patterns around skipping, inclusion and non-regulated exons. Additionally, we report a possible DNA binding motif for HP1α and AGO1.

CHAPTER 5

Conclusions

1. There is evidence of alternative splicing events regulated upon AGO1 depletion. These events show shorter upstream introns, which suggests that RNAPII elongation changes may be a response on different genic structural features.

2. There is bioinformatical evidence of miRNAs overlapping and targeting introns flanking alternative exons. Moreover, there is evidence of AGO1 dependent alternative splicing events with predicted miRNA targets.

3. Some of the alternative splicing events are AGO1 dependent but DCR independent, which suggests that the possible transcriptional gene silencing mediated regulation of alternative splicing, involves sRNAs other than siRNAs and miRNAs.

4. AGO1 follows a non-random distribution in the genome and shows enrichment in specific regions in different cells lines. While MCF7 shows enrichment over CpG islands and 5'UTRs, MCF10A only shows enrichment over the 5'UTR. Interestingly there is a high enrichment in the first 300 nt of the first introns, AGO1 which suggests a downstream effect on transcription dependent alternative splicing.

5. AGO1 localizes with sRNAs from different classes that are mainly present in intronic regions. There is a considerable proportion of sRNAs that overlap in antisense to regulated genes and mainly colocalize in the first introns. There is also overlap between AGO1 and sRNAs in the promoters and 5'UTRs. The results suggest that active promoters with AGO1 could be sites of sense and antisense sRNA generation.

6. AGO1 colocalizes with silencing histone modifications H3K27me3 and H3K9me2, which would suggest a gene silencing role of AGO1. Howevwe, there is a detectable enrichment of AGO1 with H3K36me3, suggesting that AGO1 activity requires active transcription. Moreover, there is AGO1 presence around TSSs with RNAPII in highly expressed genes that might be associated to the loading of short RNAs to act elsewhere in trans.

7. The proportion of genes with AGO1 clusters in their promoters that

change their expression levels upon AGO1 knockdown is not enough to attribute to AGO1 a role in the control of transcription initiation, neither negative nor positive.

8. RNA-Seq revealed that AGO1 depletion promotes skipping and inclusion of 334 and 401 alternative exons respectively. Additionally, RNA-seq data indicated that AGO1 regulated constitutive splicing negatively. AGO1 depletion correlates with splicing efficiency, where AGO1 density and alternative exons inclusion changes are related in the sense of exon skipping.

9. For some of the alternative splicing events endogenous sRNAs may be involved, in some cases antisense to the gene direction. This suggests an endogenous sRNA mediated regulation of alternative splicing dependent of AGO1.

10. We have provided evidence for a chromatin code associated to splicing differences between MCF7 and MCF10A cell lines. We have derived a chromatin RNA-map that shows a strong association between HP1α and CTCF activity around regulated exons, as well as some association of AGO1, RNAPII and histone marks.

11. AGO1 activity in the downstream intron, near the 3'ss, correlates with splicing changes in MCF7 compared to MCF10A patterns, providing further indication that AGO1 association to chromatin could be implicated in splicing regulation.

12. There is a strong reciprocal interplay between HP1α and CTCF, directed in the sense of exon inclusion in MCF7, whereas the relation between CTCF and 5metC is antagonistic. The results show a genome-wide association of chromatin and alternative splicing and that explains part of the inclusion and skipping changes between MCF7 and MCF10A change.

13. We report possible DNA binding motifs for HP1α and AGO1, and we recover the previously known DNA binding motif for CTCF. The AGO1 motif suggests a non RNAi mediated regulation of alternative splicing mechanism with a direct binding of AGO1 to DNA.

14. Our model is based on continuous relative changes and can explain about 68% of the splicing changes. We propose that the relative change in the histone marks or any other chromatin-related signal provides a better descriptor of the association between chromatin and splicing regulation than using a singgle condition. We show that the chromatin code is define by the combination of different marks and factors that differentiate patterns of skipping, inclusion and non-regulated exons.

References

Adami, G. and Babiss, L. E. (1991). DNA template effect on RNA splicing: two copies of the same gene in the same nucleus are processed differently. *EMBO J.*, 10(11):3457–3465. [PubMed Central:PMC453074] [PubMed:1915302].

Agirre, E. and Eyras, E. (2011). Databases and resources for human small non-coding RNAs. *Hum. Genomics*, 5(3):192–199. [PubMed:21504869].

Ahlenstiel, C. L., Lim, H. G., Cooper, D. A., Ishida, T., Kelleher, A. D., and Suzuki, K. (2012). Direct evidence of nuclear Argonaute distribution during transcriptional silencing links the actin cytoskeleton to nuclear RNAi machinery in human cells. *Nucleic Acids Res.*, 40(4):1579–1595. [PubMed Central:PMC3287199] [DOI:10.1093/nar/gkr891] [PubMed:22064859].

Allo, M., Buggiano, V., Fededa, J. P., Petrillo, E., Schor, I., de la Mata, M., Agirre, E., Plass, M., Eyras, E., Elela, S. A., Klinck, R., Chabot, B., and Kornblihtt, A. R. (2009). Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat. Struct. Mol. Biol.*, 16(7):717–724. [DOI:10.1038/nsmb.1620] [PubMed:19543290].

Allo, M. and Kornblihtt, A. R. (2010). Gene silencing: small RNAs control RNA polymerase II elongation. *Curr. Biol.*, 20(17):R704–707. [DOI:10.1016/j.cub.2010.07.013] [PubMed:20833310].

Allo, M., Schor, I. E., Munoz, M. J., de la Mata, M., Agirre, E., Valcarcel, J., Eyras, E., and Kornblihtt, A. R. (2010). Chromatin and alternative splicing. *Cold Spring Harb. Symp. Quant. Biol.*, 75:103–111. [DOI:10.1101/sqb.2010.75.023] [PubMed:21289049].

Althammer, S., Gonzalez-Vallinas, J., Ballare, C., Beato, M., and Eyras, E. (2011). Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data. *Bioinformatics*, 27(24):3333–3340. [PubMed Central:PMC3232367] [DOI:10.1093/bioinformatics/btr570] [PubMed:21994224].

Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllensten, U., Cavelier, L., and Feuk, L. (2011). Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat. Struct. Mol. Biol.*, 18(12):1435–1440. [DOI:10.1038/nsmb.2143] [PubMed:22056773].

Ameyar-Zazoua, M., Rachez, C., Souidi, M., Robin, P., Fritsch, L., Young, R., Morozova, N., Fenouil, R., Descostes, N., Andrau, J. C., Mathieu, J., Hamiche, A., Ait-Si-Ali, S., Muchardt, C., Batsche, E., and Harel-Bellan, A. (2012). Argonaute proteins couple chromatin silencing to alternative splicing. *Nat. Struct. Mol. Biol.* [DOI:10.1038/nsmb.2373] [PubMed:22961379].

Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C., and Komorowski, J. (2009). Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.*, 19(10):1732–1741. [PubMed Central:PMC2765275] [DOI:10.1101/gr.092353.109] [PubMed:19687145].

Auboeuf, D., Honig, A., Berget, S. M., and O'Malley, B. W. (2002). Coordinate regulation of transcription and splicing by steroid receptor coregulators. *Science*, 298(5592):416–419. [DOI:10.1126/science.1073734] [PubMed:12376702].

Audic, S. and Claverie, J. M. (1997). The significance of digital gene expression profiles. *Genome Res.*, 7(10):986–995. [PubMed:9331369].

Ayyanathan, K., Lechner, M. S., Bell, P., Maul, G. G., Schultz, D. C., Yamada, Y., Tanaka, K., Torigoe, K., and Rauscher, F. J. (2003). Regulated

recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. *Genes Dev.*, 17(15):1855–1869. [PubMed Central:PMC196232] [DOI:10.1101/gad.1102803] [PubMed:12869583].

Bailey, T. L. and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*, 2:28–36. [PubMed:7584402].

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*, 410(6824):120–124. [DOI:10.1038/35065138] [PubMed:11242054].

Batsche, E., Yaniv, M., and Muchardt, C. (2006). The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat. Struct. Mol. Biol.*, 13(1):22–29. [DOI:10.1038/nsmb1030] [PubMed:16341228].

Bauren, G. and Wieslander, L. (1994). Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell*, 76(1):183–192. [PubMed:8287477].

Beckmann, J. S. and Trifonov, E. N. (1991). Splice junctions follow a 205-base ladder. *Proc. Natl. Acad. Sci. U.S.A.*, 88(6):2380–2383. [PubMed Central:PMC51235] [PubMed:2006175].

Berget, S. M. (1995). Exon recognition in vertebrate splicing. *J. Biol. Chem.*, 270(6):2411–2414. [PubMed:7852296].

Berget, S. M., Moore, C., and Sharp, P. A. (1977). Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. U.S.A.*, 74(8):3171–3175. [PubMed Central:PMC431482] [PubMed:269380].

Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409(6818):363–366. [DOI:10.1038/35053110] [PubMed:11201747].

Beyer, A. L. and Osheim, Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.*, 2(6):754–765. [PubMed:3138163].

Bieberstein, N. I., Carrillo Oesterreich, F., Straube, K., and Neugebauer, K. M. (2012). First exon length controls active chromatin signatures and transcription. *Cell Rep*, 2(1):62–68. [DOI:10.1016/j.celrep.2012.05.019] [PubMed:22840397].

Blanchette, M. and Chabot, B. (1997). A highly stable duplex structure sequesters the 5′ splice site region of hnRNP A1 alternative exon 7B. *RNA*, 3(4):405–419. [PubMed Central:PMC1369492] [PubMed:9085847].

Boguski, M. S., Lowe, T. M., and Tolstoshev, C. M. (1993). dbEST–database for "expressed sequence tags". *Nat. Genet.*, 4(4):332–333. [DOI:10.1038/ng0893-332] [PubMed:8401577].

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J.*, 17(1):170–180. [PubMed Central:PMC1170368] [DOI:10.1093/emboj/17.1.170] [PubMed:9427751].

Boutz, P. L., Chawla, G., Stoilov, P., and Black, D. L. (2007). MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. *Genes Dev.*, 21(1):71–84. [PubMed Central:PMC1759902] [DOI:10.1101/gad.1500707] [PubMed:17210790].

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978). Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. U.S.A.*, 75(10):4853–4857. [PubMed Central:PMC336219] [PubMed:283395].

Breathnach, R. and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.*, 50:349–383. [DOI:10.1146/annurev.bi.50.070181.002025] [PubMed:6791577].

Breitbart, R. E., Andreadis, A., and Nadal-Ginard, B. (1987). Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu. Rev. Biochem.*, 56:467–495. [DOI:10.1146/annurev.bi.56.070187.002343] [PubMed:3304142].

Brodsky, A. S., Meyer, C. A., Swinburne, I. A., Hall, G., Keenan, B. J., Liu, X. S., Fox, E. A., and Silver, P. A. (2005). Genomic mapping of RNA

polymerase II reveals sites of co-transcriptional regulation in human cells. *Genome Biol.*, 6(8):R64. [PubMed Central:PMC1273631] [DOI:10.1186/gb-2005-6-8-r64] [PubMed:16086846].

Brody, Y., Neufeld, N., Bieberstein, N., Causse, S. Z., Bohnlein, E. M., Neugebauer, K. M., Darzacq, X., and Shav-Tal, Y. (2011). The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS Biol.*, 9(1):e1000573. [PubMed Central:PMC3019111] [DOI:10.1371/journal.pbio.1000573] [PubMed:21264352].

Buhler, M. and Moazed, D. (2007). Transcription and RNAi in heterochromatic gene silencing. *Nat. Struct. Mol. Biol.*, 14(11):1041–1048. [DOI:10.1038/nsmb1315] [PubMed:17984966].

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. *Mol. Cell*, 36(4):541–546. [PubMed Central:PMC3232742] [DOI:10.1016/j.molcel.2009.10.019] [PubMed:19941815].

Buratti, E. and Baralle, F. E. (2004). Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol. Cell. Biol.*, 24(24):10505–10514. [PubMed Central:PMC533984] [DOI:10.1128/MCB.24.24.10505-10514.2004] [PubMed:15572659].

Caceres, J. F. and Kornblihtt, A. R. (2002). Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.*, 18(4):186–193. [PubMed:11932019].

Carmell, M. A., Xuan, Z., Zhang, M. Q., and Hannon, G. J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.*, 16(21):2733–2742. [DOI:10.1101/gad.1026102] [PubMed:12414724].

Carthew, R. W. and Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell*, 136(4):642–655. [PubMed Central:PMC2675692] [DOI:10.1016/j.cell.2009.01.035] [PubMed:19239886].

Castanotto, D., Tommasi, S., Li, M., Li, H., Yanow, S., Pfeifer, G. P., and Rossi, J. J. (2005). Short hairpin RNA-directed cytosine (CpG) methylation

of the RASSF1A gene promoter in HeLa cells. *Mol. Ther.*, 12(1):179–183. [DOI:10.1016/j.ymthe.2005.03.003] [PubMed:15963934].

Castel, S. E. and Martienssen, R. A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat. Rev. Genet.*, 14(2):100–112. [DOI:10.1038/nrg3355] [PubMed:23329111].

Catterall, J. F., O'Malley, B. W., Robertson, M. A., Staden, R., Tanaka, Y., and Brownlee, G. G. (1978). Nucleotide sequence homology at 12 intron–exon junctions in the chick ovalbumin gene. *Nature*, 275(5680):510–513. [PubMed:692731].

Cellini, A., Felder, E., and Rossi, J. J. (1986). Yeast pre-messenger RNA splicing efficiency depends on critical spacing requirements between the branch point and 3' splice site. *EMBO J.*, 5(5):1023–1030. [PubMed Central:PMC1166896] [PubMed:3013610].

Cernilogar, F. M., Onorati, M. C., Kothe, G. O., Burroughs, A. M., Parsi, K. M., Breiling, A., Lo Sardo, F., Saxena, A., Miyoshi, K., Siomi, H., Siomi, M. C., Carninci, P., Gilmour, D. S., Corona, D. F., and Orlando, V. (2011). Chromatin-associated RNA interference components contribute to transcriptional regulation in Drosophila. *Nature*, 480(7377):391–395. [DOI:10.1038/nature10492] [PubMed:22056986].

Chen, M. and Manley, J. L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.*, 10(11):741–754. [PubMed Central:PMC2958924] [DOI:10.1038/nrm2777] [PubMed:19773805].

Chi, S. W., Zang, J. B., Mele, A., and Darnell, R. B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, 460(7254):479–486. [PubMed Central:PMC2733940] [DOI:10.1038/nature08170] [PubMed:19536157].

Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, 12(1):1–8. [PubMed:902310].

Colgan, D. F. and Manley, J. L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes Dev.*, 11(21):2755–2766. [PubMed:9353246].

Core, L. J., Waterfall, J. J., and Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 322(5909):1845–1848. [PubMed Central:PMC2833333] [DOI:10.1126/science.1162228] [PubMed:19056941].

Corvelo, A. and Eyras, E. (2008). Exon creation and establishment in human genes. *Genome Biol.*, 9(9):R141. [PubMed Central:PMC2592719] [DOI:10.1186/gb-2008-9-9-r141] [PubMed:18811936].

Cramer, P., Caceres, J. F., Cazalla, D., Kadener, S., Muro, A. F., Baralle, F. E., and Kornblihtt, A. R. (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. [PubMed:10488340].

Cramer, P., Pesce, C. G., Baralle, F. E., and Kornblihtt, A. R. (1997). Functional association between promoter structure and transcript alternative splicing. *Proc. Natl. Acad. Sci. U.S.A.*, 94(21):11456–11460. [PubMed Central:PMC23504] [PubMed:9326631].

Cuddapah, S., Jothi, R., Schones, D. E., Roh, T. Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.*, 19(1):24–32. [PubMed Central:PMC2612964] [DOI:10.1101/gr.082800.108] [PubMed:19056695].

David, C. J. and Manley, J. L. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.*, 24(21):2343–2364. [PubMed Central:PMC2964746] [DOI:10.1101/gad.1973010] [PubMed:21041405].

de Almeida, S. F. and Carmo-Fonseca, M. (2010). Cotranscriptional RNA checkpoints. *Epigenomics*, 2(3):449–455. [DOI:10.2217/epi.10.21] [PubMed:22121903].

de Almeida, S. F., Grosso, A. R., Koch, F., Fenouil, R., Carvalho, S., Andrade, J., Levezinho, H., Gut, M., Eick, D., Gut, I., Andrau, J. C., Ferrier, P., and

Carmo-Fonseca, M. (2011). Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat. Struct. Mol. Biol.*, 18(9):977–983. [DOI:10.1038/nsmb.2123] [PubMed:21792193].

de la Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A. R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell*, 12(2):525–532. [PubMed:14536091].

de Wit, E., Greil, F., and van Steensel, B. (2007). High-resolution mapping reveals links of HP1 with active and inactive chromatin components. *PLoS Genet.*, 3(3):e38. [PubMed Central:PMC1808074] [DOI:10.1371/journal.pgen.0030038] [PubMed:17335352].

Dhami, P., Saffrey, P., Bruce, A. W., Dillon, S. C., Chiang, K., Bonhoure, N., Koch, C. M., Bye, J., James, K., Foad, N. S., Ellis, P., Watkins, N. A., Ouwehand, W. H., Langford, C., Andrews, R. M., Dunham, I., and Vetrie, D. (2010). Complex exon-intron marking by histone modifications is not determined solely by nucleosome distribution. *PLoS ONE*, 5(8):e12339. [PubMed Central:PMC2925886] [DOI:10.1371/journal.pone.0012339] [PubMed:20808788].

Ding, L. and Han, M. (2007). GW182 family proteins are crucial for microRNA-mediated gene silencing. *Trends Cell Biol.*, 17(8):411–416. [DOI:10.1016/j.tcb.2007.06.003] [PubMed:17766119].

Dower, K. and Rosbash, M. (2002). T7 RNA polymerase-directed transcripts are processed in yeast and link 3' end formation to mRNA nuclear export. *RNA*, 8(5):686–697. [PubMed Central:PMC1370288] [PubMed:12022234].

Dreyfuss, G., Matunis, M. J., Pinol-Roma, S., and Burd, C. G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.*, 62:289–321. [DOI:10.1146/annurev.bi.62.070193.001445] [PubMed:8352591].

Dumesic, P. A., Natarajan, P., Chen, C., Drinnenberg, I. A., Schiller, B. J., Thompson, J., Moresco, J. J., Yates, J. R., Bartel, D. P., and Madhani, H. D. (2013). Stalled spliceosomes are a signal for RNAi-mediated genome defense. *Cell*, 152(5):957–968. [DOI:10.1016/j.cell.2013.01.046] [PubMed:23415457].

Egloff, S., Dienstbier, M., and Murphy, S. (2012). Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends Genet.*, 28(7):333–341. [DOI:10.1016/j.tig.2012.03.007] [PubMed:22622228].

Elkayam, E., Kuhn, C. D., Tocilj, A., Haase, A. D., Greene, E. M., Hannon, G. J., and Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. *Cell*, 150(1):100–110. [DOI:10.1016/j.cell.2012.05.017] [PubMed:22682761].

Enerly, E., Sheng, Z., and Li, K. B. (2005). Natural antisense as potential regulator of alternative initiation, splicing and termination. *In Silico Biol. (Gedrukt)*, 5(4):367–377. [PubMed:16268781].

Enroth, S., Bornelov, S., Wadelius, C., and Komorowski, J. (2012). Combinations of histone modifications mark exon inclusion levels. *PLoS ONE*, 7(1):e29911. [PubMed Central:PMC3252363] [DOI:10.1371/journal.pone.0029911] [PubMed:22242188].

Eperon, L. P., Graham, I. R., Griffiths, A. D., and Eperon, I. C. (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. [PubMed:2840206].

Essien, K., Vigneau, S., Apreleva, S., Singh, L. N., Bartolomei, M. S., and Hannenhalli, S. (2009). CTCF binding site classes exhibit distinct evolutionary, genomic, epigenomic and transcriptomic features. *Genome Biol.*, 10(11):R131. [PubMed Central:PMC3091324] [DOI:10.1186/gb-2009-10-11-r131] [PubMed:19922652].

Fejes-Toth, K., Sotirova, V., Sachidanandam, R., Assaf, G., Hannon, G. J., Kapranov, P., Foissac, S., Willingham, A. T., Duttagupta, R., Dumais, E., and Gingeras, T. R. (2009). Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. *Nature*, 457(7232):1028–1032. [PubMed Central:PMC2719882] [DOI:10.1038/nature07759] [PubMed:19169241].

Filion, G. J., van Bemmel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D., Brugman, W., de Castro, I. J., Kerkhoven, R. M., Bussemaker,

H. J., and van Steensel, B. (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell*, 143(2):212–224. [PubMed Central:PMC3119929] [DOI:10.1016/j.cell.2010.09.009] [PubMed:20888037].

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, 391(6669):806–811. [DOI:10.1038/35888] [PubMed:9486653].

Gagnon, K. T. and Corey, D. R. (2012). Argonaute and the nuclear RNAs: new pathways for RNA-mediated control of gene expression. *Nucleic Acid Ther*, 22(1):3–16. [PubMed Central:PMC3318256] [DOI:10.1089/nat.2011.0330] [PubMed:22283730].

Gonzalez, S., Pisano, D. G., and Serrano, M. (2008). Mechanistic principles of chromatin remodeling guided by siRNAs and miRNAs. *Cell Cycle*, 7(16):2601–2608. [PubMed:18719372].

Gornemann, J., Kotovic, K. M., Hujer, K., and Neugebauer, K. M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol. Cell*, 19(1):53–63. [DOI:10.1016/j.molcel.2005.05.007] [PubMed:15989964].

Gowher, H., Brick, K., Camerini-Otero, R. D., and Felsenfeld, G. (2012). Vezf1 protein binding sites genome-wide are associated with pausing of elongating RNA polymerase II. *Proc. Natl. Acad. Sci. U.S.A.*, 109(7):2370–2375. [PubMed Central:PMC3289347] [DOI:10.1073/pnas.1121538109] [PubMed:22308494].

Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. *RNA*, 6(9):1197–1211. [PubMed Central:PMC1369994] [PubMed:10999598].

Graveley, B. R. (2001). Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.*, 17(2):100–107. [PubMed:11173120].

Green, V. A. and Weinberg, M. S. (2011). Small RNA-induced transcriptional gene regulation in mammals mechanisms, therapeutic applica-

tions, and scope within the genome. *Prog Mol Biol Transl Sci*, 102:11–46. [DOI:10.1016/B978-0-12-415795-8.00005-2] [PubMed:21846568].

Grewal, S. I. and Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. *Science*, 301(5634):798–802. [DOI:10.1126/science.1086887] [PubMed:12907790].

Griffiths-Jones, S. (2004). The microRNA Registry. *Nucleic Acids Res.*, 32(Database issue):D109–111. [PubMed Central:PMC308757] [DOI:10.1093/nar/gkh023] [PubMed:14681370].

Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., and Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, 34(Database issue):D140–144. [PubMed Central:PMC1347474] [DOI:10.1093/nar/gkj112] [PubMed:16381832].

Griffiths-Jones, S., Saini, H. K., van Dongen, S., and Enright, A. J. (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, 36(Database issue):D154–158. [PubMed Central:PMC2238936] [DOI:10.1093/nar/gkm952] [PubMed:17991681].

Guang, S., Bochner, A. F., Burkhart, K. B., Burton, N., Pavelec, D. M., and Kennedy, S. (2010). Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature*, 465(7301):1097–1101. [PubMed Central:PMC2892551] [DOI:10.1038/nature09095] [PubMed:20543824].

Hamilton, A. J. and Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, 286(5441):950–952. [PubMed:10542148].

Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R., and Hannon, G. J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science*, 293(5532):1146–1150. [DOI:10.1126/science.1064023] [PubMed:11498593].

Haussecker, D. and Proudfoot, N. J. (2005). Dicer-dependent turnover of intergenic transcripts from the human beta-globin gene cluster.

Mol. Cell. Biol., 25(21):9724–9733. [PubMed Central:PMC1265824] [DOI:10.1128/MCB.25.21.9724-9733.2005] [PubMed:16227618].

He, Y., Vogelstein, B., Velculescu, V. E., Papadopoulos, N., and Kinzler, K. W. (2008). The antisense transcriptomes of human cells. *Science*, 322(5909):1855–1857. [PubMed Central:PMC2824178] [DOI:10.1126/science.1163853] [PubMed:19056939].

Hertel, K. J. (2008). Combinatorial control of exon recognition. *J. Biol. Chem.*, 283(3):1211–1215. [DOI:10.1074/jbc.R700035200] [PubMed:18024426].

Hiragami-Hamada, K., Shinmyozu, K., Hamada, D., Tatsu, Y., Uegaki, K., Fujiwara, S., and Nakayama, J. (2011). N-terminal phosphorylation of HP1alpha promotes its chromatin binding. *Mol. Cell. Biol.*, 31(6):1186–1200. [PubMed Central:PMC3067897] [DOI:10.1128/MCB.01012-10] [PubMed:21245376].

Hirose, Y., Tacke, R., and Manley, J. L. (1999). Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev.*, 13(10):1234–1239. [PubMed Central:PMC316731] [PubMed:10346811].

Hock, J. and Meister, G. (2008). The Argonaute protein family. *Genome Biol.*, 9(2):210. [PubMed Central:PMC2374724] [DOI:10.1186/gb-2008-9-2-210] [PubMed:18304383].

Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., and Bustamante, C. (2009). Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science*, 325(5940):626–628. [PubMed Central:PMC2775800] [DOI:10.1126/science.1172926] [PubMed:19644123].

Hon, G., Wang, W., and Ren, B. (2009). Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput. Biol.*, 5(11):e1000566. [PubMed Central:PMC2775352] [DOI:10.1371/journal.pcbi.1000566] [PubMed:19918365].

Huang, V., Place, R. F., Portnoy, V., Wang, J., Qi, Z., Jia, Z., Yu, A., Shuman, M., Yu, J., and Li, L. C. (2012). Upregulation of Cyclin B1 by miRNA and its implications in cancer. *Nucleic Acids Res.*, 40(4):1695–1707. [PubMed Central:PMC3287204] [DOI:10.1093/nar/gkr934] [PubMed:22053081].

Hubbard, T. J., Aken, B. L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Clarke, L., Coates, G., Fairley, S., Fitzgerald, S., Fernandez-Banet, J., Gordon, L., Graf, S., Haider, S., Hammond, M., Holland, R., Howe, K., Jenkinson, A., Johnson, N., Kahari, A., Keefe, D., Keenan, S., Kinsella, R., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Megy, K., Meidl, P., Overduin, B., Parker, A., Pritchard, B., Rios, D., Schuster, M., Slater, G., Smedley, D., Spooner, W., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wilder, S., Zadissa, A., Birney, E., Cunningham, F., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Herrero, J., Kasprzyk, A., Proctor, G., Smith, J., Searle, S., and Flicek, P. (2009). Ensembl 2009. *Nucleic Acids Res.*, 37(Database issue):D690–697. [PubMed Central:PMC2686571] [DOI:10.1093/nar/gkn828] [PubMed:19033362].

Huff, J. T., Plocik, A. M., Guthrie, C., and Yamamoto, K. R. (2010). Reciprocal intronic and exonic histone modification regions in humans. *Nat. Struct. Mol. Biol.*, 17(12):1495–1499. [PubMed Central:PMC3057557] [DOI:10.1038/nsmb.1924] [PubMed:21057525].

Hui, J., Hung, L. H., Heiner, M., Schreiner, S., Neumuller, N., Reither, G., Haas, S. A., and Bindereif, A. (2005). Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. *EMBO J.*, 24(11):1988–1998. [PubMed Central:PMC1142610] [DOI:10.1038/sj.emboj.7600677] [PubMed:15889141].

Ingelbrecht, I., Van Houdt, H., Van Montagu, M., and Depicker, A. (1994). Posttranscriptional silencing of reporter transgenes in tobacco correlates with DNA methylation. *Proc. Natl. Acad. Sci. U.S.A.*, 91(22):10502–10506. [PubMed Central:PMC45049] [PubMed:7937983].

Ip, J. Y., Schmidt, D., Pan, Q., Ramani, A. K., Fraser, A. G., Odom, D. T., and Blencowe, B. J. (2011). Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome Res.*, 21(3):390–401. [PubMed Central:PMC3044853] [DOI:10.1101/gr.111070.110] [PubMed:21163941].

Irvine, D. V., Zaratiegui, M., Tolia, N. H., Goto, D. B., Chitwood, D. H., Vaughn, M. W., Joshua-Tor, L., and Martienssen, R. A. (2006). Argonaute

slicing is required for heterochromatic silencing and spreading. *Science*, 313(5790):1134–1137. [DOI:10.1126/science.1128813] [PubMed:16931764].

Janowski, B. A., Huffman, K. E., Schwartz, J. C., Ram, R., Nordsell, R., Shames, D. S., Minna, J. D., and Corey, D. R. (2006). Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nat. Struct. Mol. Biol.*, 13(9):787–792. [DOI:10.1038/nsmb1140] [PubMed:16936728].

John, B., Enright, A. J., Aravin, A., Tuschl, T., Sander, C., and Marks, D. S. (2004). Human MicroRNA targets. *PLoS Biol.*, 2(11):e363. [PubMed Central:PMC521178] [DOI:10.1371/journal.pbio.0020363] [PubMed:15502875].

Jones, L., Ratcliff, F., and Baulcombe, D. C. (2001). RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. *Curr. Biol.*, 11(10):747–757. [PubMed:11378384].

Joshua-Tor, L. and Hannon, G. J. (2011). Ancestral roles of small RNAs: an Ago-centric perspective. *Cold Spring Harb Perspect Biol*, 3(10):a003772. [DOI:10.1101/cshperspect.a003772] [PubMed:20810548].

Kalsotra, A., Wang, K., Li, P. F., and Cooper, T. A. (2010). MicroRNAs coordinate an alternative splicing network during mouse postnatal heart development. *Genes Dev.*, 24(7):653–658. [PubMed Central:PMC2849122] [DOI:10.1101/gad.1894310] [PubMed:20299448].

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., Stadler, P. F., Hertel, J., Hackermuller, J., Hofacker, I. L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., Tammana, H., and Gingeras, T. R. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, 316(5830):1484–1488. [DOI:10.1126/science.1138341] [PubMed:17510325].

Karginov, F. V. and Hannon, G. J. (2010). The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell*, 37(1):7–19. [PubMed Central:PMC2819186] [DOI:10.1016/j.molcel.2009.12.033] [PubMed:20129051].

Kessler, O., Jiang, Y., and Chasin, L. A. (1993). Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. *Mol. Cell. Biol.*, 13(10):6211–6222. [PubMed Central:PMC364680] [PubMed:8413221].

Ketting, R. F. (2011). The many faces of RNAi. *Dev. Cell*, 20(2):148–161. [DOI:10.1016/j.devcel.2011.01.012] [PubMed:21316584].

Khodor, Y. L., Rodriguez, J., Abruzzi, K. C., Tang, C. H., Marr, M. T., and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev.*, 25(23):2502–2512. [PubMed Central:PMC3243060] [DOI:10.1101/gad.178962.111] [PubMed:22156210].

Kim, D. H., Saetrom, P., Sn?ve, O., and Rossi, J. J. (2008). MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.*, 105(42):16230–16235. [PubMed Central:PMC2571020] [DOI:10.1073/pnas.0808830105] [PubMed:18852463].

Kim, D. H., Villeneuve, L. M., Morris, K. V., and Rossi, J. J. (2006). Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat. Struct. Mol. Biol.*, 13(9):793–797. [DOI:10.1038/nsmb1142] [PubMed:16936726].

Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D. L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc. Natl. Acad. Sci. U.S.A.*, 108(33):13564–13569. [PubMed Central:PMC3158196] [DOI:10.1073/pnas.1109475108] [PubMed:21807997].

Kim, T. K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., Harmin, D. A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P. F., Kreiman, G., and Greenberg, M. E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, 465(7295):182–187. [PubMed Central:PMC3020079] [DOI:10.1038/nature09033] [PubMed:20393465].

Kishore, S., Khanna, A., Zhang, Z., Hui, J., Balwierz, P. J., Stefan, M., Beach, C., Nicholls, R. D., Zavolan, M., and Stamm, S. (2010). The snoRNA

MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. *Hum. Mol. Genet.*, 19(7):1153–1164. [PubMed Central:PMC2838533] [DOI:10.1093/hmg/ddp585] [PubMed:20053671].

Kishore, S. and Stamm, S. (2006a). Regulation of alternative splicing by snoRNAs. *Cold Spring Harb. Symp. Quant. Biol.*, 71:329–334. [DOI:10.1101/sqb.2006.71.024] [PubMed:17381313].

Kishore, S. and Stamm, S. (2006b). The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science*, 311(5758):230–232. [DOI:10.1126/science.1118265] [PubMed:16357227].

Klattenhoff, C. and Theurkauf, W. (2008). Biogenesis and germline functions of piRNAs. *Development*, 135(1):3–9. [DOI:10.1242/dev.006486] [PubMed:18032451].

Klinck, R., Bramard, A., Inkel, L., Dufresne-Martin, G., Gervais-Bird, J., Madden, R., Paquet, E. R., Koh, C., Venables, J. P., Prinos, P., Jilaveanu-Pelmus, M., Wellinger, R., Rancourt, C., Chabot, B., and Abou Elela, S. (2008). Multiple alternative splicing markers for ovarian cancer. *Cancer Res.*, 68(3):657–663. [DOI:10.1158/0008-5472.CAN-07-2580] [PubMed:18245464].

Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X. S., and Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.*, 41(3):376–381. [PubMed Central:PMC2648722] [DOI:10.1038/ng.322] [PubMed:19182803].

Komarnitsky, P., Cho, E. J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.*, 14(19):2452–2460. [PubMed Central:PMC316976] [PubMed:11018013].

Kornberg, R. D. and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*, 98(3):285–294. [PubMed:10458604].

Kornblihtt, A. R. (2005). Promoter usage and alternative splicing. *Curr. Opin. Cell Biol.*, 17(3):262–268. [DOI:10.1016/j.ceb.2005.04.014] [PubMed:15901495].

Kornblihtt, A. R. (2006). Chromatin, transcript elongation and alternative splicing. *Nat. Struct. Mol. Biol.*, 13(1):5–7. [DOI:10.1038/nsmb0106-5] [PubMed:16395314].

Kornblihtt, A. R., de la Mata, M., Fededa, J. P., Munoz, M. J., and Nogues, G. (2004). Multiple links between transcription and splicing. *RNA*, 10(10):1489–1498. [PubMed Central:PMC1370635] [DOI:10.1261/rna.7100104] [PubMed:15383674].

Kotovic, K. M., Lockshon, D., Boric, L., and Neugebauer, K. M. (2003). Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. *Mol. Cell. Biol.*, 23(16):5768–5779. [PubMed Central:PMC166328] [PubMed:12897147].

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, 128(4):693–705. [DOI:10.1016/j.cell.2007.02.005] [PubMed:17320507].

Kowalczyk, M. S., Hughes, J. R., Garrick, D., Lynch, M. D., Sharpe, J. A., Sloane-Stanley, J. A., McGowan, S. J., De Gobbi, M., Hosseini, M., Vernimmen, D., Brown, J. M., Gray, N. E., Collavin, L., Gibbons, R. J., Flint, J., Taylor, S., Buckle, V. J., Milne, T. A., Wood, W. G., and Higgs, D. R. (2012). Intragenic enhancers act as alternative promoters. *Mol. Cell*, 45(4):447–458. [DOI:10.1016/j.molcel.2011.12.021] [PubMed:22264824].

Kozomara, A. and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, 39(Database issue):D152–157. [PubMed Central:PMC3013655] [DOI:10.1093/nar/gkq1027] [PubMed:21037258].

Kwon, S. H. and Workman, J. L. (2008). The heterochromatin protein 1 (HP1) family: put away a bias toward HP1. *Mol. Cells*, 26(3):217–227. [PubMed:18664736].

Kwon, S. H. and Workman, J. L. (2011). The changing faces of HP1: From heterochromatin formation and gene silencing to euchromatic gene expression: HP1 acts as a positive regulator of transcription. *Bioessays*, 33(4):280–289. [DOI:10.1002/bies.201000138] [PubMed:21271610].

Lacadie, S. A. and 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. [DOI:10.1016/j.molcel.2005.05.006] [PubMed:15989965].

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science*, 294(5543):853–858. [DOI:10.1126/science.1064921] [PubMed:11679670].

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, 10(3):R25. [PubMed Central:PMC2690996] [DOI:10.1186/gb-2009-10-3-r25] [PubMed:19261174].

Lau, N. C., Lim, L. P., Weinstein, E. G., and Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science*, 294(5543):858–862. [DOI:10.1126/science.1065062] [PubMed:11679671].

Le Thomas, A., Rogers, A. K., Webster, A., Marinov, G. K., Liao, S. E., Perkins, E. M., Hur, J. K., Aravin, A. A., and Toth, K. F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.*, 27(4):390–399. [PubMed Central:PMC3589556] [DOI:10.1101/gad.209841.112] [PubMed:23392610].

Lee, R. C. and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. *Science*, 294(5543):862–864. [DOI:10.1126/science.1065329] [PubMed:11679672].

Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L., and Steitz, J. A. (1980). Are snRNPs involved in splicing? *Nature*, 283(5743):220–224.

Li, B., Carey, M., and Workman, J. L. (2007). The role of chromatin during transcription. *Cell*, 128(4):707–719. [DOI:10.1016/j.cell.2007.01.015] [PubMed:17320508].

Liang, K. and Keles, S. (2012). Normalization of ChIP-seq data with control. *BMC Bioinformatics*, 13:199. [PubMed Central:PMC3475056] [DOI:10.1186/1471-2105-13-199] [PubMed:].

Lin, H. and Yin, H. (2008). A novel epigenetic mechanism in Drosophila somatic cells mediated by Piwi and piRNAs. *Cold Spring Harb. Symp. Quant. Biol.*, 73:273–281. [PubMed Central:PMC2810500] [DOI:10.1101/sqb.2008.73.056] [PubMed:19270080].

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X. D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. *Nat. Struct. Mol. Biol.*, 15(8):819–826. [PubMed Central:PMC2574591] [DOI:10.1038/nsmb.1461] [PubMed:18641664].

Lindroth, A. M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I., Jenuwein, T., Khorasanizadeh, S., and Jacobsen, S. E. (2004). Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *EMBO J.*, 23(21):4286–4296. [PubMed Central:PMC524392] [DOI:10.1038/sj.emboj.7600430] [PubMed:15457214].

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2004). Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat. Struct. Mol. Biol.*, 11(6):576–577. [DOI:10.1038/nsmb777] [PubMed:15156196].

Listerman, I., Sapra, A. K., and Neugebauer, K. M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat. Struct. Mol. Biol.*, 13(9):815–822. [DOI:10.1038/nsmb1135] [PubMed:16921380].

Loomis, R. J., Naoe, Y., Parker, J. B., Savic, V., Bozovsky, M. R., Macfarlan, T., Manley, J. L., and Chakravarti, D. (2009). Chromatin binding of SRp20 and ASF/SF2 and dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation. *Mol. Cell*, 33(4):450–461. [PubMed Central:PMC2667802] [DOI:10.1016/j.molcel.2009.02.003] [PubMed:19250906].

Lorincz, M. C., Dickerson, D. R., Schmitt, M., and Groudine, M. (2004). Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat. Struct. Mol. Biol.*, 11(11):1068–1075. [DOI:10.1038/nsmb840] [PubMed:15467727].

Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell*, 144(1):16–26. [PubMed Central:PMC3038581] [DOI:10.1016/j.cell.2010.11.056] [PubMed:21215366].

Luco, R. F. and Misteli, T. (2011). More than a splicing code: integrating the role of RNA, chromatin and non-coding RNA in alternative splicing regulation. *Curr. Opin. Genet. Dev.*, 21(4):366–372. [DOI:10.1016/j.gde.2011.03.004] [PubMed:21497503].

Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science*, 327(5968):996–1000. [PubMed Central:PMC2913848] [DOI:10.1126/science.1184208] [PubMed:20133523].

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389(6648):251–260. [DOI:10.1038/38444] [PubMed:9305837].

Ma, J. B., Ye, K., and Patel, D. J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, 429(6989):318–322. [DOI:10.1038/nature02519] [PubMed:15152257].

Maison, C. and Almouzni, G. (2004). HP1 and the dynamics of heterochromatin maintenance. *Nat. Rev. Mol. Cell Biol.*, 5(4):296–304. [DOI:10.1038/nrm1355] [PubMed:15071554].

Makeyev, E. V., Zhang, J., Carrasco, M. A., and Maniatis, T. (2007). The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell*, 27(3):435–448. [PubMed Central:PMC3139456] [DOI:10.1016/j.molcel.2007.07.015] [PubMed:17679093].

Manley, J. L. and Tacke, R. (1996). SR proteins and splicing control. *Genes Dev.*, 10(13):1569–1579. [PubMed:8682289].

Marais, G., Nouvellet, P., Keightley, P. D., and Charlesworth, B. (2005). Intron size and exon evolution in Drosophila. *Genetics*, 170(1):481–

485. [PubMed Central:PMC1449718] [DOI:10.1534/genetics.104.037333] [PubMed:15781704].

Matlin, A. J., Clark, F., and Smith, C. W. (2005). Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.*, 6(5):386–398. [DOI:10.1038/nrm1645] [PubMed:15956978].

Matzke, M. A., Primig, M., Trnovsky, J., and Matzke, A. J. (1989). Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J.*, 8(3):643–649. [PubMed Central:PMC400855] [PubMed:16453872].

Mauger, D. M., Lin, C., and Garcia-Blanco, M. A. (2008). hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc. *Mol. Cell. Biol.*, 28(17):5403–5419. [PubMed Central:PMC2519734] [DOI:10.1128/MCB.00739-08] [PubMed:18573884].

Mayr, C. and Bartel, D. P. (2009). Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, 138(4):673–684. [PubMed Central:PMC2819821] [DOI:10.1016/j.cell.2009.06.016] [PubMed:19703394].

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D. L. (1997a). 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev.*, 11(24):3306–3318. [PubMed Central:PMC316822] [PubMed:9407024].

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M., and Bentley, D. L. (1997b). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, 385(6614):357–361. [DOI:10.1038/385357a0] [PubMed:9002523].

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell*, 15(2):185–197. [DOI:10.1016/j.molcel.2004.07.007] [PubMed:15260970].

Meister, G. and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. *Nature*, 431(7006):343–349. [DOI:10.1038/nature02873] [PubMed:15372041].

Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A., and Matzke, A. J. (2000). Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.*, 19(19):5194–5201. [PubMed Central:PMC302106] [DOI:10.1093/emboj/19.19.5194] [PubMed:11013221].

Meyer, L. R., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Kuhn, R. M., Wong, M., Sloan, C. A., Rosenbloom, K. R., Roe, G., Rhead, B., Raney, B. J., Pohl, A., Malladi, V. S., Li, C. H., Lee, B. T., Learned, K., Kirkup, V., Hsu, F., Heitner, S., Harte, R. A., Haeussler, M., Guruvadoo, L., Goldman, M., Giardine, B. M., Fujita, P. A., Dreszer, T. R., Diekhans, M., Cline, M. S., Clawson, H., Barber, G. P., Haussler, D., and Kent, W. J. (2013). The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res.*, 41(Database issue):D64–69. [PubMed Central:PMC3531082] [DOI:10.1093/nar/gks1048] [PubMed:23155063].

Middendorf, M., Kundaje, A., Wiggins, C., Freund, Y., and Leslie, C. (2004). Predicting genetic regulatory response using classification. *Bioinformatics*, 20 Suppl 1:i232–240. [DOI:10.1093/bioinformatics/bth923] [PubMed:15262804].

Millhouse, S. and Manley, J. L. (2005). The C-terminal domain of RNA polymerase II functions as a phosphorylation-dependent splicing activator in a heterologous protein. *Mol. Cell. Biol.*, 25(2):533–544. [PubMed Central:PMC543425] [DOI:10.1128/MCB.25.2.533-544.2005] [PubMed:15632056].

Misteli, T. and Spector, D. L. (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol. Cell*, 3(6):697–705. [PubMed:10394358].

Moazed, D. (2009). Small RNAs in transcriptional gene silencing and genome defence. *Nature*, 457(7228):413–420. [PubMed Central:PMC3246369] [DOI:10.1038/nature07756] [PubMed:19158787].

Morris, K. V., Chan, S. W., Jacobsen, S. E., and Looney, D. J. (2004). Small interfering RNA-induced transcriptional gene silencing in human cells. *Science*, 305(5688):1289–1292. [DOI:10.1126/science.1101372] [PubMed:15297624].

Morris, K. V., Santoso, S., Turner, A. M., Pastori, C., and Hawkins, P. G. (2008). Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genet.*, 4(11):e1000258. [PubMed Central:PMC2576438] [DOI:10.1371/journal.pgen.1000258] [PubMed:19008947].

Moshkovich, N., Nisha, P., Boyle, P. J., Thompson, B. A., Dale, R. K., and Lei, E. P. (2011). RNAi-independent role for Argonaute2 in CTCF/CP190 chromatin insulator function. *Genes Dev.*, 25(16):1686–1701. [PubMed Central:PMC3165934] [DOI:10.1101/gad.16651211] [PubMed:21852534].

Mount, S. M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Res.*, 10(2):459–472. [PubMed Central:PMC326150] [PubMed:7063411].

Munoz, M. J., Perez Santangelo, M. S., Paronetto, M. P., de la Mata, M., Pelisch, F., Boireau, S., Glover-Cutter, K., Ben-Dov, C., Blaustein, M., Lozano, J. J., Bird, G., Bentley, D., Bertrand, E., and Kornblihtt, A. R. (2009). DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell*, 137(4):708–720. [DOI:10.1016/j.cell.2009.03.010] [PubMed:19450518].

Murphy, D., Dancis, B., and Brown, J. R. (2008). The evolution of core proteins involved in microRNA biogenesis. *BMC Evol. Biol.*, 8:92. [PubMed Central:PMC2287173] [DOI:10.1186/1471-2148-8-92] [PubMed:18366743].

Myers, R. M., Stamatoyannopoulos, J., Snyder, M., Dunham, I., Hardison, R. C., Bernstein, B. E., Gingeras, T. R., Kent, W. J., Birney, E., Wold, B., Crawford, G. E., Bernstein, B. E., Epstein, C. B., Shoresh, N., Ernst, J., Mikkelsen, T. S., Kheradpour, P., Zhang, X., Wang, L., Issner, R., Coyne, M. J., Durham, T., Ku, M., Truong, T., Ward, L. D., Altshuler, R. C., Lin, M. F., Kellis, M., Gingeras, T. R., Davis, C. A., Kapranov, P., Dobin, A., Zaleski, C., Schlesinger, F., Batut, P., Chakrabortty, S., Jha, S., Lin, W., Drenkow, J., Wang, H., Bell, K., Gao, H., Bell, I., Dumais, E., Dumais, J.,

Antonarakis, S. E., Ucla, C., Borel, C., Guigo, R., Djebali, S., Lagarde, J., Kingswood, C., Ribeca, P., Sammeth, M., Alioto, T., Merkel, A., Tilgner, H., Carninci, P., Hayashizaki, Y., Lassmann, T., Takahashi, H., Abdelhamid, R. F., Hannon, G., Fejes-Toth, K., Preall, J., Gordon, A., Sotirova, V., Reymond, A., Howald, C., Graison, E., Chrast, J., Ruan, Y., Ruan, X., Shahab, A., Ting Poh, W., Wei, C. L., Crawford, G. E., Furey, T. S., Boyle, A. P., Sheffield, N. C., Song, L., Shibata, Y., Vales, T., Winter, D., Zhang, Z., London, D., Wang, T., Birney, E., Keefe, D., Iyer, V. R., Lee, B. K., McDaniell, R. M., Liu, Z., Battenhouse, A., Bhinge, A. A., Lieb, J. D., Grasfeder, L. L., Showers, K. A., Giresi, P. G., Kim, S. K., Shestak, C., Myers, R. M., Pauli, F., Reddy, T. E., Gertz, J., Partridge, E. C., Jain, P., Sprouse, R. O., Bansal, A., Pusey, B., Muratet, M. A., Varley, K. E., Bowling, K. M., Newberry, K. M., Nesmith, A. S., Dilocker, J. A., Parker, S. L., Waite, L. L., Thibeault, K., Roberts, K., Absher, D. M., Wold, B., Mortazavi, A., Williams, B., Marinov, G., Trout, D., Pepke, S., King, B., McCue, K., Kirilusha, A., DeSalvo, G., Fisher-Aylor, K., Amrhein, H., Vielmetter, J., Sherlock, G., Sidow, A., Batzoglou, S., Rauch, R., Kundaje, A., Libbrecht, M., Margulies, E. H., Parker, S. C., Elnitski, L., Green, E. D., Hubbard, T., Harrow, J., Searle, S., Kokocinski, F., Aken, B., Frankish, A., Hunt, T., Despacio-Reyes, G., Kay, M., Mukherjee, G., Bignell, A., Saunders, G., Boychenko, V., Van Baren, M., Brown, R. H., Khurana, E., Balasubramanian, S., Zhang, Z., Lam, H., Cayting, P., Robilotto, R., Lu, Z., Guigo, R., Derrien, T., Tanzer, A., Knowles, D. G., Mariotti, M., James Kent, W., Haussler, D., Harte, R., Diekhans, M., Kellis, M., Lin, M., Kheradpour, P., Ernst, J., Reymond, A., Howald, C., Graison, E. A., Chrast, J., Tress, M., Rodriguez, J. M., Snyder, M., Landt, S. G., Raha, D., Shi, M., Euskirchen, G., Grubert, F., Kasowski, M., Lian, J., Cayting, P., Lacroute, P., Xu, Y., Monahan, H., Patacsil, D., Slifer, T., Yang, X., Charos, A., Reed, B., Wu, L., Auerbach, R. K., Habegger, L., Hariharan, M., Rozowsky, J., Abyzov, A., Weissman, S. M., Gerstein, M., Struhl, K., Lamarre-Vincent, N., Lindahl-Allen, M., Miotto, B., Moqtaderi, Z., Fleming, J. D., Newburger, P., Farnham, P. J., Frietze, S., O'Geen, H., Xu, X., Blahnik, K. R., Cao, A. R., Iyengar, S., Stamatoyannopoulos, J. A., Kaul, R., Thurman, R. E., Wang, H., Navas, P. A., Sandstrom, R., Sabo, P. J., Weaver, M., Canfield, T., Lee, K., Neph, S., Roach, V., Reynolds, A., Johnson, A., Rynes, E., Giste, E., Vong, S., Neri, J., Frum, T., Johnson, E. M., Nguyen,

E. D., Ebersol, A. K., Sanchez, M. E., Sheffer, H. H., Lotakis, D., Haugen, E., Humbert, R., Kutyavin, T., Shafer, T., Dekker, J., Lajoie, B. R., Sanyal, A., James Kent, W., Rosenbloom, K. R., Dreszer, T. R., Raney, B. J., Barber, G. P., Meyer, L. R., Sloan, C. A., Malladi, V. S., Cline, M. S., Learned, K., Swing, V. K., Zweig, A. S., Rhead, B., Fujita, P. A., Roskin, K., Karolchik, D., Kuhn, R. M., Haussler, D., Birney, E., Dunham, I., Wilder, S. P., Keefe, D., Sobral, D., Herrero, J., Beal, K., Lukk, M., Brazma, A., Vaquerizas, J. M., Luscombe, N. M., Bickel, P. J., Boley, N., Brown, J. B., Li, Q., Huang, H., Gerstein, M., Habegger, L., Sboner, A., Rozowsky, J., Auerbach, R. K., Yip, K. Y., Cheng, C., Yan, K. K., Bhardwaj, N., Wang, J., Lochovsky, L., Jee, J., Gibson, T., Leng, J., Du, J., Hardison, R. C., Harris, R. S., Song, G., Miller, W., Haussler, D., Roskin, K., Suh, B., Wang, T., Paten, B., Noble, W. S., Hoffman, M. M., Buske, O. J., Weng, Z., Dong, X., Wang, J., Xi, H., Tenenbaum, S. A., Doyle, F., Penalva, L. O., Chittur, S., Tullius, T. D., Parker, S. C., White, K. P., Karmakar, S., Victorsen, A., Jameel, N., Bild, N., Grossman, R. L., Snyder, M., Landt, S. G., Yang, X., Patacsil, D., Slifer, T., Dekker, J., Lajoie, B. R., Sanyal, A., Weng, Z., Whitfield, T. W., Wang, J., Collins, P. J., Trinklein, N. D., Partridge, E. C., Myers, R. M., Giddings, M. C., Chen, X., Khatun, J., Maier, C., Yu, Y., Gunawardena, H., Risk, B., Feingold, E. A., Lowdon, R. F., Dillon, L. A., Good, P. J., Harrow, J., and Searle, S. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.*, 9(4):e1001046. [PubMed Central:PMC3079585] [DOI:10.1371/journal.pbio.1001046] [PubMed:21526222].

Nahkuri, S., Taft, R. J., and Mattick, J. S. (2009). Nucleosomes are preferentially positioned at exons in somatic and sperm cells. *Cell Cycle*, 8(20):3420–3424. [PubMed:19823040].

Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., and Grewal, S. I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science*, 292(5514):110–113. [DOI:10.1126/science.1060118] [PubMed:11283354].

Nishi, K., Nishi, A., Nagasawa, T., and Ui-Tei, K. (2013). Human TNRC6A is an Argonaute-navigator protein for microRNA-mediated gene silenc-

ing in the nucleus. *RNA*, 19(1):17–35. [PubMed Central:PMC3527724] [DOI:10.1261/rna.034769.112] [PubMed:23150874].

Nogues, G., Kadener, S., Cramer, P., Bentley, D., and Kornblihtt, A. R. (2002). Transcriptional activators differ in their abilities to control alternative splicing. *J. Biol. Chem.*, 277(45):43110–43114. [DOI:10.1074/jbc.M208418200] [PubMed:12221105].

Ohrt, T., Mutze, J., Staroske, W., Weinmann, L., Hock, J., Crell, K., Meister, G., and Schwille, P. (2008). Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. *Nucleic Acids Res.*, 36(20):6439–6449. [PubMed Central:PMC2582625] [DOI:10.1093/nar/gkn693] [PubMed:18842624].

Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., and Sharp, P. A. (1986). Splicing of messenger RNA precursors. *Annu. Rev. Biochem.*, 55:1119–1150.

Padgett, R. A., Konarska, M. M., Grabowski, P. J., Hardy, S. F., and Sharp, P. A. (1984). Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science*, 225(4665):898–903. [PubMed:6206566].

Pagani, F., Stuani, C., Zuccato, E., Kornblihtt, A. R., and Baralle, F. E. (2003). Promoter architecture modulates CFTR exon 9 skipping. *J. Biol. Chem.*, 278(3):1511–1517. [DOI:10.1074/jbc.M209676200] [PubMed:12421814].

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, 40(12):1413–1415. [DOI:10.1038/ng.259] [PubMed:18978789].

Pandya-Jones, A. and Black, D. L. (2009). Co-transcriptional splicing of constitutive and alternative exons. *RNA*, 15(10):1896–1908. [PubMed Central:PMC2743041] [DOI:10.1261/rna.1714509] [PubMed:19656867].

Parker, J. S., Roe, S. M., and Barford, D. (2005). Structural insights into mRNA recognition from a PIWI domain-siRNA guide com-

plex. *Nature*, 434(7033):663–666. [PubMed Central:PMC2938470] [DOI:10.1038/nature03462] [PubMed:15800628].

Perales, R. and Bentley, D. (2009). "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol. Cell*, 36(2):178–191. [PubMed Central:PMC2770090] [DOI:10.1016/j.molcel.2009.09.018] [PubMed:19854129].

Persson, H., Kvist, A., Vallon-Christersson, J., Medstrand, P., Borg, A., and Rovira, C. (2009). The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. *Nat. Cell Biol.*, 11(10):1268–1271. [DOI:10.1038/ncb1972] [PubMed:19749744].

Phatnani, H. P. and Greenleaf, A. L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.*, 20(21):2922–2936. [DOI:10.1101/gad.1477006] [PubMed:17079683].

Phillips, J. E. and Corces, V. G. (2009). CTCF: master weaver of the genome. *Cell*, 137(7):1194–1211. [PubMed Central:PMC3040116] [DOI:10.1016/j.cell.2009.06.001] [PubMed:19563753].

Piacentini, L., Fanti, L., Negri, R., Del Vescovo, V., Fatica, A., Altieri, F., and Pimpinelli, S. (2009). Heterochromatin protein 1 (HP1a) positively regulates euchromatic gene expression through RNA transcript association and interaction with hnRNPs in Drosophila. *PLoS Genet.*, 5(10):e1000670. [PubMed Central:PMC2743825] [DOI:10.1371/journal.pgen.1000670] [PubMed:19798443].

Place, R. F., Li, L. C., Pookot, D., Noonan, E. J., and Dahiya, R. (2008). MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc. Natl. Acad. Sci. U.S.A.*, 105(5):1608–1613. [PubMed Central:PMC2234192] [DOI:10.1073/pnas.0707594105] [PubMed:18227514].

Pradeepa, M. M., Sutherland, H. G., Ule, J., Grimes, G. R., and Bickmore, W. A. (2012). Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genet.*, 8(5):e1002717. [PubMed Central:PMC3355077] [DOI:10.1371/journal.pgen.1002717] [PubMed:22615581].

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M. S., Mapendano, C. K., Schierup, M. H., and Jensen, T. H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. *Science*, 322(5909):1851–1854. [DOI:10.1126/science.1164096] [PubMed:19056938].

Pruitt, K. D., Tatusova, T., Brown, G. R., and Maglott, D. R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.*, 40(Database issue):D130–135. [PubMed Central:PMC3245008] [DOI:10.1093/nar/gkr1079] [PubMed:22121212].

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842. [PubMed Central:PMC2832824] [DOI:10.1093/bioinformatics/btq033] [PubMed:20110278].

Richardson, J. E. (2006). fjoin: simple and efficient computation of feature overlaps. *J. Comput. Biol.*, 13(8):1457–1464. [DOI:10.1089/cmb.2006.13.1457] [PubMed:17061921].

Robb, G. B., Brown, K. M., Khurana, J., and Rana, T. M. (2005). Specific and potent RNAi in the nucleus of human cells. *Nat. Struct. Mol. Biol.*, 12(2):133–137. [DOI:10.1038/nsmb886] [PubMed:15643423].

Robberson, B. L., Cote, G. J., and Berget, S. M. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.*, 10(1):84–94. [PubMed Central:PMC360715] [PubMed:2136768].

Roberts, G. C., Gooding, C., Mak, H. Y., Proudfoot, N. J., and Smith, C. W. (1998). Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Res.*, 26(24):5568–5572. [PubMed Central:PMC148035] [PubMed:9837984].

Robson-Dixon, N. D. and Garcia-Blanco, M. A. (2004). MAZ elements alter transcription elongation and silencing of the fibroblast growth factor receptor 2 exon IIIb. *J. Biol. Chem.*, 279(28):29075–29084. [DOI:10.1074/jbc.M312747200] [PubMed:15126509].

Ross-Innes, C. S., Brown, G. D., and Carroll, J. S. (2011). A co-ordinated interaction between CTCF and ER in breast cancer cells. *BMC Genomics*, 12:593. [PubMed Central:PMC3248577] [DOI:10.1186/1471-2164-12-593] [PubMed:22142239].

Rozowsky, J., Euskirchen, G., Auerbach, R. K., Zhang, Z. D., Gibson, T., Bjornson, R., Carriero, N., Snyder, M., and Gerstein, M. B. (2009). PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nat. Biotechnol.*, 27(1):66–75. [PubMed Central:PMC2924752] [DOI:10.1038/nbt.1518] [PubMed:19122651].

Sabin, L. R., Delas, M. J., and Hannon, G. J. (2013). Dogma derailed: the many influences of RNA on the genome. *Mol. Cell*, 49(5):783–794. [DOI:10.1016/j.molcel.2013.02.010] [PubMed:23473599].

Saint-Andre, V., Batsche, E., Rachez, C., and Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1 favor inclusion of alternative exons. *Nat. Struct. Mol. Biol.*, 18(3):337–344. [DOI:10.1038/nsmb.1995] [PubMed:21358630].

Sasaki, T., Shiohama, A., Minoshima, S., and Shimizu, N. (2003). Identification of eight members of the Argonaute family in the human genome small star, filled. *Genomics*, 82(3):323–330. [PubMed:12906857].

Schirle, N. T. and MacRae, I. J. (2012). The crystal structure of human Argonaute2. *Science*, 336(6084):1037–1040. [DOI:10.1126/science.1221551] [PubMed:22539551].

Schmidt, D., Wilson, M. D., Ballester, B., Schwalie, P. C., Brown, G. D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C. P., Mackay, S., Talianidis, I., Flicek, P., and Odom, D. T. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science*, 328(5981):1036–1040. [PubMed Central:PMC3008766] [DOI:10.1126/science.1186176] [PubMed:20378774].

Schor, I. E., Rascovan, N., Pelisch, F., Allo, M., and Kornblihtt, A. R. (2009). Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc. Natl.*

Acad. Sci. U.S.A., 106(11):4325–4330. [PubMed Central:PMC2657401] [DOI:10.1073/pnas.0810666106] [PubMed:19251664].

Schroeder, S. C., Schwer, B., Shuman, S., and Bentley, D. (2000). Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev.*, 14(19):2435–2440. [PubMed Central:PMC316982] [PubMed:11018011].

Schwartz, J. C., Younger, S. T., Nguyen, N. B., Hardy, D. B., Monia, B. P., Corey, D. R., and Janowski, B. A. (2008). Antisense transcripts are targets for activating small RNAs. *Nat. Struct. Mol. Biol.*, 15(8):842–848. [PubMed Central:PMC2574822] [DOI:10.1038/nsmb.1444] [PubMed:18604220].

Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.*, 16(9):990–995. [DOI:10.1038/nsmb.1659] [PubMed:19684600].

Seila, A. C., Core, L. J., Lis, J. T., and Sharp, P. A. (2009). Divergent transcription: a new feature of active promoters. *Cell Cycle*, 8(16):2557–2564. [PubMed:19597342].

Selth, L. A., Sigurdsson, S., and Svejstrup, J. Q. (2010). Transcript Elongation by RNA Polymerase II. *Annu. Rev. Biochem.*, 79:271–293. [DOI:10.1146/annurev.biochem.78.062807.091425] [PubMed:20367031].

Seraphin, B., Kretzner, L., and 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. [PubMed Central:PMC457124] [PubMed:3056718].

Shahi, P., Loukianiouk, S., Bohne-Lang, A., Kenzelmann, M., Kuffer, S., Maertens, S., Eils, R., Grone, H. J., Gretz, N., and Brors, B. (2006). Argonaute–a database for gene regulation by mammalian microRNAs. *Nucleic Acids Res.*, 34(Database issue):D115–118. [PubMed Central:PMC1347455] [DOI:10.1093/nar/gkj093] [PubMed:16381827].

Shukla, G. C., Singh, J., and Barik, S. (2011a). MicroRNAs: Processing, Maturation, Target Recognition and Regulatory Functions. *Mol Cell Pharmacol*, 3(3):83–92. [PubMed Central:PMC3315687] [PubMed:22468167].

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011b). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature*, 479(7371):74–79. [DOI:10.1038/nature10442] [PubMed:21964334].

Siliciano, P. G. and Guthrie, C. (1988). 5′ splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. *Genes Dev.*, 2(10):1258–1267. [PubMed:3060402].

Sims, R. J., Belotserkovskaya, R., and Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. *Genes Dev.*, 18(20):2437–2468. [DOI:10.1101/gad.1235904] [PubMed:15489290].

Sims, R. J., Millhouse, S., Chen, C. F., Lewis, B. A., Erdjument-Bromage, H., Tempst, P., Manley, J. L., and Reinberg, D. (2007). Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol. Cell*, 28(4):665–676. [PubMed Central:PMC2276655] [DOI:10.1016/j.molcel.2007.11.010] [PubMed:18042460].

Sirand-Pugnet, P., Durosay, P., Clouet d'Orval, B. C., Brody, E., and Marie, J. (1995). beta-Tropomyosin pre-mRNA folding around a muscle-specific exon interferes with several steps of spliceosome assembly. *J. Mol. Biol.*, 251(5):591–602. [PubMed:7666413].

Sisodia, S. S., Sollner-Webb, B., and Cleveland, D. W. (1987). Specificity of RNA maturation pathways: RNAs transcribed by RNA polymerase III are not substrates for splicing or polyadenylation. *Mol. Cell. Biol.*, 7(10):3602–3612. [PubMed Central:PMC368014] [PubMed:3683396].

Smith, C. W. and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.*, 25(8):381–388. [PubMed:10916158].

Smolle, M., Venkatesh, S., Gogol, M. M., Li, H., Zhang, Y., Florens, L., Washburn, M. P., and Workman, J. L. (2012). Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat. Struct. Mol. Biol.* [DOI:10.1038/nsmb.2312] [PubMed:22922743].

Song, J. J., Smith, S. K., Hannon, G. J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*, 305(5689):1434–1437. [DOI:10.1126/science.1102514] [PubMed:15284453].

Spies, N., Nielsen, C. B., Padgett, R. A., and Burge, C. B. (2009). Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell*, 36(2):245–254. [PubMed Central:PMC2786773] [DOI:10.1016/j.molcel.2009.10.008] [PubMed:19854133].

Strahl, B. D. and Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, 403(6765):41–45. [DOI:10.1038/47412] [PubMed:10638745].

Subtil-Rodriguez, A. and Reyes, J. C. (2010). BRG1 helps RNA polymerase II to overcome a nucleosomal barrier during elongation, in vivo. *EMBO Rep.*, 11(10):751–757. [PubMed Central:PMC2948185] [DOI:10.1038/embor.2010.131] [PubMed:20829883].

Sugiyama, T., Cam, H., Verdel, A., Moazed, D., and Grewal, S. I. (2005). RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc. Natl. Acad. Sci. U.S.A.*, 102(1):152–157. [PubMed Central:PMC544066] [DOI:10.1073/pnas.0407641102] [PubMed:15615848].

Sun, Z., Asmann, Y. W., Kalari, K. R., Bot, B., Eckel-Passow, J. E., Baker, T. R., Carr, J. M., Khrebtukova, I., Luo, S., Zhang, L., Schroth, G. P., Perez, E. A., and Thompson, E. A. (2011). Integrated analysis of gene expression, CpG island methylation, and gene copy number in breast cancer cells by deep sequencing. *PLoS ONE*, 6(2):e17490. [PubMed Central:PMC3045451] [DOI:10.1371/journal.pone.0017490] [PubMed:21364760].

Suzuki, K., Shijuuku, T., Fukamachi, T., Zaunders, J., Guillemin, G., Cooper, D., and Kelleher, A. (2005). Prolonged transcriptional silencing and CpG methylation induced by siRNAs targeted to the HIV-1 promoter region. *J RNAi Gene Silencing*, 1(2):66–78. [PubMed Central:PMC2737205] [PubMed:19771207].

Taft, R. J., Glazov, E. A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G. J., Lassmann, T., Forrest, A. R., Grimmond, S. M., Schroder, K., Irvine, K.,

Arakawa, T., Nakamura, M., Kubosaki, A., Hayashida, K., Kawazu, C., Murata, M., Nishiyori, H., Fukuda, S., Kawai, J., Daub, C. O., Hume, D. A., Suzuki, H., Orlando, V., Carninci, P., Hayashizaki, Y., and Mattick, J. S. (2009). Tiny RNAs associated with transcription start sites in animals. *Nat. Genet.*, 41(5):572–578. [DOI:10.1038/ng.312] [PubMed:19377478].

Taft, R. J., Pang, K. C., Mercer, T. R., Dinger, M., and Mattick, J. S. (2010). Non-coding RNAs: regulators of disease. *J. Pathol.*, 220(2):126–139. [DOI:10.1002/path.2638] [PubMed:19882673].

Taliaferro, J. M., Aspden, J. L., Bradley, T., Marwha, D., Blanchette, M., and Rio, D. C. (2013). Two new and distinct roles for Drosophila Argonaute-2 in the nucleus: alternative pre-mRNA splicing and transcriptional repression. *Genes Dev.*, 27(4):378–389. [PubMed Central:PMC3589555] [DOI:10.1101/gad.210708.112] [PubMed:23392611].

Tennyson, C. N., Klamut, H. J., and Worton, R. G. (1995). The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat. Genet.*, 9(2):184–190. [DOI:10.1038/ng0295-184] [PubMed:7719347].

Tilgner, H., Knowles, D. G., Johnson, R., Davis, C. A., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T. R., and Guigo, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.*, 22(9):1616–1625. [DOI:10.1101/gr.134445.111] [PubMed:22955974].

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcarcel, J., and Guigo, R. (2009). Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.*, 16(9):996–1001. [DOI:10.1038/nsmb.1658] [PubMed:19684599].

Ting, A. H., Schuebel, K. E., Herman, J. G., and Baylin, S. B. (2005). Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. *Nat. Genet.*, 37(8):906–910. [PubMed Central:PMC2659476] [DOI:10.1038/ng1611] [PubMed:16025112].

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., Freier, S. M., Bennett, C. F., Sharma, A., Bubulya, P. A., Blencowe, B. J., Prasanth, S. G., and Prasanth, K. V. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell*, 39(6):925–938. [DOI:10.1016/j.molcel.2010.08.011] [PubMed:20797886].

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B. J., and Darnell, R. B. (2006). An RNA map predicting Nova-dependent splicing regulation. *Nature*, 444(7119):580–586. [DOI:10.1038/nature05304] [PubMed:17065982].

van Wolfswinkel, J. C. and Ketting, R. F. (2010). The role of small non-coding RNAs in genome stability and chromatin organization. *J. Cell. Sci.*, 123(Pt 11):1825–1839. [DOI:10.1242/jcs.061713] [PubMed:20484663].

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S. I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. *Science*, 303(5658):672–676.

Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I., and Martienssen, R. A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, 297(5588):1833–1837. [DOI:10.1126/science.1074973] [PubMed:12193640].

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., and Burge, C. B. (2008a). Alternative isoform regulation in human tissue transcriptomes. *Nature*, 456(7221):470–476. [PubMed Central:PMC2593745] [DOI:10.1038/nature07509] [PubMed:18978772].

Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T., and Patel, D. J. (2008b). Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature*, 456(7224):921–926. [PubMed Central:PMC2765400] [DOI:10.1038/nature07666] [PubMed:19092929].

Wang, Y., Sheng, G., Juranek, S., Tuschl, T., and Patel, D. J. (2008c). Structure of the guide-strand-containing argonaute silencing complex. *Nature*, 456(7219):209–213. [DOI:10.1038/nature07315] [PubMed:18754009].

Wang, Z. and Burge, C. B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*, 14(5):802–813. [PubMed Central:PMC2327353] [DOI:10.1261/rna.876308] [PubMed:18369186].

Weinberg, M. S., Villeneuve, L. M., Ehsani, A., Amarzguioui, M., Aagaard, L., Chen, Z. X., Riggs, A. D., Rossi, J. J., and Morris, K. V. (2006). The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. *RNA*, 12(2):256–262. [PubMed Central:PMC1370905] [DOI:10.1261/rna.2235106] [PubMed:16373483].

Welboren, W. J., van Driel, M. A., Janssen-Megens, E. M., van Heeringen, S. J., Sweep, F. C., Span, P. N., and Stunnenberg, H. G. (2009). ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. *EMBO J.*, 28(10):1418–1428. [PubMed Central:PMC2688537] [DOI:10.1038/emboj.2009.88] [PubMed:19339991].

Will, C. L. and Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb Perspect Biol*, 3(7). [DOI:10.1101/cshperspect.a003707] [PubMed:21441581].

Yan, K. S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M. M. (2003). Structure and conserved RNA binding of the PAZ domain. *Nature*, 426(6965):468–474. [DOI:10.1038/nature02129] [PubMed:14615802].

Yeo, G. W., Coufal, N. G., Liang, T. Y., Peng, G. E., Fu, X. D., and Gage, F. H. (2009). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.*, 16(2):130–137. [PubMed Central:PMC2735254] [DOI:10.1038/nsmb.1545] [PubMed:19136955].

Yin, H. and Lin, H. (2007). An epigenetic activation role of Piwi and a Piwi-associated piRNA in Drosophila melanogaster. *Nature*, 450(7167):304–308. [DOI:10.1038/nature06263] [PubMed:17952056].

Yin, Q. F., Yang, L., Zhang, Y., Xiang, J. F., Wu, Y. W., Carmichael, G. G., and Chen, L. L. (2012). Long Noncoding RNAs with snoRNA Ends. *Mol. Cell*, 48(2):219–230.

Younger, S. T. and Corey, D. R. (2011). Transcriptional gene silencing in mammalian cells by miRNA mimics that target gene promoters. *Nucleic Acids Res.*, 39(13):5682–5691. [PubMed Central:PMC3141263] [DOI:10.1093/nar/gkr155] [PubMed:21427083].

Yuan, Y. R., Pei, Y., Ma, J. B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H. Y., Dauter, Z., Tuschl, T., and Patel, D. J. (2005). Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol. Cell*, 19(3):405–419. [DOI:10.1016/j.molcel.2005.07.011] [PubMed:16061186].

Zaratiegui, M., Irvine, D. V., and Martienssen, R. A. (2007). Noncoding RNAs and gene silencing. *Cell*, 128(4):763–776. [DOI:10.1016/j.cell.2007.02.016] [PubMed:17320512].

Zhang, G., Taneja, K. L., Singer, R. H., and Green, M. R. (1994). Localization of pre-mRNA splicing in mammalian nuclei. *Nature*, 372(6508):809–812. [PubMed:7997273].

Zhao, T., Heyduk, T., Allis, C. D., and Eissenberg, J. C. (2000). Heterochromatin protein 1 binds to nucleosomes and DNA in vitro. *J. Biol. Chem.*, 275(36):28332–28338. [DOI:10.1074/jbc.M003493200] [PubMed:10882726].

Zhou, H. L., Hinman, M. N., Barron, V. A., Geng, C., Zhou, G., Luo, G., Siegel, R. E., and Lou, H. (2011). Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner. *Proc. Natl. Acad. Sci. U.S.A.*, 108(36):E627–635. [PubMed Central:PMC3169152] [DOI:10.1073/pnas.1103344108] [PubMed:21808035].

Zhou, Y., Lu, Y., and Tian, W. (2012). Epigenetic features are significantly associated with alternative splicing. *BMC Genomics*, 13:123. [PubMed Central:PMC3362759] [DOI:10.1186/1471-2164-13-123] [PubMed:22455468].

Zilberman, D., Cao, X., and Jacobsen, S. E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science*, 299(5607):716–719. [DOI:10.1126/science.1079695] [PubMed:12522258].

Zong, X., Tripathi, V., and Prasanth, K. V. (2011). RNA splicing control: yet another gene regulatory role for long nuclear noncoding RNAs. *RNA Biol*, 8(6):968–977. [PubMed Central:PMC3256421] [DOI:10.4161/rna.8.6.17606] [PubMed:21941126].

CHAPTER 6

Appendices

Contents

6.1 Appendix A: Databases and resources for human small non-coding RNAs

In this section I attached the review that I published about human sRNA databases in Human Genomics journal (Agirre and Eyras, 2011).

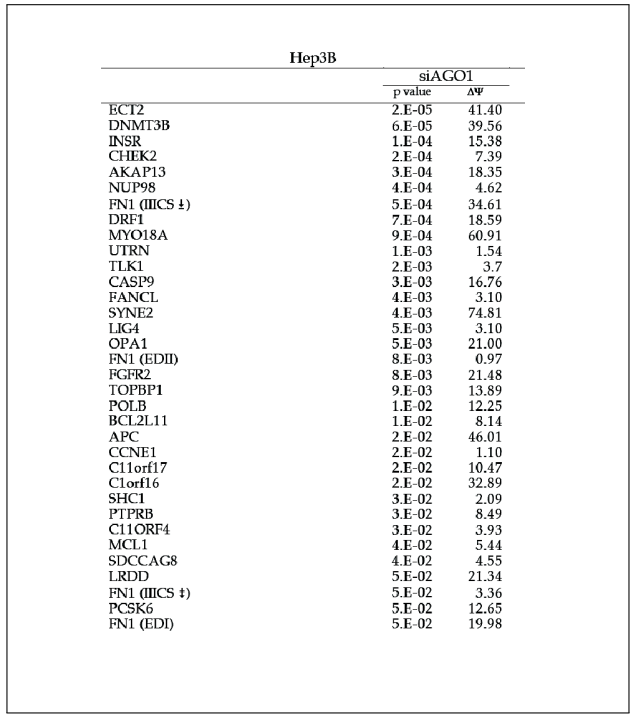
Abstract

Recent advances in high-throughput sequencing have facilitated the genome-wide studies of small non-coding RNAs (sRNAs). Numerous studies have highlighted the role of various classes of sRNAs at different levels of gene regulation and disease. The fast growth of sequence data and the diversity of sRNA species have prompted the need to organise them in annotation databases. There are currently several databases that collect sRNA data. Various tools are provided for access, with special emphasis on the well-characterised family of micro-RNAs. The striking heterogeneity of the new classes of sRNAs and the lack of sufficient functional annotation, however, make integration of these datasets a difficult task. This review describes the currently available databases for human sRNAs that are accessible via the internet, and some of the large datasets for human sRNAs from highthroughput sequencing experiments that are so far only available as supplementary data in publications. Some of the main issues related to the integration and annotation of sRNA datasets are also discussed.

- Article abstract :`http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3500172/?report=abstract`
- Full text :`http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=21504869`
- PDF :`http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3500172/pdf/1479-7364-5-3-192.pdf`

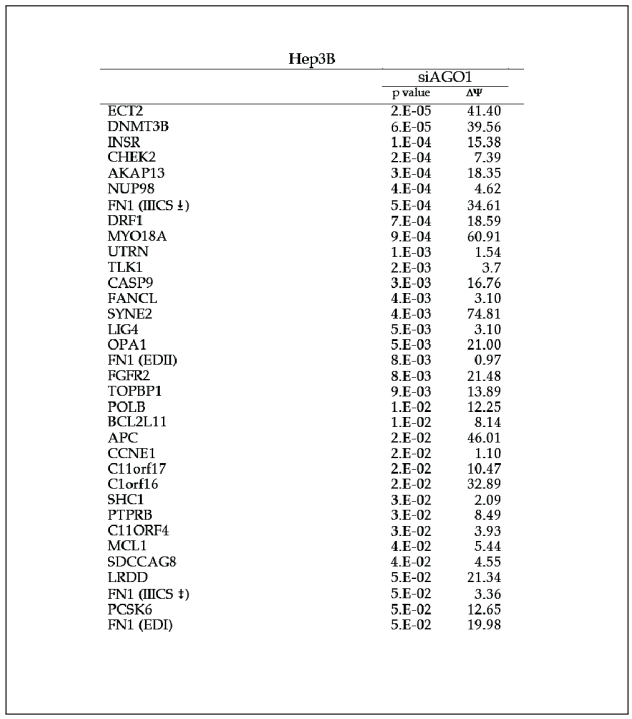
Agirre E, Eyras E. Databases and resources for human small non-coding RNAs. Hum Genomics. 2011 Mar; 5(3):192-199.

6.2 Appendix B: Supplementary figures and tables



Hep3B	siAGO1	
	p value	ΔΨ
ECT2	2.E-05	41.40
DNMT3B	6.E-05	39.56
INSR	1.E-04	15.38
CHEK2	2.E-04	7.39
AKAP13	3.E-04	18.35
NUP98	4.E-04	4.62
FN1 (IIICS ‡)	5.E-04	34.61
DRF1	7.E-04	18.59
MYO18A	9.E-04	60.91
UTRN	1.E-03	1.54
TLK1	2.E-03	3.7
CASP9	3.E-03	16.76
FANCL	4.E-03	3.10
SYNE2	4.E-03	74.81
LIG4	5.E-03	3.10
OPA1	5.E-03	21.00
FN1 (EDII)	8.E-03	0.97
FGFR2	8.E-03	21.48
TOPBP1	9.E-03	13.89
POLB	1.E-02	12.25
BCL2L11	1.E-02	8.14
APC	2.E-02	46.01
CCNE1	2.E-02	1.10
C11orf17	2.E-02	10.47
C1orf16	2.E-02	32.89
SHC1	3.E-02	2.09
PTPRB	3.E-02	8.49
C11ORF4	3.E-02	3.93
MCL1	4.E-02	5.44
SDCCAG8	4.E-02	4.55
LRDD	5.E-02	21.34
FN1 (IIICS ‡)	5.E-02	3.36
PCSK6	5.E-02	12.65
FN1 (EDI)	5.E-02	19.98

Figure 6.1: For each splicing event, the ratio of long isoform versus total isoform concentrations, expressed as a percentage and termed the percent splicing index, ψ, was determined fot the Hep3B cell line. The absolute ψ change between control and AGO1 depleted samples, termed $\Delta\psi$, is reported for each gene



Hep3B

	siAGO1	
	p value	ΔΨ
ECT2	2.E-05	41.40
DNMT3B	6.E-05	39.56
INSR	1.E-04	15.38
CHEK2	2.E-04	7.39
AKAP13	3.E-04	18.35
NUP98	4.E-04	4.62
FN1 (IIICS ‡)	5.E-04	34.61
DRF1	7.E-04	18.59
MYO18A	9.E-04	60.91
UTRN	1.E-03	1.54
TLK1	2.E-03	3.7
CASP9	3.E-03	16.76
FANCL	4.E-03	3.10
SYNE2	4.E-03	74.81
LIG4	5.E-03	3.10
OPA1	5.E-03	21.00
FN1 (EDII)	8.E-03	0.97
FGFR2	8.E-03	21.48
TOPBP1	9.E-03	13.89
POLB	1.E-02	12.25
BCL2L11	1.E-02	8.14
APC	2.E-02	46.01
CCNE1	2.E-02	1.10
C11orf17	2.E-02	10.47
C1orf16	2.E-02	32.89
SHC1	3.E-02	2.09
PTPRB	3.E-02	8.49
C11ORF4	3.E-02	3.93
MCL1	4.E-02	5.44
SDCCAG8	4.E-02	4.55
LRDD	5.E-02	21.34
FN1 (IIICS ‡)	5.E-02	3.36
PCSK6	5.E-02	12.65
FN1 (EDI)	5.E-02	19.98

Figure 6.2: For each splicing event, the ratio of long isoform versus total isoform concentrations, expressed as a percentage and termed the percent splicing index, ψ, was determined for the Hela cell line. The absolute ψ change between control and AGO1 depleted samples, termed $\Delta\psi$, is reported for each gene.

Table 6.1: Predicted miRNA targets in ASEs affected by siAGO1 knockdown in HeLa.

Mature miRNA annotation source MiRbase database (Griffiths-Jones et al., 2008; Kozomara and Griffiths-Jones, 2011; Griffiths-Jones et al., 2006; Griffiths-Jones, 2004)

miRNA	chr	start	end	strand	score	RefseqID
hsa-miR-610	chr15	42460467	42482376	+	166.00	CASC4
hsa-miR-1294	chr15	42460467	42482376	+	173.00	CASC4
hsa-miR-187	chr11	114554706	114585503	-	166.00	CADM1
hsa-miR-127-3p	chr15	84002826	84008797	+	165.00	AKAP13
hsa-miR-1913	chr17	41420299	41423080	+	167.00	MAPT
hsa-miR-146b-5p	chr15	42460467	42482376	+	171.00	CASC4
hsa-miR-519b-5p	chr19	35000289	35003448	+	166.00	CCNE1
hsa-miR-219-2-3p	chr10	74869666	74874488	-	168.00	PPP3CB
hsa-miR-196b	chr11	114554706	114585503	-	167.00	CADM1
hsa-miR-1278	chr10	74869666	74874488	-	166.00	PPP3CB
hsa-miR-891a	chr11	114554706	114585503	-	166.00	CADM1
hsa-miR-101	chr11	114554706	114585503	-	172.00	CADM1
hsa-miR-1913	chr11	8889347	8889566	+	167.00	C11orf17
hsa-miR-146b-5p	chr1	241608789	241645626	+	171.00	SDCCAG8
hsa-miR-516b	chr11	4061305	4064282	+	166.00	STIM1
hsa-miR-634	chr19	35000289	35003448	+	167.00	CCNE1
hsa-miR-28-5p	chr11	114554706	114585503	-	166.00	CADM1
hsa-miR-1537	chr15	84000023	84006622	+	175.00	AKAP13
hsa-miR-499-5p	chr15	84000023	84006622	+	166.00	AKAP13
hsa-miR-518c	chr17	22993502	22996466	+	178.00	LGALS9
hsa-miR-541	chr15	84000023	84006622	+	166.00	AKAP13
hsa-miR-518c	chr17	22994774	22996466	+	178.00	LGALS9
hsa-miR-425	chr1	241608789	241645626	+	170.00	SDCCAG8
hsa-miR-520e	chr19	35000289	35003448	+	169.00	CCNE1
hsa-miR-663	chr5	176452392	176452938	+	168.00	FGFR4
hsa-miR-758	chr22	27471997	27477228	+	171.00	HSC20
hsa-miR-487b	chr1	241608789	241645626	+	166.00	SDCCAG8
hsa-miR-125a-3p	chr22	27471997	27477228	+	165.00	HSC20
hsa-miR-519b-5p	chr11	114554706	114585503	-	170.00	CADM1
hsa-miR-1276	chr1	241608789	241645626	+	167.00	SDCCAG8

hsa-miR-875-5p	chr6	167347046	167355886	+	169.00	FGFR1OP
hsa-miR-196b	chr15	42482545	42492825	+	188.00	CASC4
hsa-miR-518d-3p	chr17	22993502	22996466	+	165.00	LGALS9
hsa-miR-518f	chr17	22993502	22996466	+	168.00	LGALS9
hsa-miR-412	chr6	167347046	167355886	+	167.00	FGFR1OP
hsa-miR-1261	chr17	30334685	30337119	+	169.00	LIG3
hsa-miR-1197	chr17	40171319	40174229	+	168.00	DBF4B
hsa-miR-519c-5p	chr19	35000289	35003448	+	166.00	CCNE1
hsa-miR-631	chr6	45588123	45620931	+	174.00	RUNX2
hsa-miR-933	chr11	114554706	114585503	-	179.00	CADM1
hsa-miR-1197	chr1	241608789	241645626	+	168.00	SDCCAG8
hsa-miR-223	chr15	42482545	42492825	+	167.00	CASC4
hsa-miR-499-5p	chr15	84000023	84002771	+	166.00	AKAP13
hsa-miR-412	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-miR-200a	chr17	30334685	30337119	+	166.00	LIG3
hsa-miR-518c	chr17	22994774	22996466	+	178.00	LGALS9
hsa-miR-516b	chr15	84006689	84008797	+	168.00	AKAP13
hsa-miR-873	chr17	30331755	30334133	+	175.00	LIG3
hsa-miR-873	chr17	40174347	40179976	+	169.00	DBF4B
hsa-miR-339-3p	chr5	176452392	176452938	+	168.00	FGFR4
hsa-miR-1537	chr15	84000023	84002771	+	175.00	AKAP13
hsa-miR-371-5p	chr11	4060789	4061068	+	170.00	STIM1
hsa-miR-1300	chr17	40174347	40179976	+	174.00	DBF4B
hsa-miR-127-3p	chr15	84000023	84006622	+	165.00	AKAP13
hsa-miR-518a-3p	chr15	42482545	42492825	+	173.00	CASC4
hsa-miR-409-5p	chr11	4061305	4064282	+	170.00	STIM1
hsa-miR-758	chr19	35000289	35003448	+	170.00	CCNE1
hsa-miR-934	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-miR-632	chr11	8890657	8892948	+	165.00	C11orf17
hsa-miR-371-5p	chr6	45588123	45620931	+	168.00	RUNX2
hsa-miR-371-5p	chr6	30971271	30972376	+	172.00	DDR1
hsa-miR-483-5p	chr11	114554706	114585503	-	168.00	CADM1
hsa-miR-875-5p	chr15	42460467	42482376	+	168.00	CASC4
hsa-miR-1300	chr6	45588123	45620931	+	167.00	RUNX2
hsa-miR-148b	chr6	45588123	45620931	+	169.00	RUNX2
hsa-miR-518f	chr17	22994774	22996466	+	168.00	LGALS9

hsa-miR-483-5p	chr6	30970428	30971159	+	167.00	DDR1
hsa-miR-1231	chr11	114554706	114585503	-	167.00	CADM1
hsa-miR-483-5p	chr6	30970428	30972376	+	167.00	DDR1
hsa-miR-361-5p	chr1	241608789	241645626	+	170.00	SDCCAG8
hsa-miR-218	chr11	4061305	4064282	+	168.00	STIM1
hsa-miR-28-5p	chr19	35000289	35003448	+	166.00	CCNE1
hsa-miR-219-2-3p	chr11	114554706	114585503	-	170.00	CADM1
hsa-miR-371-5p	chr6	30971271	30972376	+	172.00	DDR1
hsa-miR-1913	chr6	167347046	167355886	+	175.00	FGFR1OP
hsa-miR-202	chr15	84000023	84002771	+	170.00	AKAP13
hsa-miR-513a-5p	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-miR-516b	chr15	84002826	84008797	+	168.00	AKAP13
hsa-miR-610	chr11	114554706	114585503	-	173.00	CADM1
hsa-miR-371-5p	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-miR-518f	chr17	22994774	22996466	+	168.00	LGALS9
hsa-miR-193a-3p	chr11	63812517	63812650	+	165.00	GPR137
hsa-miR-518d-3p	chr17	22994774	22996466	+	165.00	LGALS9
hsa-miR-591	chr11	114554706	114585503	-	167.00	CADM1
hsa-miR-200a	chr1	241645755	241647892	+	175.00	SDCCAG8
hsa-miR-541	chr15	84002826	84008797	+	166.00	AKAP13
hsa-miR-361-5p	chr6	45588123	45620931	+	172.00	RUNX2
hsa-miR-200a	chr22	27471997	27477228	+	166.00	HSC20
hsa-miR-1282	chr1	241608789	241645626	+	168.00	SDCCAG8
hsa-miR-1913	chr2	111595237	111597780	+	177.00	BCL2L11
hsa-miR-516b	chr1	241608789	241645626	+	196.00	SDCCAG8
hsa-miR-519c-5p	chr6	167347046	167355886	+	166.00	FGFR1OP
hsa-miR-193a-3p	chr19	35003596	35004468	+	170.00	CCNE1
hsa-miR-1294	chr6	30969180	30970261	+	172.00	DDR1
hsa-miR-1295	chr17	30331755	30334133	+	165.00	LIG3
hsa-miR-507	chr15	42482545	42492825	+	170.00	CASC4
hsa-miR-518d-3p	chr17	22994774	22996466	+	165.00	LGALS9
hsa-miR-483-5p	chr6	30970428	30972376	+	167.00	DDR1
hsa-miR-875-5p	chr1	241608789	241645626	+	178.00	SDCCAG8
hsa-miR-200b	chr6	45588123	45620931	+	167.00	RUNX2
hsa-miR-632	chr17	22993502	22996466	+	166.00	LGALS9
hsa-miR-483-5p	chr6	30970428	30971159	+	167.00	DDR1

hsa-miR-1280	chr15	84002826	84008797	+	169.00	AKAP13
hsa-miR-10b	chr11	114554706	114585503	-	178.00	CADM1
hsa-miR-147	chr1	241608789	241645626	+	166.00	SDCCAG8
hsa-miR-10a	chr11	114554706	114585503	-	180.00	CADM1
hsa-miR-192	chr15	42460467	42482376	+	171.00	CASC4
hsa-miR-1276	chr6	167344372	167346985	+	169.00	FGFR1OP
hsa-miR-152	chr1	241608789	241645626	+	166.00	SDCCAG8
hsa-miR-483-5p	chr10	74869666	74874488	-	167.00	PPP3CB
hsa-miR-1913	chr11	4061305	4064282	+	171.00	STIM1
hsa-miR-1197	chr11	114554706	114585503	-	171.00	CADM1
hsa-miR-516b	chr6	45588123	45620931	+	171.00	RUNX2
hsa-miR-937	chr2	111595237	111597780	+	169.00	BCL2L11
hsa-miR-660	chr6	45588123	45620931	+	169.00	RUNX2
hsa-miR-1287	chr6	45588123	45620931	+	172.00	RUNX2
hsa-miR-10a	chr15	42460467	42482376	+	172.00	CASC4
hsa-miR-1261	chr1	241608789	241645626	+	168.00	SDCCAG8
hsa-miR-210	chr6	167347046	167355886	+	169.00	FGFR1OP
hsa-miR-361-5p	chr15	42460467	42482376	+	165.00	CASC4
hsa-miR-770-5p	chr17	40171319	40174229	+	166.00	DBF4B
hsa-miR-129-5p	chr17	41423279	41424662	+	173.00	MAPT
hsa-miR-507	chr1	241608789	241645626	+	172.00	SDCCAG8
hsa-miR-1914	chr11	8890657	8892948	+	168.00	C11orf17
hsa-miR-615-3p	chr17	40174347	40179976	+	169.00	DBF4B
hsa-miR-1280	chr15	84006689	84008797	+	169.00	AKAP13
hsa-miR-876-3p	chr10	74869666	74874488	-	167.00	PPP3CB
hsa-miR-202	chr15	84000023	84006622	+	170.00	AKAP13
hsa-miR-1308	chr22	27471997	27477228	+	165.00	HSC20
hsa-miR-371-5p	chr6	30970428	30972376	+	172.00	DDR1
hsa-miR-1539	chr2	111595237	111597780	+	167.00	BCL2L11
hsa-miR-524-5p	chr5	162824387	162827287	+	166.00	HMMR
hsa-miR-299-5p	chr15	42460467	42482376	+	166.00	CASC4

hsa-miR-1913	chr11	63812517	63812650	+	175.00	GPR137
hsa-miR-425	chr15	42460467	42482376	+	170.00	CASC4
hsa-miR-202	chr4	87891302	87893185	+	167.00	PTPN13
hsa-miR-1322	chr1	241608789	241645626	+	171.00	SDCCAG8
hsa-miR-412	chr17	40174347	40179976	+	165.00	DBF4B
hsa-miR-208b	chr6	167347046	167355886	+	166.00	FGFR1OP
hsa-miR-541	chr11	114554706	114585503	-	167.00	CADM1
hsa-miR-877	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-miR-663	chr17	41420299	41423080	+	165.00	MAPT
hsa-miR-1275	chr11	63812517	63812650	+	165.00	GPR137
hsa-miR-1275	chr17	30331755	30334133	+	175.00	LIG3
hsa-miR-412	chr2	111595237	111597780	+	167.00	BCL2L11
hsa-miR-200a	chr11	114554706	114585503	-	170.00	CADM1
hsa-miR-634	chr2	111595237	111597780	+	176.00	BCL2L11
hsa-miR-891a	chr11	8890657	8892948	+	171.00	C11orf17
hsa-miR-371-5p	chr6	30970428	30972376	+	172.00	DDR1
hsa-miR-519b-5p	chr6	167347046	167355886	+	166.00	FGFR1OP
hsa-miR-519c-5p	chr11	114554706	114585503	-	170.00	CADM1

Table 6.2: Predicted miRNA targets in ASEs affected by siAGO1 knockdown in Hep3B.

Mature miRNA annotation source MiRbase database (Griffiths-Jones et al., 2008; Kozomara and Griffiths-Jones, 2011; Griffiths-Jones et al., 2006; Griffiths-Jones, 2004)

miRNA	chr	start	end	strand	score	RefseqID
hsa-miR-193a-3p	chr19	35003596	35004468	+	170.00	CCNE1
hsa-miR-127-3p	chr15	84002826	84008797	+	165.00	AKAP13
hsa-miR-30b	chr6	145183848	145190176	+	182.00	UTRN
hsa-miR-142-3p	chr15	84002826	84008797	+	167.00	AKAP13
hsa-miR-1539	chr2	111595237	111597780	+	167.00	BCL2L11
hsa-miR-877	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-miR-142-3p	chr15	84000023	84006622	+	167.00	AKAP13
hsa-miR-642	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-miR-892a	chr20	30853938	30856804	+	170.00	DNMT3B
hsa-miR-499-5p	chr15	84000023	84002771	+	166.00	AKAP13
hsa-miR-634	chr19	35000289	35003448	+	167.00	CCNE1
hsa-miR-133a	chr5	112228329	112230999	+	165.00	SRP19
hsa-miR-1537	chr15	84000023	84006622	+	175.00	AKAP13
hsa-miR-499-5p	chr15	84000023	84006622	+	166.00	AKAP13
hsa-miR-16	chr3	194818321	194819351	+	167.00	OPA1
hsa-miR-30d	chr15	84000023	84006622	+	165.00	AKAP13
hsa-miR-30b	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-miR-30d	chr6	145183848	145190176	+	171.00	UTRN
hsa-miR-301b	chr3	194819420	194826574	+	173.00	OPA1
hsa-miR-513a-5p	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-miR-151-3p	chr11	8890657	8892948	+	182.00	C11orf17
hsa-miR-142-3p	chr3	194819420	194826574	+	167.00	OPA1
hsa-miR-1537	chr15	84000023	84002771	+	175.00	AKAP13
hsa-miR-127-3p	chr15	84000023	84006622	+	165.00	AKAP13
hsa-miR-140-5p	chr20	30853938	30856804	+	173.00	DNMT3B
hsa-miR-133b	chr5	112228329	112230999	+	165.00	SRP19
hsa-miR-934	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-miR-520e	chr19	35000289	35003448	+	169.00	CCNE1
hsa-miR-634	chr2	111595237	111597780	+	176.00	BCL2L11
hsa-miR-16	chr3	194818321	194819351	+	167.00	OPA1

Table 6.3: Predicted miRNA targets in ASEs affected by siAGO1 knockdown in HeLa.

Mature miRNA annotation source Argonaute database (Shahi et al., 2006)

miRNA	chr	start	end	strand	score	RefseqID
hsa-miR-142-3p	chr3	194819420	194826574	+	167.00	OPA1
hsa-miR-152	chr1	241608789	241645626	+	166.00	SDCCAG8
hsa-miR-151-3p	chr20	30853938	30856804	+	174.00	DNMT3B
hsa-miR-611	chr5	112228329	112230999	+	165.00	SRP19
hsa-miR-642	chr20	30857795	30859228	+	186.00	DNMT3B
hsa-miR-770-5p	chr17	40171319	40174229	+	166.00	DBF4B
hsa-miR-154	chr20	30853938	30856804	+	165.00	DNMT3B
hsa-miR-1282	chr1	241608789	241645626	+	168.00	SDCCAG8
hsa-miR-615-3p	chr17	40174347	40179976	+	169.00	DBF4B
hsa-miR-937	chr2	111595237	111597780	+	169.00	BCL2L11
hsa-miR-1914	chr11	8890657	8892948	+	168.00	C11orf17
hsa-miR-1280	chr15	84006689	84008797	+	169.00	AKAP13
hsa-miR-1910	chr8	42315745	42321627	+	165.00	POLB
hsa-miR-668	chr3	194819420	194826574	+	168.00	OPA1
hsa-miR-1322	chr1	241608789	241645626	+	171.00	SDCCAG8
hsa-mir-522*	chr11	114554706	114585503	-	170.00	CADM1
hsa-mir-340*	chr6	45588123	45620931	+	166.00	RUNX2
hsa-mir-875-5p	chr6	167347046	167355886	+	169.00	FGFR1OP
hsa-mir-522*	chr19	35000289	35003448	+	166.00	CCNE1
hsa-mir-196b	chr15	42482545	42492825	+	188.00	CASC4
hsa-mir-148b	chr6	45588123	45620931	+	169.00	RUNX2
hsa-mir-132*	chr17	22993502	22996466	+	174.00	LGALS9
hsa-mir-519c-5p	chr19	35000289	35003448	+	166.00	CCNE1
hsa-mir-522*	chr6	167347046	167355886	+	166.00	FGFR1OP
hsa-mir-16-1*	chr11	114554706	114585503	-	166.00	CADM1
hsa-mir-132*	chr6	30970428	30972376	+	166.00	DDR1
hsa-mir-519a-1*	chr19	35000289	35003448	+	166.00	CCNE1
hsa-mir-516b-1	chr11	4061305	4064282	+	166.00	STIM1
hsa-mir-412	chr17	40174347	40179976	+	165.00	DBF4B

hsa-mir-218-1	chr11	4061305	4064282	+	168.00	STIM1
hsa-mir-147	chr1	241608789	241645626	+	166.00	SDCCAG8
hsa-mir-24-1*	chr17	22993502	22996466	+	169.00	LGALS9
hsa-mir-937	chr2	111595237	111597780	+	169.00	BCL2L11
hsa-mir-520e	chr19	35000289	35003448	+	169.00	CCNE1
hsa-mir-758	chr22	27471997	27477228	+	171.00	HSC20
hsa-mir-425	chr1	241608789	241645626	+	170.00	SDCCAG8
hsa-mir-518e*	chr11	114554706	114585503	-	170.00	CADM1
hsa-mir-92a-2*	chr17	40174347	40179976	+	165.00	DBF4B
hsa-mir-196b	chr11	114554706	114585503	-	167.00	CADM1
hsa-mir-518a-1-3p	chr15	42482545	42492825	+	173.00	CASC4
hsa-mir-518c	chr17	22993502	22996466	+	178.00	LGALS9
hsa-mir-202	chr4	87891302	87893185	+	167.00	PTPN13
hsa-mir-339-3p	chr5	176452392	176452938	+	168.00	FGFR4
hsa-mir-183*	chr15	42482545	42492825	+	167.00	CASC4
hsa-mir-154*	chr15	42460467	42482376	+	167.00	CASC4
hsa-mir-154*	chr19	35003596	35004468	+	173.00	CCNE1
hsa-mir-634	chr2	111595237	111597780	+	176.00	BCL2L11
hsa-mir-632	chr11	8890657	8892948	+	165.00	C11orf17
hsa-mir-770-5p	chr17	40171319	40174229	+	166.00	DBF4B
hsa-mir-523*	chr6	167347046	167355886	+	166.00	FGFR1OP
hsa-mir-24-1*	chr17	22993502	22994677	+	169.00	LGALS9
hsa-mir-200a	chr22	27471997	27477228	+	166.00	HSC20
hsa-mir-933	chr11	114554706	114585503	-	179.00	CADM1
hsa-mir-610	chr11	114554706	114585503	-	173.00	CADM1
hsa-mir-193a-3p	chr11	63812517	63812650	+	165.00	GPR137
hsa-mir-183*	chr17	40174347	40179976	+	165.00	DBF4B
hsa-mir-891a	chr11	114554706	114585503	-	166.00	CADM1
hsa-mir-148b*	chr1	241608789	241645626	+	168.00	SDCCAG8
hsa-mir-541	chr11	114554706	114585503	-	167.00	CADM1
hsa-mir-202	chr15	84000023	84002771	+	170.00	AKAP13

hsa-mir-376a-1*	chr15	42482545	42492825	+	170.00	CASC4
hsa-mir-516b-1	chr15	84002826	84008797	+	168.00	AKAP13
hsa-mir-223	chr15	42482545	42492825	+	167.00	CASC4
hsa-mir-16-1*	chr17	40174347	40179976	+	178.00	DBF4B
hsa-mir-507	chr1	241608789	241645626	+	172.00	SDCCAG8
hsa-mir-183*	chr1	241608789	241645626	+	166.00	SDCCAG8
hsa-mir-132*	chr6	30971271	30972376	+	166.00	DDR1
hsa-mir-129-1-5p	chr17	41423279	41424662	+	173.00	MAPT
hsa-mir-412	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-mir-487b	chr1	241608789	241645626	+	166.00	SDCCAG8
hsa-mir-132*	chr6	167347046	167355886	+	165.00	FGFR1OP
hsa-mir-507	chr15	42482545	42492825	+	170.00	CASC4
hsa-mir-361-5p	chr15	42460467	42482376	+	165.00	CASC4
hsa-mir-516b-2	chr1	241608789	241645626	+	196.00	SDCCAG8
hsa-mir-361-5p	chr1	241608789	241645626	+	170.00	SDCCAG8
hsa-mir-412	chr6	167347046	167355886	+	167.00	FGFR1OP
hsa-mir-299-5p	chr15	42460467	42482376	+	166.00	CASC4
hsa-mir-96*	chr15	42482545	42492825	+	165.00	CASC4
hsa-mir-518c	chr17	22994774	22996466	+	178.00	LGALS9
hsa-mir-132*	chr17	22994774	22996466	+	174.00	LGALS9
hsa-mir-409-5p	chr11	4061305	4064282	+	170.00	STIM1
hsa-mir-183*	chr15	84000023	84006622	+	166.00	AKAP13
hsa-mir-513-1-5p	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-mir-192	chr15	42460467	42482376	+	171.00	CASC4
hsa-mir-193a-3p	chr19	35003596	35004468	+	170.00	CCNE1
hsa-mir-28-5p	chr19	35000289	35003448	+	166.00	CCNE1
hsa-mir-376a-1*	chr1	241608789	241645626	+	166.00	SDCCAG8
hsa-mir-499-5p	chr15	84000023	84002771	+	166.00	AKAP13
hsa-mir-92a-2*	chr6	45588123	45620931	+	167.00	RUNX2
hsa-mir-571	chr11	114554706	114585503	-	169.00	CADM1
hsa-mir-221*	chr6	45588123	45620931	+	166.00	RUNX2
hsa-mir-518e*	chr19	35000289	35003448	+	166.00	CCNE1
hsa-mir-16-1*	chr6	30971271	30972376	+	166.00	DDR1
hsa-mir-219-2-3p	chr10	74869666	74874488	-	168.00	PPP3CB
hsa-mir-10a	chr11	114554706	114585503	-	180.00	CADM1
hsa-mir-873	chr17	30331755	30334133	+	175.00	LIG3
hsa-mir-181a-1*	chr2	111595237	111597780	+	167.00	BCL2L11
hsa-mir-634	chr19	35000289	35003448	+	167.00	CCNE1
hsa-mir-591	chr11	114554706	114585503	-	167.00	CADM1

hsa-mir-19b-2*	chr6	167347046	167355886	+	172.00	FGFR1OP
hsa-mir-146b-5p	chr15	42460467	42482376	+	171.00	CASC4
hsa-mir-934	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-mir-516b-2	chr15	84002826	84008797	+	168.00	AKAP13
hsa-mir-132*	chr17	22994774	22996466	+	174.00	LGALS9
hsa-mir-891a	chr11	8890657	8892948	+	171.00	C11orf17
hsa-mir-708*	chr6	45588123	45620931	+	181.00	RUNX2
hsa-mir-523*	chr11	114554706	114585503	-	170.00	CADM1
hsa-mir-92a-2*	chr11	114554706	114585503	-	171.00	CADM1
hsa-mir-125a-3p	chr22	27471997	27477228	+	165.00	HSC20
hsa-mir-16-1*	chr10	74869666	74874488	-	167.00	PPP3CB
hsa-mir-541	chr15	84002826	84008797	+	166.00	AKAP13
hsa-mir-877	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-mir-412	chr2	111595237	111597780	+	167.00	BCL2L11
hsa-mir-96*	chr6	45588123	45620931	+	167.00	RUNX2
hsa-mir-19b-2*	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-mir-200a	chr11	114554706	114585503	-	170.00	CADM1
hsa-mir-208b	chr6	167347046	167355886	+	166.00	FGFR1OP
hsa-mir-183*	chr6	45588123	45620931	+	166.00	RUNX2
hsa-mir-210	chr6	167347046	167355886	+	169.00	FGFR1OP
hsa-mir-758	chr19	35000289	35003448	+	170.00	CCNE1
hsa-mir-27b*	chr15	42482545	42492825	+	181.00	CASC4
hsa-mir-660	chr6	45588123	45620931	+	169.00	RUNX2
hsa-mir-516b-2	chr6	45588123	45620931	+	171.00	RUNX2
hsa-mir-218-2	chr11	4061305	4064282	+	168.00	STIM1
hsa-mir-200b	chr6	45588123	45620931	+	167.00	RUNX2
hsa-mir-523*	chr19	35000289	35003448	+	166.00	CCNE1
hsa-mir-23b*	chr1	241608789	241645626	+	175.00	SDCCAG8
hsa-mir-708*	chr11	114554706	114585503	-	168.00	CADM1
hsa-mir-101-1	chr11	114554706	114585503	-	172.00	CADM1
hsa-mir-10b	chr11	114554706	114585503	-	178.00	CADM1
hsa-mir-367*	chr10	74869666	74874488	-	166.00	PPP3CB
hsa-mir-518f	chr17	22993502	22996466	+	168.00	LGALS9
hsa-mir-632	chr17	22993502	22996466	+	166.00	LGALS9
hsa-mir-127-3p	chr15	84002826	84008797	+	165.00	AKAP13
hsa-mir-513-2-5p	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-mir-875-5p	chr15	42460467	42482376	+	168.00	CASC4

Table 6.4: Predicted miRNA targets in ASEs affected by siAGO1 knockdown in Hep3B.

Mature miRNA annotation source Argonaute database (Shahi et al., 2006)

miRNA	chr	start	end	strand	score	RefseqID
hsa-mir-142-3p	chr3	194819420	194826574	+	167.00	OPA1
hsa-mir-513-1-5p	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-mir-193a-3p	chr19	35003596	35004468	+	170.00	CCNE1
hsa-mir-301b	chr3	194819420	194826574	+	173.00	OPA1
hsa-mir-668	chr3	194819420	194826574	+	168.00	OPA1
hsa-mir-193a-3p	chr11	63812517	63812650	+	165.00	GPR137
hsa-mir-133a-1	chr20	30853938	30856804	+	167.00	DNMT3B
hsa-mir-499-5p	chr15	84000023	84002771	+	166.00	AKAP13
hsa-mir-142-3p	chr3	194819420	194826574	+	167.00	OPA1
hsa-mir-16-2	chr3	194818321	194819351	+	167.00	OPA1
hsa-mir-133b	chr5	112228329	112230999	+	165.00	SRP19
hsa-mir-148b*	chr1	241608789	241645626	+	168.00	SDCCAG8
hsa-mir-30a*	chr1	241608789	241645626	+	173.00	SDCCAG8
hsa-mir-30d	chr15	84000023	84006622	+	165.00	AKAP13
hsa-mir-301b	chr3	194819420	194826574	+	173.00	OPA1
hsa-mir-16-1	chr3	194818321	194819351	+	167.00	OPA1
hsa-mir-152	chr1	241608789	241645626	+	166.00	SDCCAG8
hsa-mir-26b*	chr6	145183848	145190176	+	171.00	UTRN
hsa-mir-181a-1*	chr2	111595237	111597780	+	167.00	BCL2L11
hsa-mir-151-3p	chr20	30853938	30856804	+	174.00	DNMT3B
hsa-mir-634	chr19	35000289	35003448	+	167.00	CCNE1
hsa-mir-499-5p	chr15	84000023	84006622	+	166.00	AKAP13
hsa-mir-30a*	chr5	112228329	112230999	+	165.00	SRP19
hsa-mir-892a	chr20	30853938	30856804	+	170.00	DNMT3B
hsa-mir-611	chr5	112228329	112230999	+	165.00	SRP19
hsa-mir-615-3p	chr17	40174347	40179976	+	169.00	DBF4B
hsa-mir-668	chr3	194819420	194826574	+	168.00	OPA1
hsa-mir-642	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-mir-23b*	chr1	241608789	241645626	+	175.00	SDCCAG8
hsa-mir-625*	chr1	241608789	241645626	+	171.00	SDCCAG8
hsa-mir-30d	chr15	84002826	84008797	+	165.00	AKAP13
hsa-mir-411*	chr6	145183848	145190176	+	168.00	UTRN
hsa-mir-642	chr20	30857795	30859228	+	186.00	DNMT3B
hsa-mir-133b	chr20	30853938	30856804	+	171.00	DNMT3B

hsa-mir-487b	chr1	241608789	241645626	+	166.00	SDCCAG8
hsa-mir-934	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-mir-30a*	chr15	84006689	84008797	+	165.00	AKAP13
hsa-mir-30d	chr6	145183848	145190176	+	171.00	UTRN
hsa-mir-30a*	chr15	84002826	84008797	+	165.00	AKAP13
hsa-mir-140-5p	chr20	30853938	30856804	+	173.00	DNMT3B
hsa-mir-935	chr1	181782096	181784965	+	168.00	SMG7
hsa-mir-133a-2	chr20	30853938	30856804	+	167.00	DNMT3B
hsa-mir-142-3p	chr15	84000023	84006622	+	167.00	AKAP13
hsa-mir-26b	chr1	241608789	241645626	+	169.00	SDCCAG8
hsa-mir-634	chr2	111595237	111597780	+	176.00	BCL2L11
hsa-mir-16-2	chr3	194818321	194819351	+	167.00	OPA1
hsa-mir-770-5p	chr17	40171319	40174229	+	166.00	DBF4B
hsa-mir-127-3p	chr15	84000023	84006622	+	165.00	AKAP13
hsa-mir-222	chr3	194826686	194832094	+	166.00	OPA1
hsa-mir-154	chr20	30853938	30856804	+	165.00	DNMT3B
hsa-mir-877	chr1	241608789	241645626	+	167.00	SDCCAG8
hsa-mir-133a-2	chr5	112228329	112230999	+	165.00	SRP19

6.3 Appendix C

Peer-reviewed publications

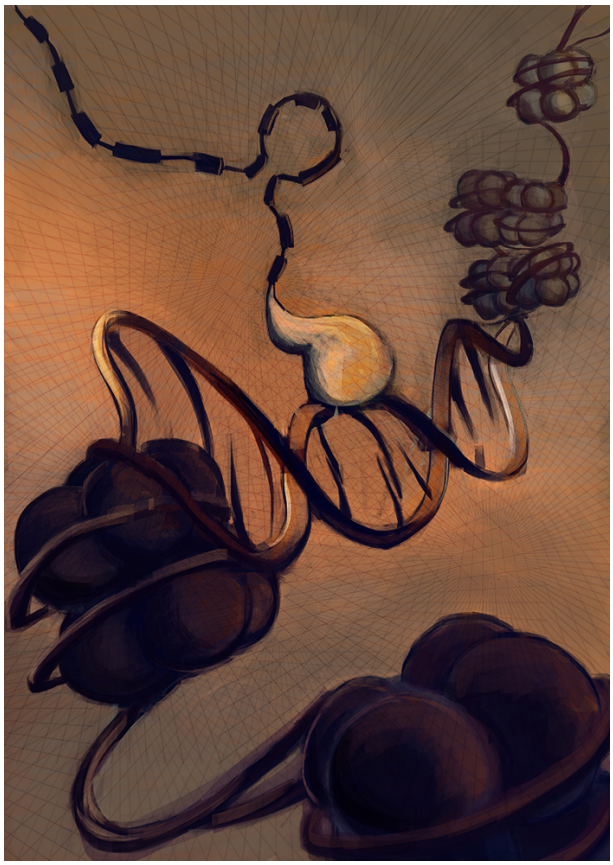
Alló M, Buggiano V, Fededa JP, Petrillo E, Schor I, de la Mata M, **Agirre E**, Plass M, Eyras E, Elela SA,Klinck R, Chabot B, Kornblihtt AR. (2009) Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat Struct Mol Biol*. 16(7):717-24.

Plass M, **Agirre E**, Reyes D, Camara F, Eyras E. (2008). Co-evolution of the branch site and SR proteins in eukaryotes. *Trens Genet*. 24(12):590-4.

Review articles

Agirre E Eyras E. (2011). Databases and resources for human small non-coding RNAs. *Human genomics*. Vol 5, no3. 102-199.

Alló M, Schor IE, Muoz MJ, de la Mata M, **Agirre E**, Valcrcel J, Eyras E, Kornblihtt AR. (2011). Chromatin and alternative splicing. *Cold Spring Harb Symp Quant Biol*. 75:103-11.



Remember this: ”Be it a rock or a grain of sand, in water they sink as the same.”

Cover design Amagoia Agirre (amagoia.agirre@gmail.com)